Social Context in Massive Open Online Courses

by

Oleksandra Poquet

Specialist degree with distinction; M.A.

A thesis submitted for the degree of

Doctor of Philosophy

School of Education

Division of Education, Arts and Social Sciences



University of
South Australia

March, 2017

Abstract

Since 2008, millions of learners worldwide have participated in massive open online courses (MOOCs). Despite the massive uptake, the quality of instruction and learning in MOOCs remains contested. However, designing for quality learning in MOOCs is complicated by the open participation, and the scale of diversity in learner backgrounds, motivations and the characteristics of open online environments.

The complexities associated with learning in MOOCs are not well aligned to the formal educational models and theories. In particular, contemporary theories of teaching and learning in formal education are based on the development of interpersonal interactions. The emphasis on peer and instructor relationships has been shown to impact on academic performance, as well as learner motivation and satisfaction. In MOOCs, anonymity, volume of interactions and shifting boundaries of the cohort cast doubt if similar learning outcomes can be achieved. This raises questions related to what models and theories of learning and teaching are effective in such emergent, open and scaled education settings.

This thesis explores the social context in MOOCs by analysing patterns and quality of interpersonal interactions occurring within four edX MOOC forums ($N_{ForumPosters1}$=7168; $N_{ForumPosters2}$=8348; $N_{ForumPosters3}$=17537; $N_{ForumPosters4}$=15796). Initial analysis delineates the sub-group of posters comparable to the learning groups in formal education settings. The thesis then examines the structure and nature of this sub-group using network analysis, as well as qualitative content and thematic analyses. The integration of the quantitative and qualitative methods allows for a novel conceptualisation of the community emerging from this group of posters in the MOOC setting. The indicators quantifying the development of the social context are validated by using statistical network modelling.

The outcomes of the thesis provide for a new conceptualisation of how social context evolves in MOOC forums and the associated indicators. The social context in MOOCs develops at multiple levels: an emergent forum core community is situated within other interactions devoid of the same relational quality. Interactions by the forum core posters

who may evolve into a community represent only a fraction of the total forum activity. These interactions have the potential to evolve into interpersonal relationships as defined by the shared temporal history and underpinned by the relational quality of attachment. The findings suggest that forum core develops as an identity-based community that is characterised by the group-level attachment rather than bond-based attachment indicative of formal education. Analysis validates the indicators of reciprocity to reflect the network structure in these communities that differentiate the extent of group cohesion and solidarity underpinning them.

The results have significance for the theory, practice and evaluation of open online educational environments. The study supports that cognitive and emotional learning outcomes enabled by the social context can be facilitated in MOOCs. The study found that the MOOC social context does differ qualitatively from formal education context. Researchers and educationalists should account for these differences in any applied analytical frameworks, research methods, and learning designs. To advance the group-level analysis in open education settings, the thesis presents suggested future directions for learning analytics research.

Table of Contents

Abstract iii
Table of Contents v
List of Figures xi
List of Tables xii
Acronyms xiii
Candidate's Declaration xv
Acknowledgements xvii
Published Work xix

CHAPTER 1. INTRODUCTION

1.1 Background 1
1.2 Problem Formulation 3
1.3 Study's Aim, Objectives and Research Question 9
1.4 Study Design 11
1.5 Structure of the Thesis 13
1.6 Summary 15

CHAPTER 2. LITERATURE REVIEW

2.1 The Evolution of MOOCs 16
2.1.1 Drivers for Online Education Adoption 17
2.1.2 The Emergence of the MOOCs 20
2.1.3 Public Debate 23
2.1.4 Resolution of the Tensions Around Scale in MOOCs 25
2.1.5 Latest Trends for MOOCs 28
2.1.6 Section Summary 31
2.2 Formal Perspectives on Non-Formal Courses 31
2.2.1 Learner Success in MOOCs 33
2.2.2 Interpersonal Interactions and Performance 34
2.2.3 Classifying MOOC Learners 36
2.2.4 Recognising the Differences Between MOOCs and Formal Courses 37
2.2.5 Section Summary 41

2.3 The Role of Interpersonal Interactions 42
2.3.1 Interpersonal Interactions and Learning 43
2.3.2 Interpersonal Interactions in Distance Education 49
2.3.3 Section Summary 52
2.4 MOOC Research on Interpersonal Interactions 53
2.4.1 Can Relationships Form in MOOCs? 54
2.4.2 Are There Communities? 57
2.4.3 What Network Structure Explains MOOC Forum Activity? 59
2.4.4 Section Summary 62
2.5 Chapter Summary 64

CHAPTER 3. METHODOLOGY
3.1 Researcher's Assumptions 67
3.1.1 Research Domain: Learning Analytics 67
3.1.2 Research Orientation: Deweyan Pragmatism 69
3.2 Research Questions 71
3.3 Design and Methods 73
3.4 Description of Analysed Courses 75
3.4.1 FP101x – Functional Programming MOOC 77
3.4.2 ET3034Tux – Solar Energy MOOC 78
3.4.3 CTB3365x – Water Treatment MOOC 78
3.4.4 EX101x – Taking Data to the Max MOOC (with Excel) 79
3.5 Ethics 80
3.6 Chapter Summary 80

CHAPTER 4. EXPLORATION OF FORUM LOG DATA
4.1 Introduction 82
4.2 Framing Forum Participation 83
4.2.1 Forum Use in MOOCs 84
4.2.2 Measuring Regularity of Interactions 89
4.2.3 The Role of the Environment: Receiving Timely Replies 90

4.3 Results of the Data Analysis 92

4.3.1 Describing Participation in Analysed Courses 92

4.3.2 Typology of Forum Use 96

4.3.3 How Much is the Forum Used by Different Types of Forum Participants? 100

4.4 Discussion 103

4.5 Conclusion 106

CHAPTER 5. POTENTIAL FOR RELATIONSHIP FORMATION IN edX FORUMS

5.1 Introduction 107

5.2 Research Questions 108

5.3 Methods 109

5.3.1 Network Construction 109

5.3.2 Temporal Network Boundaries 113

5.3.3 Descriptive Network Analysis 115

5.3.4 Statistical Analysis of Network Measures 118

5.4 Results 119

5.4.1 Network Analysis of Co-occurrence Networks 119

5.4.2 Co-Occurrence Over the Course of a MOOC's Progression 128

5.4.3 Effect of Time on Degree and Edge Weight Accumulation across different Forum Participation Types 131

5.4.4 Differences in Frequency of Co-Occurrence 134

5.4.5 Summary of Results 136

5.5 Discussion 137

5.6 Conclusion 140

CHAPTER 6. FORUM CORE AND GROUPS IN FORMAL ONLINE COURSES

6.1 Introduction 141

6.2 Analysing Student-Produced Text in CMC and MOOCs 142

6.2.1 CMC and the Development of Social Context 143

6.2.2 Automated Content Analysis in MOOCs 146

6.2.3 Summary 150
6.3 Research Questions 151
6.4 Methods 151
6.4.1 Unit of Analysis 152
6.4.2 Manual Content Analysis in MOOCs 152
6.4.3 Coding Framework 154
6.4.4 Process of Coding and Validity 159
6.4.5 Inter-rater Reliability 160
6.5 Data Wrangling 161
6.6 Results 162
6.6.1 Dominant Discussion Types 163
6.6.2 The Change of Discussions Throughout the Course 166
6.6.3 Discussions Central to Network Formation 174
6.6.4 Summary of Findings from Research Questions 177
6.7 Discussion 178
6.7.1 Forum Core as a Curriculum-Driven Group 179
6.7.2 Relational and Affective Aspects of Forum Core Development 180
6.7.3 Comparing Forum Core with Groups in Formal Online Courses 183
6.8 Conclusion 184

CHAPTER 7. IDENTITY FORMATION PROCESSES IN FORUM CORE

7.1 Introduction 186
7.2 Literature Review 187
7.2.1 Understanding Attachment in Group Formation 187
7.2.2 Participation and Communities of Practice 189
7.3 Research Question 191
7.4 Methods 192
7.5 Forum Core Discussions Beyond the Curriculum 192
7.5.1 Discussing Assessment 193
7.5.2 Discussing Teacher, Staff and Course Organisation 195
7.5.3 Discussing Communication, Rules and Norms 198

7.5.4 Summary 205
7.6 Discussion 205
7.6.1 Forum Core as an Emergent Community of Practice (CoP) 205
7.6.1 Mechanisms of Identity Formation in a Forum Core 206
7.6.1.1 Negotiating Control as a Mechanism of Relatedness 207
7.6.1.2 Negotiating Identity Through Situational Norms 208
7.6.1.3 Wenger's Modes of Identification 209
7.6.2 Implications 210
7.7 Conclusions 211

CHAPTER 8. MODELLING A FORUM CORE AS AN IDENTITY-BASED COMMUNITY

8.1 Introduction 213
8.2 Analysis of Network Structure 214
8.2.1 Formalist and Relational Network Analysis 215
8.2.2 Network Exchange Patterns in Online Communities 217
8.3 Methods 220
8.3.1 Network Construction 221
8.3.2 Exponential Random Graph Modelling 221
8.3.3 Modelled Configurations and Effects 223
8.4 Results 227
8.5 Discussion 234
8.6 Conclusions 238

CHAPTER 9. CONCLUSION

9.1 Summary of the Findings 240
9.2 The Contribution of the Thesis 243
9.3 Implications 247
9.3.1 Transfer of Theories and Models 248
9.3.2 Methodological Implications 249
9.3.2 Implications for Designing Technology and Evaluating Interventions 250

9.3.4 Implications for Pedagogy 250

9.4 Limitations **251**

9.5 Future Research **252**

9.5.1 Relational Model of the Components of Social Learning in a MOOC Setting 253

9.5.2 Indicators of Identity-Based Communities 254

9.5.3 Temporal Analysis of the Social Context Development 255

9.6 Final Note **256**

REFERENCES…….…….257
APPENDIX A……..…….298
APPENDIX B……...….300
APPENDIX C……....312
APPENDIX D……....315
APPENDIX E……322
APPENDIX F……326
APPENDIX G……....327

List of Figures

Figure 1.1 Schematic representation of the Biggs Presage-Product-Process Model for student learning in higher education (simplified from Biggs & Tang, 2011) 4
Figure 3.1 Schematic representation of the mixed method sequential design of the research. 74
Figure 4.1 Relative break-down of main forum user types, per course. 99
Figure 4.2 Forum access by different forum users (Table 4.7) in FP 101
Figure 4.3 Forum access by different forum users (Table 4.7) in Solar 102
Figure 5.1 Centralisation of the networks of the entire cohort, regular posters and forum core, across four MOOCs 122
Figure 5.2 Degree distribution across the networks, with outliers removed. 125
Figure 5.3 Edge weight distribution across networks, with outliers removed 126
Figure 5.4 Distribution of average edge weight per person in each network, outliers removed 127
Figure 5.5 Accumulation of edges and ties in the co-occurrence networks in a Solar Course. 129
Figure 5.6 Visualised effects of time over normalised degree and normalised edges for different participation types 134
Figure 6.1 Ratio of the sum of ST discussion and SN discussions to the sum of CT and CN discussions threads at comparable time stages, in four MOOC forums 169
Figure 6.2 Ratio of CT to CN conversations at comparable time stages, in four MOOC forums. 169
Figure 6.3 Counts of ST discussion threads at comparable time stages, in four MOOC forums. 171
Figure 6.4 Counts of SN discussion threads at comparable time stages, in four MOOC forums. 172
Figure 8.1 Visualisation of the configurations for the ERGM 224
Figure 8.2 Cross-course comparison of the social processes driving the formation of a forum core. 230
Figure 9.1 Multi-Level Relationships in a MOOC Forum. 244
Figure 9.2 Relational model of the components of social learning in a MOOC setting 253

List of Tables

Table 1.1 Study design 12
Table 3.1 Overview of data sources and methods 74
Table 3.2 Overview of the courses 76
Table 4.1 Number of forum posts, per week, in four edX MOOCs 93
Table 4.2 Number of participants posting, per week, in four edX MOOCs 94
Table 4.3 Number of posters, according to the total number of weeks they posted in forums (consecutive or non-consecutive), in four edX MOOCs 94
Table 4.4 Number of un-replied posts, per week, in four edX MOOCs 94
Table 4.5 Number of posters, according to the total number of weeks they posted in forums, and received replies (at any point, by the end of the course) 95
Table 4.6 Number of posters, according to the total number of weeks they posted in forums, and received replies in the same thematic week of their posts 95
Table 4.7 Typology of forum users 96
Table 5.1 Number of posters in analysed courses 111
Table 5.2 Example of one row from a mongo forum file output (over 16000 rows, extracted information is bolded, student information removed) 112
Table 5.3 Temporal boundaries for longitudinal analysis of network development 114
Table 5.4 Descriptive statistics of co-occurrence networks in four edX MOOCs 121
Table 5.5 Bootstrapped coefficients for mixed effects models predicting degree and edge weight from forum participation type and time of the MOOC 132
Table 5.6 Bootstrapped means for co-occurrence frequencies in various posting groups 136
Table 6.1 Coding framework overview 154
Table 6.2 Overview of coded content 160
Table 6.3 Overview of the nature of discussion threads exchanged within the forum core. 163
Table 6.4 Overview of the counts of discussion types throughout the course. 167
Table 6.5 Central conversations from forum core posters 174
Table 7.1 Number of non-curriculum discussions analysed across the Four MOOCs 182
Table 8.1 Descriptive statistics of structural network configurations 224
Table 8.2 Descriptive statistics of the three clusters representing forum core posting activity levels in forums 227
Table 8.3 Outputs for final ERG models in four edX MOOCs 229
Table 9.1 Indicators of the social context applied in the thesis 246
Table 9.2 Overview of indicators in the analysed MOOCs 247

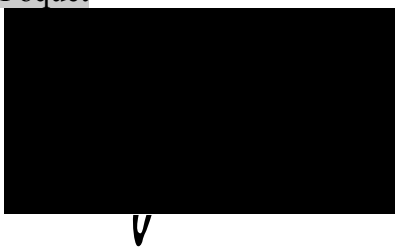
Acronyms

AI	Artificial intelligence
AIC	Akaike information criterion
ANOVA	Analysis of variance
CMC	Computer-mediated communication
CN	Content non-task
CoP	Community of practice
CT	Content-task
CTA	Community teaching assistant
DE	Distance education
EDM	Educational data mining
ENP	Electronic network of practice
ERGM	Exponential random graph modelling
FP	Functional programming
FUN	France Universite Numerique
ICT	Information and communication technologies
INF	Informational and administrative queries
LA	Learning analytics
LMS	Learning management system
MCMC	Monte Carlo Markov Chain
META	Metacognitive discussion threads
MIT	Massachusetts Institute of Technology
MLE	Maximum Likelihood Estimation
MIX	Mixed discussion threads
MOOC	Massive open online course
OCW	OpenCourseWare
OE	Online education
OER	Open educational resource
Q&A	Questions and answers
RQ	Research question
SIDE	Social identity deindividuation effects
SJSU	San Jose State University
SN	Social non-task
ST	Social-task
STEM	Science, Technology, Engineering and Mathematics
THE	Times Higher Education
TUD	Delft University of Technology
UNESCO	United Nations Educational, Scientific and Cultural Organisation

Declaration

This thesis presents work carried out by myself and does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; to the best of my knowledge it does not contain any materials previously published or written by another person except where due reference is made in the text; and all substantive contributions by others to the work presented, including jointly authored publications, are clearly acknowledged.

Oleksandra Poquet

[Signature]  March 8, 2017

Acknowledgments

Like in every significant undertaking, there were people who accompanied me during the long uphill climb and whose presence helped me continue.

I would like to acknowledge the guidance and support from my supervisor Professor Shane Dawson. I am grateful for his patience and trust in this research project, for always getting that critical feedback when it was most needed, for challenging my assumptions, and helping learn to write a little better.

I would like to acknowledge the support of my husband Pierre and his share in carrying the weight of every word.

I want to thank Dr. Pieter de Vries and Dr. Thieme Hennis for their leap of faith in 2013, for great collaboration and friendship.

I would like to acknowledge the Commonwealth Government funding I received under the Research Training Program (RTP), without which this research would not have been possible.

Certainly, there were others who inspired me, who I learnt from, and who were a part of my academic formation. I want to thank Srećko Joksimović, Nia Dowell, Vitomir Kovanović for the friendship and learning I experienced through working with them, and Professor Dragan Gašević for linking me to his team when I just have started; Dr. Monica Behrend, for her incredible kindness to all research students; Rosa Lisa Iannone, for her editing work on my text; MingKun Lou, for her down-to-earth advice; Dr. Marisa Ponti, for the kindness of strangers in connecting me to the Delft University of Technology; to the LAK Doctoral Consortium members (2015), Interlab researchers, and UniSA visiting researchers, for the inspiration, as they all have contributed to my thinking. I also want to acknowledge the research fellowship opportunity I received from the Open Online Learning at Delft University of Technology, and to thank Dr. Jan-Paul van Staalduinen in his help during the project.

Published Work

The following peer-reviewed publications, manuscripts under review, and a technical report were produced from the analyses undertaken in this thesis. This work reflects the development of my ideas about learning and social activity in MOOCs. First author papers formed the basis of individual chapters, other publications either contributed with evidence to tangential arguments, or were used to understand MOOC data, and the application of social network analysis in MOOCs. Experiences and thinking through these papers have contributed to my understanding of MOOCs and the associated literature.

Poquet, O., Dawson, S., Dowell, N. (2017). How effective is your facilitation? Group-level analytics of MOOC forums. *Proceedings of the 7th International Conference on Learning Analytics and Knowledge (LAK'17*), 208–217. http://dx.doi.org/10.1145/3027385.3027404

Pardo, A., **Poquet, O**., Martinez-Maldonaldo, R., Dawson, S. (2017). Promotion of data-driven feedback in EDM and LA. In C. Lang & G. Siemens (Eds.), *Handbook of learning analytics.* [To be published in March, 2017]

Poquet, O. & Dawson, S. (2016). Untangling MOOC learner networks. *Proceedings of 6th International Learning Analytics and Knowledge Conference (LAK'16),* 208–217. http://dx.doi.org/10.1145/2883851.2883919

Hennis, T., Topolovec, S., **Poquet, O**., de Vries, P. (2016). *Who is the learner: Profiling the engineering MOOC student.* Paper presented at the 44th Annual Conference of the European Society for Engineering Education (SEFI), Tampere, Finland. Retrieved from http://goo.gl/yl2PJB

Poquet, O. (2016). *Social Learning in DelftX MOOCs.* [Internal Research Report for Open Online Education and Research]. Delft, the Netherlands: TUD.

Skrypnyk, O., Joksimovic, S., Kovanovic, V., Gasevic, D., Dawson, S. (2015). Roles of course facilitators, learners, and technology in the flow of information of a cMOOC. *The International Review of Research in Open and Distributed Learning, 16*(3).

Poquet, O. & Dawson, S. (2015). Analysis of MOOC forum participation. In T. Reiners, B.R. von Konsky, D. Gibson, V. Chang, L. Irving, & K. Clarke (Eds.), *Globally connected, digitally enabled (Ascilite 2015*) (pp. 224–234). Perth: Australasian Society for Computers in Learning in Tertiary Education.

Dowell, N., **Skrypnyk, O**., Joksimovic, S., Graesser, A., Dawson, S., Gasevic, D., de Vries P., Hennis T., Kovanovic, V. (2015)**.** *Modelling learners' social centrality and performance through language and discourse.* Paper presented at the 8th International Conference on Educational Data Mining (EDM), Madrid, Spain. Retrieved from https://eric.ed.gov/?id=ED560532

Hennis, T. A**., Skrypnyk, O**., & de Vries, P. (2015) *Who is the learner in the DelftX engineering MOOCs.* Paper presented at the International Joint Conference on the Learner in Engineering Education (IJCLEE 2015), Orleans, France. Retrieved from http://bit.ly/2nhNixh

Joksimovic, S., Dowell N., **Skrypnyk, O.,** Kovanovic V., Gasevic, D., Dawson, S., Graesser, A. (2015)**.** How do you connect? Analysis of Social Capital Accumulation in connectivist MOOCs. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge (LAK 15),* 64–68. http://dx.doi.org/10.1145/2723576.2723604

Skrypnyk, O., Hennis, T., de Vries, P. (2015). Reconsidering retention in MOOCs: The relevance of formal assessment and pedagogy. *Proceedings of the European Stakeholders Summit on experiences and best practices in and around MOOCs, Mons, Belgium* (pp. 166–173*)*. Retrieved from http://www.emoocs2015.eu/sites/default/files/Papers.pdf

Joksimovic S., Kovanovic V., **Skrypnyk O.,** Gasevic D., Dawson S., Siemens G. (2015). The history and state of online learning. In G. Siemens, D. Gasevic, & S. Dawson (Eds.), *Preparing for the digital university: a review of the history and current state of distance, blended, and online learning*. Athabasca University. Retrieved from http://linkresearchlab.org/PreparingDigitalUniversity.pdf

Kovanovic V, Joksimovic S., Skrypnyk O., Gasevic D., Dawson S., Siemens G. (2015). The history and state of distance learning. In G. Siemens, D. Gasevic, & S. Dawson (Eds.), *Preparing for the digital university: a review of the history and current state of distance, blended, and online learning*. Athabasca University. Retrieved from http://linkresearchlab.org/PreparingDigitalUniversity.pdf

Skrypnyk O., Joksimovic S., Kovanovic V., Dawson S., Gasevic D., Siemens G. The history and state of blended learning. In G. Siemens, D. Gasevic, & S. Dawson (Eds.), *Preparing for the digital university: a review of the history and current state of distance, blended, and online learning*. Athabasca University. Retrieved from http://linkresearchlab.org/PreparingDigitalUniversity.pdf

Skrypnyk, O[1]**.**, Hennis, T., de Vries, P. (2014) *Understanding social learning behaviors of xMOOC completers.* Paper presented at the 42 Conference of the European Society for Engineering Education (SEFI), Birmingham, UK. Retrieved from http://goo.gl/yl2PJB

Manuscripts Under Review

Joksimovic, S., **Poquet, O.,** Kovanovic, V., Dowell, N., Mills, C., Gasevic, D., Dawson, S., Graesser, A., Brooks, C. How do we model learning at scale? A systematic review of literature. Manuscript under review.

Poquet, O., Kovanovic, V., Hennis, T., de Vries, P., Joksimovic, S., Gasevic, D., Dawson, S. Social presence in Massive Open Online Courses. Manuscript under review.

Kovanovic, V., Joksimovic, S., **Poquet, O.,** Hennis, T., Cukis, I., de Vries, P., Hatala, M., Dawson, S., Siemens, G., Gasevic, D. Exploring Communities of Inquiry in Massive Open Online Courses. Manuscript under review.

Joksimovic, S., Dowell, N., **Skrypnyk O.**, Kovanovic, V., Gasevic, D., Dawson, S., Graesser, A. (2016). Exploring the accumulation of social capital in cMOOC through language and discourse. Manuscript under review.

[1] Prior to 2016 I published under the last name Skrypnyk.

1 Introduction

1.1 Background

Massive Open Online Courses (MOOCs) are education offerings freely available to anybody over the Internet. From their onset, MOOCs attracted much attention from social and technological commentators for their perceived potential to disrupt the higher education (DBIS, 2013, p.19). The public and mainstream media viewed MOOCs as a disruptive force that could revolutionise higher education due to their unprecedented scale, openness and decreased costs for the learner. For example, in the report by Ernst and Young (2012) the introduction of MOOCs was predicted to rapidly result in the demise of the traditional brick-and-mortar universities. Although such hyperbole has not come to fruition, MOOCs have brought about a new way of thinking about education and the scale of education offerings delivered via the Internet.

In 2008 the first open online course on connectivism (CCK08) was offered by a group of Canadian educators. This first MOOC used an innovative pedagogy to integrate web 2.0 technologies into the constructivist practices of learners' co-creation of knowledge. Courses adopting a similar pedagogical design became known as cMOOCs or connectivist MOOCs. Although over 2000 learners participated in this first open online course, cMOOCs did not gain traction with the general public beyond those already interested in online education. However, this changed quickly in 2011 when a model of an open course different from earlier cMOOC versions was trialled by Sebastian Thrun. The Stanford professor simply extended his university course offering on Artificial Intelligence (AI MOOC) as an open offering (i.e. without fees to external learners). Such an opportunity to learn from a leading expert at no cost appealed to hundreds of thousands of individuals. This experiment demonstrated the MOOC's potential capacity to accommodate unprecedented enrolment numbers, thereby rapidly attracting media attention and entrepreneurial interest. Several high-profile universities, such as

Stanford, Massachusetts Institute of Technology (MIT) and Harvard, along with venture capitalists commenced the roll-out of various MOOC initiatives that sought to replicate the AI MOOC experience. These subsequently named xMOOCs mirrored the design of traditional university courses based on lecture broadcasting and quiz-based assessment. xMOOC provisions gave rise to claims about the potential of MOOCs to transform higher education and break down the traditional power structures held by universities. Bulfin, Pangrazio and Selwyn (2014, para.12) described the media's portrayal of MOOCs as:

> A key theme here was disruption of the established order—'the education world has been thrown into disarray' (*Sunday Times*, 1.12.2013), with MOOCs 'disrupting centuries-old models' (*Chronicle of Higher Education*, 28.10.2013). These were courses that marked a new educational direction—'path breaking' (*Chronicle of Higher Education*, 25.3.2013) and representing 'learning's new path' (*Australian,* 17.8.2013). The momentum of these changes was described in forceful terms as a 'revolution' (*New York Times*, 16.5.2012).

Contrary to predictions made by social and education commentators (Ernst and Young, 2012; Shirky, 2012), higher education has preserved its societal weight and role. As summarised by the head of a UK MOOC provider in an interview for the Times Higher Education, the claims that MOOCs "were going … to transform education … to destroy the university system … to put educators out of business" were exaggerated (Parr, 2014, para. 6). Rather, the head noted that the on-going higher educational transformation was a result of the technological advances and the Internet, with MOOCs comprising a small part of this wider change. Such a position is well-aligned with the position of the educational technology researchers who have argued that MOOCs are a legacy of distance education (Moe, 2014). Through such a perspective, open courses can be viewed as a model of online education, rather than a unique transformative phenomenon.

Due to the lack of formal accreditation processes, MOOCs are more likely to complement formal higher education models, rather than act as their replacements. MOOCs are not the full-fledged components of degree programs integrated into national higher education systems. Instead of formal credentials, MOOCs provide free or low-cost structured opportunities to learn for self-fulfilment or upskilling. Such unaccredited learning opportunities are intentional from the learner perspective, and structured through the learning objectives, course time and support. These course features position MOOCs within the domain of non-formal education (Eraut, 2000;

Livingstone, 2001; Myers, Conte, & Rubenson, 2014). Non-formal offerings are unaccredited courses where "learners opt to acquire further knowledge or skill by studying voluntarily with a teacher who assists their self-determined interests, by using an organised curriculum, as is the case in many adult education courses and workshops" (CEC, 2000, p. 2). The significance of non-formal education has been acknowledged through the political agendas that presented upskilling in a global economy as an individual responsibility of citizens (Milana, 2012). As MOOCs are characteristic of non-formal education (Czerniewicz, Deacon, Fife, Small, & Walji, 2015; Witthaus et al., 2016), this Doctoral thesis approaches them as affordable non-formal offerings that serve to help individuals fulfil their learning needs.

A strong and ongoing public demand for non-formal learning opportunities is evident from the unprecedented uptake of MOOCs by individuals worldwide. By 2017 MOOC regional providers have been established in North America, Latin America, Australasia, Europe, the Middle East and Africa reaching out to millions of learners. Reports note some 4.5 million enrolments for edX, a US-based MOOC provider (Chung & Ho, 2016), over 25 million international learners for Coursera, another US company (Zhenghao et al., 2015), some 5 million users for the UK-based Futurelearn (FutureLearn, 2016), some 2 million Spanish-speaking registrants for MiriadaX (Telecompaper, 2016), and over 5 million users for the Chinese MOOC platform XuentangX (Shah, 2016), among others. According to the MOOC providers, the majority of registrants are adults who are studying, working or seeking employment (Chung & Ho, 2016; Stich & Reeves, 2017; Zhenghao et al., 2015). The high MOOC enrolment among labour market entrants and working professionals supports the value of affordable alternate education offerings in serving a need around life-long and professional learning.

1.2 Problem Formulation

Although MOOCs have been delivered to millions of individuals worldwide, questions as to how to promote optimal student learning outcomes and enhance student experiences in MOOCs remain unaddressed. As further detailed in Chapter 2, until now MOOC delivery and research have relied on the practices and approaches inherited from formal higher education in online and face-to-face settings. Yet, this thesis argues that due to the non-formal nature of MOOCs and their characteristics of diversity, scale and

openness, the application of formal educational frameworks to MOOC contexts may be inadequate.

Theories of teaching and learning used in formal education rely on a familiar learner setting. Biggs (in Biggs & Tang, 2011; Biggs & Telfer, 1987) abstracted the factors comprising the university setting into a model of student learning in the university context (the presage-process-product model, see Figure 1.1). The presage stage describes personal and situational factors existing prior a student's entry into a learning situation. Personal factors include the student's prior knowledge, ability and preferred approaches to learning. Situational factors include a teacher's chosen design, subject area, and course structure, among others. To facilitate effective learning in formal education, ideally a teacher needs to ensure that the learner's prior knowledge and background (presage) align with learning activities (process) (Biggs & Tang, 2011). The teacher's task also includes aligning the learning process with the intended learning outcome (product).

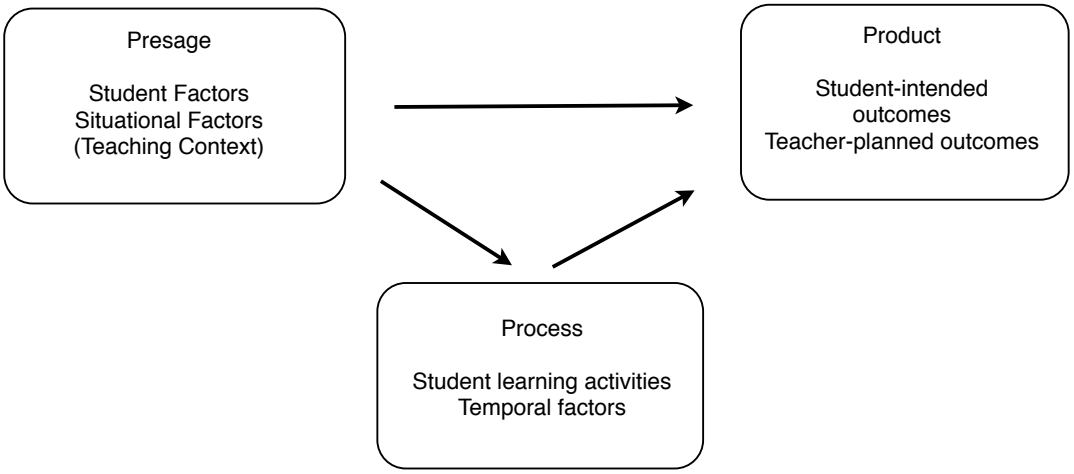


Figure 1.1 Schematic representation of the Biggs Presage-Product-Process Model for student learning in higher education (simplified from Biggs & Tang, 2011)

Massive open online educational settings, however, provide a context for learning that differs from the formal university setting. An exponentially growing body of research offers insights into MOOCs related to:

- individual learner demographics (e.g. DeBoer, Stump, Pritchard, Seaton, & Breslow, 2013; Dillahunt, Wang, & Teasley, 2014; Hansen & Reich, 2015; Macleod, Haywood, Woodgate, & Alkhatnai, 2015);
- performance (e.g. Champaign et al., 2014; Greene, Oswald, & Pomerantz, 2015; Jiang, Warschauer, Williams, O'Dowd, & Schenke, 2014; Kennedy, Coffrin, de Barba, & Corrin, 2015);
- participation (e.g. Andersen & Ponti, 2014; Anh, Butler, & Alam, 2013; C. Milligan, Littlejohn, & Margaryan, 2013; R. Kizilcec, Schneider, Cohen, & McFarland, 2014), interaction (e.g. Eynon, 2014; Shillair & Wash, 2015; Sinha, 2014; D. Yang, Wen, Kumar, Xing, & Rosé, 2014);
- perceptions and experiences (e.g. Cao, 2014; Kop, Fournier, & Mak, 2011; Veletsianos, 2013); and
- and motivation (e.g. Kizilcec & Schneider, 2015; Schneider & Kizilcec, 2014; Zheng, Rosson, Shih, & Carroll, 2015).

In particular, educational research illuminated features that make MOOCs a distinct educational setting, including the (1) diversity of learner backgrounds, (2) motivations, and (3) variations in learner participation patterns. These features differentiate MOOCs from formal courses and challenge the established process of designing for learning.

The first feature, the massive scale of diversity, is one of the key characteristics of the MOOC settings. The student population in MOOCs varies in learner background, age, schooling, and origin (Rachel Baker, Evans, Greenberg, & Dee, 2014; Breslow & Pritchard, 2013; Chung & Ho, 2016; DeBoer, Stump, Seaton, & Breslow, 2013; Glass, Shiokawa-Baklan, & Saltarelli, 2016). A large proportion of learners are college-educated males in developed countries, or working adults with a college degree (Hansen & Reich, 2015; Kolowich, 2013). However, researchers are yet to define a typical MOOC student due to the overall diversity of registrants (Glass et al., 2016). Schmid, Manturuk, Simpkins, Goldwasser, and Whitfield (2015) in particular delineated three groups of learners beyond the educated college students and working professionals: young people under the age of 18, older adults over 65, and people who reported not having access to higher education opportunities. Such a diversity in backgrounds also applies to diversity in learner prior knowledge and professional exposure: some registrants have extensive professional experience in a given MOOC subject, while others gain first-time exposure to new topics or disciplines through the same MOOC

(DeBoer, Ho, Stump, & Breslow, 2014; DeBoer et al., 2013; DeBoer et al., 2013; Hood, Littlejohn, & Milligan, 2015). Combined with the sheer volume of registrations, such diversity of presage factors challenges an alignment between the design of a learning process and intended learning outcomes.

The second feature is that MOOC learner motivations are also more varied than in formal settings. That is, in contrast to formal settings where receiving credit for a course is the primary outcome, not all individuals who join a MOOC are there to receive credentialing (C. Milligan, Littlejohn, & Hood, 2016; S. White, Davis, Dickens, Leon Urrutia, & Sanchez Vera, 2014). For instance, a study by Reich (2014) found that in a selected group of MOOCs from the US-based edX platform, only 9%–35% of learners actually intended to complete the course they signed up for. The diversity of learner intentions for taking a MOOC has been evidenced through learner surveys (DeBoer et al., 2013; Reich, 2014; Schneider & Kizilcec, 2014) and learner interviews (Zheng, Rosson, Shih, & Carroll, 2015). Zheng et al. (2015) identified four dimensions underpinning student motivations for enrolment: fulfilling current needs, preparing for the future, satisfying curiosity and connecting with other people. In short, the intentions of MOOC registrants are varied: individuals may enrol to sample the content, gain experience in learning at a distance, forge new professional connections, or even improve their English skills. As a result of these divergent motivations, individuals may enrol without a strong commitment to either engage in or complete a MOOC. Such a diversity in learner intentions, accompanied by divergent commitment to course participation, challenges an alignment of learning outcomes (product stage) with learning activities in the process of course design (see Figure 1.1).

The third feature typical for MOOC settings relates to the open participation. The open nature of the MOOC lies in stark contrast to formal education practices. In MOOC settings, individuals are allowed to join the course and disengage at their convenience. This permits learners with different motivations and backgrounds to join a course at different points in its progression, and engage with the materials and other MOOC-takers. Simply put, the diversity of learner backgrounds and intentions in an open educational setting results in diverse learner participation patterns (DeBoer et al., 2014; Guo & Reinecke, 2014; Kizilcec, Piech, & Schneider, 2013). Such flexibility of pathways to enter and exit a learning process changes the pace and synchronicity of the learning process in a MOOC (Mullaney, 2014). Besides individual choices around

patterns of engagement with course materials, forum participation patterns at the group-level also vary from formal settings. In MOOCs only a small fraction of learners participate in group-related activities in course forums (A. Anderson, Huttenlocher, Kleinberg, & Leskovec, 2014). Further, those who participate, drop in and out of group activities intermittently through a course's progression (Rosé et al., 2014; D. Yang, Sinha, Adamson, & Rosé, 2013). As a result, group dynamics and patterns of interactions in a MOOC differ from formal settings that have continuous interpersonal participation and obligatory assessment.

In summary, the distinct features of MOOC settings, including unprecedented scale of enrolments, high levels of learner diversity and open participation, challenge the established process of designing for learning. As such, the direct application of Biggs' presage, process and product model may not be ideal for MOOC settings. In MOOC settings, the design for learning needs to account for the diversity of learners and fluid participation patterns. In a MOOC, the alignment of the process of learning with student factors is distinct from formal course offerings. Given the multiple and diverse learner interests and prior knowledge of the subject at-hand (Hood et al., 2015; Schneider & Kizilcec, 2014), the suitability of the designed course's materials and activities for learners will vary from person to person. Furthermore, an alignment of learning outcomes with learning activities necessitates an accommodation for the variety of learner goals in an open course. In MOOCs, learner intentions are manifold, and the intended learning outcomes are more than likely not shared by all participants (Reich, 2014). Even when an instructor defines learning outcomes, the alignment of the outcomes with the learning process needs to address the fluid and changing context of MOOC interpersonal interactions (D. Yang et al., 2013). In MOOCs, it cannot be assumed that all learners will start and finish together or choose to interact with peers, as is often be the case in formal online courses (DeBoer et al., 2014; Mullaney, 2014). Since the alignment of different parts of teaching and design in MOOCs is more complex than formal online settings, existing theories and models of learning may not be fully suited for the MOOC environment. To re-apply existing practices and approaches to online learning to MOOCs, the differences between MOOC contexts and formal contexts need to be better understood.

Understanding the nature and quality of the emergent social context in an open educational environment is of essence for effective teaching online. Facilitation of

learner-to-learner and learner-to-teacher interactions is inherent in contemporary teaching practices (Goodyear, Salmon, Spector, Steeples, & Tickner, 2001; Woo & Reeves, 2008). Thus far, existing evidence-based practices of how to facilitate interpersonal interactions have been derived from research traditions analysing micro-level interactions in bounded contexts (Bernard et al., 2009; Bernard, Abrami, Wade, Borokhovski, & Lou, 2004; Borokhovski, Tamim, Bernard, Abrami, & Sokolovskaya, 2012). In formal courses, learners and teachers continuously interact through forums, obligatory assignments and tutorials, and interpersonal interactions aggregate into patterns of similar quality. That is, with the progression of a course, these interpersonal interactions are expected to develop a relational quality, i.e. an underpinning emotional connection (Campbell, 2006; Cleveland-Innes & Campbell, 2012; Rourke, Anderson, Garrison, & Archer, 2007; Rovai, 2002). If interpersonal interactions are sustained, the relational processes underpinning them may enable social support, deeper learning, and stronger motivation (Martin & Dowson, 2009). Such a development of interpersonal relationships in the classroom is implicit within contemporary socio-constructivist and socio-cultural theories and models of learning (Palincsar, 1998; Wubbels, den Brok, van Tartwijk, & Levy, 2012). Given that interpersonal patterns in MOOCs differ as a result of their fluid and open nature, the quality of the emergent social context in MOOCs is also likely to be different to formal models of education. If so, contemporary theories of learning and their models may not be well suited to facilitate an effective learning process.

Despite these evident differences between MOOCs and university courses, past research has examined open online courses through the prism of formal educational approaches. As Chapter 2 demonstrates, much of the MOOC research to date has approached the success of learning as an outcome of course completion, certification and retention (see Section 2.2.1). The process of learning has therefore often been aligned with these performance metrics. For example, learner forum participation has been studied in relation to completion and final learner grades (see Section 2.2.2). Moreover, there is a dearth of research investigating the emergent social context of a MOOC and its quality. Empirical research has demonstrated that the patterns of interpersonal interactions in MOOCs differ due to the open participation (see Section 2.2.4). These fluid boundaries of groups of learners in MOOCs call into question the degree to which trust, safety and similar socio-psychological constructs, associated with closely-knit small groups, occur

in open online learning environments (see Section 2.3 and Section 2.4). Therefore, this study examines the patterns of interpersonal interactions and social processes underpinning them in order to understand the emergent social context in a MOOC setting, as well as to establish the applicability of socio-constructivist and socio-cultural learning theories to MOOC settings.

1.3 Study's Aim, Objectives and Research Question

The aim of the thesis is to examine the nature and development of the social context in MOOCs. The social context in MOOCs emerges from the socio-emotional processes underpinning the interpersonal interactions within the open and fluid participation patterns of MOOC learner activity. Given the lack of research into the social context in a MOOC setting, the study's objectives are to:

- understand the nature of socio-emotional processes underpinning interpersonal interactions in MOOCs;
- compare the context of a formal course and an open course in online education; and
- establish lead indicators of the emergent social context.

To address these objectives, the following overarching research question (RQ) was formulated to frame the thesis:

RQ: What are the properties of social context development in MOOC settings, and what indicators capture it?

The social context refers to the emergent patterns of interpersonal interactions between learners and instructors, and their relational quality, i.e. emotional attachment. The scope of the analysis was limited to platform-based MOOCs, such as those delivered through edX and Coursera. In particular, the MOOCs analysed in this study were delivered via the edX platform. Such a choice is representative of the majority of MOOC offerings, as all MOOC providers are essentially platform-based. EDX, Coursera, Udacity, Futurelearn, among others, run MOOCs from their own platforms. In addition, the edX platform designed an open source MOOC platform – the open edX. Open edX has been adopted by some national MOOC initiatives, such as the France Universite Numerique (FUN). Universities or individual lecturers may also offer standalone MOOCs without a mediating provider. In such instances, they either use their own home-grown platforms or deliver through an established LMS, such as

Moodle or Canvas. At times, instructors may encourage the use of social media in their MOOCs, however the main body of the course remains located within a chosen platform. Essentially, the online platform is the central and driving technology that is used to facilitate the provision of open scaled online education.

The first analysis chapter (Chapter 4) explored data from forum logs. Exploratory data analysis illuminated the patterns of forum participation and use. The frequency of posting and viewing activity was used to frame forum activity for further analyses. The findings demonstrate that forum users can be delimited into groups reflecting the continuity of their posting and viewing activity. Three groups of forum users were considered relevant to the formation of the social context: occasional posters, regular posters and regular posters who receive timely replies (i.e. forum core). The remainder of the thesis is structured around successive research studies investigating these three groups of posters. The investigation progresses from quantitative data-driven analyses to the analyses of smaller datasets of learner interactions viewed through qualitative lenses. The following research questions framed each of the empirical chapters; Chapters 5–8.

Study 1 (Chapter 5). What types of forum participation are associated with opportunities for relationship formation?

Study 2 (Chapter 6). What is the role of on-task and off-task communication in the formation of a forum core?

Study 3 (Chapter 7). What were the forum core posters talking about within social-task, social non-task and mixed discussion threads?

Study 4 (Chapter 8). To what extent are the participatory patterns observed in a forum core's networks typical of identity-based communities?

The findings from Studies 1–3 (Chapters 5–7) provide for a novel conceptualisation of the social context of MOOCs. Throughout these chapters, participation activity is examined through log data, forum posting structure, and content of forum discussions. Based on the insights from these sequential analyses, the forum core is theorised as an emergent identity-based community. Study 4 (Chapter 8) uses findings from the previous chapters to formulate hypothesis about the network structure of such a community. The hypothesis is empirically tested via statistical modelling of the

network. A set of quantitative indicators reflecting the development of the identity-based communities is suggested.

1.4 Study Design

Research orientation. The thesis is situated in the domain of learning analytics. Learning analytics is a scholarly engagement into the "measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs" (Siemens, 2013a, p. 1382). Learning analytics research requires a practical orientation towards the adopted research questions. To measure, collect, analyse and report learning-related data, learning analytics takes on an interdisciplinary approach to analysis (Haythornthwaite, de Laat, & Dawson, 2013). In so doing, learning analytics draws on a diversity of methodologies (Dawson, Gašević, Siemens, & Joksimović, 2014). The practical orientation of the field's research calls for combining interpretative analyses, as well as statistical analyses, text analysis, and social network analysis, among others. Similarly, the practical orientation of the research questions in this thesis necessitates the use of both qualitative and quantitative approaches. Therefore, a pragmatist research orientation (Biesta & Burbules, 2003) and mixed method design were adopted (Johnson, Onwuegbuzie, & Turner, 2007; Tashakkori & Teddie, 1998).

Data. The study examines forum activity evolving from four MOOC courses in engineering, computer science and data analysis delivered via the edX platform by the Delft University of Technology in 2013–2014. The courses are representative of the conventional platform-based MOOCs offered through edX and had a high volume of content, offered in the forms of video lecture series, weekly released quizzes and assessments, and minimal targeted activities to promote social learning activities in the forums. The courses are comparable to standard STEM (Science, Technology, Engineering and Mathematics) courses with minimal design for social interaction. These particular courses were chosen due to their average scale of enrolment and average completion rates (Jordan, 2014), as well as vibrant forum activity. Chapter 3 provides further details regarding the individual course descriptions. An overview of the data used in the thesis is summarised in Table 1.1. Research ethics approval was obtained through the University of South Australia (see Appendix A).

Research Design. The research inquiry was conducted sequentially, as typical in explanatory or exploratory studies (Gelo, Braakman, & Benetka, 2008). The structure of the thesis follows a top-down approach whereby earlier chapters used data analyses of larger datasets, and later chapters shifted towards more granular insights in smaller datasets. First, through exploratory data analysis of forum log access data (Chapter 4), sub-groups of learners were delineated for further examination. Chapter 5 compared different sub-groups of posters in order to identify the sub-group where individuals could potentially form relationships with one another. Chapter 6 analysed the content of the discussion threads exchanged by this sub-group of posters so as to understand the nature of their exchanges. Chapter 7 adopted an interpretative lens to the selected discussion threads associated with socio-emotional processes. Throughout Chapters 4-7, each subsequent study focused on a smaller dataset, defined by the prior analyses. Through findings from Chapters 4-7, a hypothesis about the forum core's network structure was formulated and tested (Chapter 8).

Table 1.1 Study design

Study	Research Questions	Method	Data Type
Exploratory data analysis	The aim of the analysis was to gain an understanding of the dataset, and frame further analyses.	Count of user logs	Forum access logs
Study 1	What types of forum participation are associated with opportunities for relationship formation?	SNA[1]/Permutated ANOVA[2] and mixed models	Post logs and timestamps
Study 2	What is the role of on-task and off-task communication in the formation of a forum core?	CA[3] SNA, bipartite networks	Text of posts, timestamps
Study 3	What were the forum core posters talking about within social-task, social non-task and mixed discussion threads?	Nvivo coding	Text of posts
Study 4	To what extent are the patterns observed in a forum core's networks typical of identity-based communities?	ERGM[4]	Post logs and timestamps

Note: 1.SNA - Social network analysis; 2. ANOVA - Analysis of variance, 3. CA - Content analysis, 4. ERGM - Exponential random graph modelling

Methods. The study employed a mixed research design, and integrated quantitative and qualitative approaches to address the research questions (Johnson et al., 2007; Tashakkori & Teddie, 1998). The integration of methods was guided by the overall assumption that "the use of qualitative and quantitative approaches in combination provides a better understanding of research projects than either approach alone" (Creswell & Clark, 2011, p. 282). Table 1.1 also summarises the methods used in the thesis. Analyses in Chapter 4 (Exploratory Data Analysis) and Chapter 5 (Study 1) used descriptive data exploration and descriptive social network analysis. These quantitative methods narrowed down the inquiry and enabled the identification of a sub-group of posters to be studied in more detail. Chapter 6 (Study 2) and Chapter 7 (Study 3) used content analysis and thematic analysis respectively in order to examine the text shared by this sub-group. Qualitative analysis of the content shared between the posters illuminated the quality of relatedness underpinning the interactions, i.e. the emotional connection. The adopted methods gradually transition from larger amounts of data and quantitative approaches to smaller sub-sets of data and qualitative analyses. Chapter 8 (Study 4) conducted quantitative research so as to establish indicators of the social context, based on the hypothesis formulated from the qualitative insights.

1.5 Structure of the Thesis

Chapter 1 introduces the topic of the thesis and the research problem. The argument highlights the features that distinguish MOOCs from formal online education. The chapter further argues that the two contexts need to be comparable for the transfer of theories and models from formal online education to the non-formal context of MOOCs. Examining the social context in MOOCs would enable such a transfer. The chapter presents the aim and research questions, research design and the structure of the thesis.

Chapter 2 demonstrates the gap in the literature. The literature review argues that thus far, the formal educational perspective has dominated MOOC research. The chapter emphasises that insights into the social context development in MOOCs are essential for decisions about the transfer of existing theories and models to an open context. The chapter further critiques empirical studies of the social context in open online settings by highlighting the limitations in how the previous work has examined the social aspects of MOOC forums. In conclusion, three research propositions that address the

development of the social context are outlined. These research propositions form the basis for the thesis' empirical studies.

Chapter 3 positions the thesis within the learning analytics research. The chapter explains the challenges associated with designing research that measures learning-related constructs. To address methodological considerations, the chapter highlights the philosophical considerations underpinning the thesis. The research design is also presented, and the four MOOC courses used for analyses are then described in relation to their enrolment numbers, pedagogy, and course design. Further individual chapters include the details of the methods used in their analyses.

Chapter 4 presents exploratory analyses of forum log access data. The chapter explains that forum use can be framed through the regularity of posting and viewing. Regularity of posting and viewing reveals underlying learner motivation towards social learning. The chapter suggests a typology of forum use, and describes forum access over time by different types of forum users. Subsequent thesis chapters analyse three types of posters highlighted within the typology.

Chapter 5 compares the frequency of dyadic interpersonal interactions across forum posters. The frequency of interpersonal co-occurrences are approached as opportunities for relationship formation. The group with the highest interpersonal co-occurrence is referred to as the forum core. The forum core's posters experience continuity of interpersonal interactions and consequently, their interactions may evolve into relational processes.

The following chapters (Chapters 6 and 7) examine the discussion threads driving the formation of the forum core's network. Chapter 6 analyses the types of discussions underlying forum core co-occurrences. The chapter identifies the role of different types of discussions in the formation of the forum core. As different discussion threads are exchanged by the forum core, some of them are found to contain traces of social processes. Chapter 7 interprets these off-task discussion threads using the community of practice theory. Chapter 7 concludes with a conceptualisation of the forum core. This conceptualisation is used to form a hypothesis examined in Chapter 8.

Chapter 8 reports on a confirmatory study testing a hypothesis about the social processes in MOOCs. The findings confirm the forum core comprises of the same

baseline processes as observed in the electronic networks of practice. The chapter then presents indicators that capture the social context of the forum core.

Chapter 9 synthesises the results of the empirical studies so as to discuss the overarching research question. The chapter elaborates on the implications of the nature of the social context in MOOCs, and proposes future research agenda to advance its understanding.

1.6 Summary

This Doctoral thesis investigates the patterns of interpersonal interactions emergent in MOOCs. Interpersonal interactions are analysed so as to gain insights into the unfolding of relational processes associated with group formation and community development. Understanding the discrepancies between the emergent social context in a MOOC setting and in a formal online setting, can enable better course design, aid in further interpretations of learner activity and inform teacher course evaluation. The next chapter reviews the literature narrating MOOCs' emergence, critiquing the analyses of MOOC data that is undertaken through formal educational approaches, also demonstrating the gap in research of the social context in MOOCs.

2 Literature review

This chapter highlights the dearth of research that specifically addresses the social context in massive open online courses (MOOCs). The chapter demonstrates that there is a lack of understanding around how social context develops in MOOC settings. However, simple examination of a social context in a MOOC cannot be undertaken through the lens of conventional education theory and practices. In support of this argument, the literature review first narrates the historical and socio-political background of the MOOCs, their emergence and their role in the contemporary education landscape (Section 2.1). Section 2.2 critiques the mainstream analyses of MOOCs that emphasise course retention and completion outcomes. Section 2.3 maintains that open participation and scale in MOOCs generates patterns of interpersonal interactions that differ significantly from what is expected in a formal university course. The section highlights the literature gap around the quality of the social context emerging from these patterns of interactions. In MOOCs, interpersonal interactions may also vary from those in formal education courses in the quality of relatedness, namely, "greater connection on personal and emotional levels" (Martin & Dowson, 2009, p. 327). The section demonstrates the need for research investigating the social context in MOOCs. Such an understanding is essential to the use of formal education theories and models, which imply a certain emergent quality of relational processes between the students for deeper cognitive engagement, motivation and satisfaction. The final section of this chapter (Section 2.4) reviews MOOC research to illustrate scarce evidence around the relational quality of interactional patterns in these courses. The chapter concludes with three research propositions that lead to the empirical studies conducted in Chapters 5-8.

2.1 The Evolution of MOOCs

Section 2.1 explains the role of MOOCs within the current education landscape to position the research focus undertaken in this Doctoral thesis. To situate the development of MOOCs, their emergence is described within the context of expanding distance and online education. The role of the socio-economic forces influencing higher education is also acknowledged. The section presents the controversies surrounding

MOOCs and the scale of institutional and individual adoption, and discusses the value of learning opportunities offered to individuals at little to no cost.

2.1.1 Drivers for Online Education Adoption

Chapter 1 presented MOOCs as a distance education (DE) provision delivered online at little or no cost, to diverse adult learners. Implementation of the first MOOCs was facilitated by the historical and socio-economic context surrounding the development of distance and online education. There were a number of factors relevant to this context, including the evolution of technological affordances, the increase in the adult learner population and its diversity, the expansion of online education into higher education, the influence of global political agendas, and the changing nature of the university. This sub-section briefly presents some of the factors involved in broader social, economic and technical context associated with the emergence of MOOCs.

The evolution of distance education (DE) and online education (OE) took place alongside the development and adoption of new information and communication technologies (ICTs) (Bates, 1988; Garrison, 1985; Nipper, 1989). The development of technology was reflected through the type of communication it afforded, the type of information it could communicate, and the speed of communication (Lewis, Snow, Farris, & Levin, 1999). These changing technological affordances have been closely related to the types of DE practices on offer (Anderson & Dron, 2012). Research has conceptualised the shifts in technological affordances and associated DE practices through the different generations of distance education (Garrison, 1985). For instance, the first generation, namely, correspondence education, which was available from the mid-nineteenth century, used mail as the predominant mode of education delivery (Mood, 1995). Within the second generation of DE, correspondence study was complemented by mass media (television and radio), and within the third generation, synchronous audio- and video-conferencing were used. Since then, computer conferencing has further enriched DE, whereas the latest generations of DE has been defined by the use of the Internet, Web 2.0 and cloud technologies (Taylor, 2001). Throughout its evolution, emerging technology embedded in DE has provided for richer modes of conveying text, voice, image and video. Most importantly, each subsequent technological advance has made DE faster and increasingly enabled two-way as well as many-to-many modes of communication.

Distance education has always leveraged these evolving technological affordances to cater to students unable to attend on-campus courses (Keegan, 1996). These non-traditional learners (i.e. adults whose demographic profile differed from 18–24 year-old campus students) composed the target group of students for DE offerings. In the days of correspondence study, the non-traditional students were considered those living in remote areas as well as women unable or not permitted to attend higher education (Diehl, 2011). However, particularly in the past 50 years, the student demographics have grown in both size and diversity. In the period between 1970 and 1991, adult credit enrolments in the US rose by 171% (Kasworm, Sandmann, & Sissel, 2000). As well as the growing numbers of those enrolling in adult education, Kasworm and colleagues (2000) reported an increased diversity of adult learners. They observed "a 248.8 percent increase of adult students 35 years of age or older" and "a 326.8 percent in women adult students" (p. 450). Alongside the demographic changes and the democratisation of the workforce, there have been rising numbers of adult learners who are working full- and part-time. The increase in adult learner populations and the diversification of the adult learner's profile has also been observed in other OECD countries (Eurostat, 2016; McIntyre, 2005). As a result, the target student population for DE has broadened from so-called 'backdoor' learners to working adults – the fastest growing student population today (Chong, 2016). Notwithstanding all this growth, DE remained "a relatively minor, often marginalised activity conducted and promoted by a small group of educators" (Thompson, 2007, p. 4). However, this perception of DE as a sideline practice of education quickly changed with the advancement of the Internet.

The practice of DE has increasingly moved from the periphery of higher education to become an integral part of university offerings over the past 20 years. Such expansion was propelled by the rapid advancement and adoption of the Internet in the early-2000s. As a result, online education (i.e. DE delivered through the Internet), afforded more flexible learning opportunities for both on-campus and off-campus students. According to Allen and Seaman (2010), the numbers of students enrolled in online education have been increasing exponentially, compared to the growth rates of the higher education student body. This trend has shown no sign of slowing. As a result, online education has become an integral part of the education environment, especially in higher education and training (Anderson, 2011). For instance, in 2002, less than half of all US higher education institutions reported that online education was relevant to their long-

term strategy, whereas by 2015 this rose to 60%–70% (Allen & Seaman, 2003, 2004, 2010, 2013, 2014; Allen, Seaman, & Garrett, 2007; Allen, Seaman, Poulin, & Straut, 2016).

In parallel with the adoption of online education, the modern university was being transformed into an enterprise reliant on student fees and public reputation (Calhoun, 2006). This transformation prompted by the knowledge economy, an ideological order (Robertson & Keeling, 2008) based on the production, distribution and use of knowledge and information (OECD, 1996, p. 7). A repercussion of the knowledge economy was a reduction in, and destabilisation of, public funding (Robertson et al., 2012). Universities that were previously and predominantly supported by national governments through the allocation of public funds were now required to engage in additional revenue-making activities to maintain status quo (Robertson, Bonal, & Dale, 2002). Alongside the commercialisation of higher education, the public good facets of university education had shifted to becoming (financially) an individual responsibility (Connell, 2013; Marginson, 2012; Moutsios, 2011; Robertson et al., 2002, 2012) . Simply put, students had to pay higher fees to receive higher education, and universities were more reliant on such income generation. The profit-making activities of universities included marketing credentials as valuable to paying students, producing 'marketable' research, and cost-effectively rolling out education services to new markets. These included buying and selling programs through ICT, attracting more students, and maintaining institutional reputations (Connell, 2013).

In summary, prior to the emergence of MOOCs, there is a long-lasting history of integrating technology into DE in order to promote and facilitate a successful learning experiences. The emergence of MOOCs can be considered as extension of such technology-mediated offerings. However, amidst the growth of DE there has been a significant demographic shift in the learner population that is becoming increasingly diversified. Furthermore, the political contexts that have evolved over the past decades have created a need for the university to pursue student recruitment and institutional reputation. These factors are later noted among the various motivational factors for institutions to adopt and engage in the development of MOOCs (Hollands & Tirthali, 2014).

2.1.2 The Emergence of the MOOCs

As noted in Section 2.1.1, growth in the adoption of DE offerings in higher education have been driven in part by the financial pressure to increase the number of student enrolments. The political and demographic context facilitated the adoption of alternate education provisions by campus-based universities and community colleges, in parallel to their on-campus face-to-face teaching. By offering part of their accredited courses online, campus-based universities have influenced the volume of online education and increased its popularity.

In early 2000, through a number of initiatives, elite universities began experimenting with the use of online services to provide open content to the public as a means to generating revenue or enhancing institutional reputation. For example, in 2001, MIT began offering some of its teaching materials through its newly created OpenCourseWare (OCW) initiative to enable learners to use the materials without charge yet without receiving formal accreditation. By making OCW available to the general public, the institution strengthened its reputation without losing the exclusivity associated with its brand (Walsh, 2011). The public release of course content was consistent with the open educational resource movement (OER) that in part gave impetus to later MOOC developments.

The term "open educational resources" was first adopted in 2002 at the Forum of the United Nations Educational, Scientific and Cultural Organization (UNESCO). The forum highlighted the potential of OCW to widen access to education in developing countries (Daniel, 2012). The meaning of open evolved from the idea of the "learning object" (Caswell, Henson, Jensen, & Wiley, 2008), a digital material that could be reused for different pedagogical purposes. The idea of a learning object was coupled with the first open licence, introduced by Wiley in 1998 (Caswell et al., 2008). Wiley described it as "an attempt to appropriately adapt the logic of 'open source' software to the non-software world of cultural and scientific artefacts like music, literature, and images" (2009, para. 2). Wiley distinguished between free and open (i.e. at gratis – no cost), as well as libre – allowing anybody to use, change or re-use learning materials. Successful early examples of open education resources included iTunes, a platform where learners could download lectures from a variety of institutions at no cost; or OCW by MIT – the first initiative of openly providing course syllabi, homework assignments and lecture notes from most courses taught at MIT (Walsh, 2011). As

noted by Caswell et al. (2008), before MIT piloted 50 courses for its OCW project, Wikipedia had already been in operation for a year, the Internet archive for seven years, and Project Gutenberg for more than 20 years. Arguably, these projects formed the foundation of the OER movement and were instrumental to the innovators who initiated the first open online courses.

Inspired by the values associated with openness and the potential of web-based technologies to aid teaching practices, Stephen Downes and George Siemens ran the first open online course in 2008 (Siemens, 2008). This course, considered to be the first MOOC offering, was established from an expanded version of a university-run course called Connectivism and Connective Knowledge (CCK08). It was provided at no cost to participants, who may or may not have been students already enrolled at the university. The course CCK08 did not present defined course content through a centralised learning management system (LMS), as is typically associated with conventional online teaching practices. Instead, its instructors encouraged participants to explore their interests through discussions, using a diversity of social-based technologies (e.g. blogs, Twitter) to create digital products for the course, in distributed online spaces. This course attracted more than 2000 participants – an enrolment number unheard of at the time. Most likely because of its astonishing impact, courses of a similar pedagogical design adopted the title connectivist massive open online course, also known as cMOOC. These represented an opportunity for online educators to play with the affordances of emerging web-based interactive technologies.

The practice of DE has always had a close relationship with technology (Bates, 2005; Kozma, 1991; Nipper, 1989). Kozma maintained that technologies "possess particular characteristics that make them both more and less suitable for the accomplishment of certain kinds of learning tasks" (1994, para. 6). The introduction of Web 2.0 technologies has further expanded the number of possibilities on the spectrum of educational practices. Previous generations of technology (i.e. information technologies, and communication and interactive technologies) supported one-way and one-to-one delivery and access to information, and mediated two-way or one-to-many user interactions. In 2008, ICTs were combined with social software technologies. Such integrations began to expose students to the types of learning activities that were previously considered possible only in face-to-face environments. For instance, students could engage in group-based activities enabled by synchronous communication

in video classrooms, through online conferencing, and text or audio-based chats. Web 2.0 technologies also enabled students to partner with peers to work on digital artefacts without the need for physical co-location. Importantly, these emerging technologies differed in the locus of control and the kinds of interactions they supported. As articulated by Siemens et al. (2015), the change of technology has allowed for more flexibility around the course structure, student ownership of data and content, institutional integration of technology, and decentralised teaching and learning approaches.

Building on the affordances of technology, online education practitioners and researchers created cMOOCs. Siemens (2005) explained that learning came to be defined as actionable knowledge located outside and inside individuals: "[being] focused on connecting specialised information sets, and the connections that enable us to learn more are more important than our current state of knowing" (sec. 6). Hence, connectivist pedagogy tapped into the development of the supporting structure through the networks of learners distributed throughout the web. The network was to be facilitated by the connecting learners, teachers, ideas, resources and digital objects throughout a course's activities. McAuley, Stewart, Siemens and Cormier (2010) explained that connecting activities in cMOOCs were driven by "the engagement of learners who self-organize their participation according to learning goals, prior knowledge and skills, and common interests" (p. 10). Such a design allowed for learning autonomy as well as the creation of structure through dialogue. These have been the keystone elements of DE pedagogy (Moore, 1997). These elements, essential for an effective learning experience at distance, made the format popular among online education professionals; however, cMOOCs did not receive much attention beyond the online education domain.

Public attention to MOOCs was drawn by a distinctly different open online course. In 2011 a Stanford Professor, Sebastian Thrun, opened a university course on artificial intelligence (AI) to an unlimited number of online participants, globally, at no cost. Thrun's course design reflected conventional teaching approaches, such as the provision of highly structured content, set instructional videos and assignments, all of which could be accessed via a centralised online platform. The pedagogical design of the AI MOOC replicated early online education pedagogy that focused on teacher- and institution-controlled information transmission, and yet, the AI course attracted around 160,000

students from all over the world (Pappano, 2012). The vast number of students enrolled in the course was unprecedented and was considered truly massive. The sheer number of student enrolments in Thrun's AI course helped popularise MOOCs and attracted much media attention. The AI MOOC became the prototype of what was later referred to as xMOOCs, pedagogically replicating the traditional university teaching format.

The differences between the two archetypal MOOCs reflect two distinct approaches to scale and, in part, two different online teaching philosophies. The focus of cMOOCs was to enable as many peer interactions with as many learners as possible. Stewart (2013) explained that in cMOOCs, the notion of scale is implied by "the size and capacity of the course network that remixes and generates new knowledge" (sec. 3). In contrast, xMOOCs were based on the idea of economies of scale. According to Stewart (2013), the notion of scale, which focused on class size in relation to investment, was based on Spencer's (1974) perspective on economics. Simply put, scaling the content provided an opportunity to reach more learners for less cost. However, this also contributed to the implementation of a regressive pedagogy. Teaching at scale predominantly involved a transmission-based approach in lieu of more contemporary participatory models. Hence, tensions between scaling content and scaling interactions underpinned public debates around the MOOCs' emergence.

2.1.3 Public Debate

The appearance of the Stanford AI MOOC and subsequent MOOCs delivered by Coursera, EdX and Udacity stirred public attention around the potential for MOOCs to disrupt the hegemonic model of higher education (Bulfin, Pangrazio, & Selwyn, 2014; Kovanović, Joksimović, Gašević, Siemens, & Hatala, 2014; Selwyn & Bulfin, 2014; Selwyn, Bulfin, & Pangrazio, 2015). Between 1 January 2011 and 31 December 2013, ProQuest databases included some 457 MOOC-related news stories from popular media (e.g. *New York Times, Washington Post, Australian and Guardian*) as well as sources of education news (e.g. *Higher Education Chronicle* and *Times Higher Education*) (Selwyn & Bulfin, 2014). News items were polarised and contained both critique and enthusiasm for MOOC offerings (DBIS, 2013). The proponents heralded them as pioneering educational models with the potential to disrupt higher education. Clayton Christensen (2003), author of the term "disruptive innovation", referred to MOOCs as a "groundbreaking innovation" (Horn & Christensen, 2013). Other MOOC proponents,

such as Shirky, claimed that MOOCs could eventually lead to traditional brick-and-mortar universities becoming outdated, and the student debt and high drop-out rates would become obsolete (Shirky, 2012).

Popular media and news outlets positioned MOOCs within a narrative linked to the massification and commercialisation of higher education. Selwyn, Bulfin, and Pangrazio (2015) argued that the prevalent themes of media discourse around MOOCs represented the marketisation of higher education. They described MOOCs in terms such as "disruptive potential", "creative destruction", "disrupting centuries old models", "path breaking", "learning's new path", "revolution", and "game changing" (p. 182). Selwyn and colleagues (2015) noted that to convey the potential for disruption to higher education, the media discourse included the metaphors "tsunami" and "avalanche" along with a narrative of lower education costs to the learner, the prestigious reputations of the MOOC-delivering universities, and MOOCs' size and scale. Kovanović, Joksimović, Gašević, Siemens, and Hatala (2015) identified similar themes in a large-scale automated analysis of news articles. The authors analysed 3958 news items with MOOC-related keywords. They identified business and financial news media as the primary sources for addressing economic aspects of MOOC scale and explained the dominant focus on commercialisation aspects as having developed from the large investments made in their start-ups.

In contrast to the media outlets, online education experts remained critical of the public narrative around MOOC provisions. Selwyn and colleagues (2015) observed that public debates were not engaging with the issues of concern to either technology or education; that is, the public ignored issues that the field of distance and online education had been tackling for decades: quality of instruction, online learning, design for learning and student experience. Online education experts argued that to reject the historical context in which MOOCs emerged would be to invalidate prior education research, discard prior initiatives and ignore valuable lessons (Baggaley, 2013, 2014b, 2014a; Bates, 2014; Moe, 2014, 2016). For example, pedagogical decisions in xMOOCs were not as revolutionary as first claimed. Education experts pointed out that xMOOCs resembled early behavioural and cognitivist approaches to DE. Siemens (2013) commented:

> The pedagogical model that underpins these courses is one of "teacher as expert" and "learner as knowledge consumer." Learning is primarily a process of the

> learner duplicating the knowledge structure set by the course designer and the instructor teaching the course. Weekly course topics are addressed through recorded lectures … [and] interactive activities… assignments are computer-graded. Direct instructor feedback is not common, except in discussion forums where teaching assistants and the course instructor respond to student questions (p. 7).

In short, such course design resembled the structure of courses with one-way communication technology, and was focused on information transmission. Such teacher-centric DE appeared regressive rather than revolutionary. The xMOOCs contrasted strikingly with the connectivist pedagogies, which positioned individuals in the sociotechnical context of learning with technology, for example in socio-constructivism, socio-cultural or socio-material approaches to learning.

2.1.4 Resolution of the Tensions Around Scale in MOOCs

From 2013, scholarly literature shifted to a more unified approach in describing the MOOC landscape. Bayne and Ross (2014) explained that the initial binary characterisation of cMOOCs and xMOOCs was convenient for addressing the distinct underlying historical trajectories of the development of open courses. They clarified however, that to continue with such an over-simplistic categorisation would misrepresent MOOCs and constrain their future development (p. 22). To counter a widespread critique about xMOOCs being exclusively focused on information transmission, Bayne and Ross offered examples that countered mainstream opinion of what would be typical for each of the binary MOOC types. For instance, they discussed the high number of interpersonal interactions found in xMOOCs forums. To challenge the premise that cMOOCs were of better design, they pointed out the challenges of engagement in some cMOOCs and the need to support self-directedness for participation. These counter-arguments suggested that the differences between the two types of MOOCs may be less clear and are unhelpful in understanding learner and teacher experiences and enactments of pedagogy at scale (Ross, Sinclair, Knox, Bayne, & Macleod, 2014). In a similar manner, Bali (2014) urged moving beyond the "good MOOC/bad MOOC debate" towards using "research to interrogate a variety of issues about MOOCs, elearning, and education in general" (para. 1). To move forward in understanding issues common to both types of MOOCs, Margaryan, Bianco and Littlejohn (2014) analysed instructional quality in 76 randomly selected MOOCs. The authors evaluated whether or not the courses' instruction complied with modern pedagogical approaches to learning and instruction. The results of this study

demonstrated that cMOOCs and xMOOCs had different strengths. Among other things, cMOOCs tended to use more authentic learning activities, better integrated new skills and facilitated collaboration. The xMOOCs outperformed cMOOCs in organisation and presentation of course materials. However, none of the analysed courses scored more than 28 points out of a possible 72 in instructional design. The implication was that, irrespective of where the teaching model fit against the binary xMOOC/cMOOC divide, the quality of the MOOC offerings required vast improvements.

The shift to a more integrated understanding of MOOCs has been reflected both in scholarly definitions of the field and in pedagogical practices. Ferguson and Clow (2015) referred to MOOCs as a generic term for provisions that share the aspects of "scale, economic/philosophical perspective (open), location (online) and structure (course)" (p. 51). Veletsianos and Shepherdson (2016) described MOOCs as "an evolving ecosystem of online learning environments featuring open enrolment, characterised by a spectrum of course designs ranging from networks of distributed online resources (cMOOCs) to structured learning pathways centralised on digital platforms (xMOOCs)" (p. 1). Rosé and colleagues (2015) demonstrated in practice that one MOOC can integrate both centralised and distributed approaches to online learning. With the same group of students, they ran an experimental dual-layer course DALMOOC that gave students the choice of following either cMOOC and xMOOC-like design within the same course (Crosslin & Dellinger, 2015; Dawson, Joksimović, Kovanović, Gašević, & Siemens, 2015). The DALMOOC experience showed that individual learners had different preferences as to which design best suited their learning needs.

Paralleling the evolution of a more unified approach to MOOCs, earlier fears that they would disrupt or replace higher education began to dissipate. Reflecting the general sentiment, Selwyn et al. (2015) wrote that:

> the idea of "the MOOC" is clearly broad and mutable enough to be able to accommodate multiple meanings of radical change and conservative continuity … the educational form of the MOOC is neither as trivial as some of its detractors might like to think, nor is it as novel and "disruptive" as some of its supporters would claim (p. 191).

In short, MOOCs have neither disrupted nor replaced the model of higher education. The lack of impact MOOCs have had on the broader education system is well demonstrated by Udacity's failure to compete with community colleges and universities

in the delivery of remedial courses. San Jose State University (SJSU) was in partnership with Udacity to provide low-cost online courses (not MOOCs) to vulnerable students at the same cost as the community college, but the project was discontinued. Educational media source *Inside Higher Ed* reported that "no more than 51 percent of Udacity students passed any of the three classes", whereas the overall passing rate in a traditional, face-to-face SJSU class is 74% (Rivard, 2013). This experience marked the company's move to newly position itself as a means for enhancing learners' employability – a marketing ethos that Udacity has maintained since.

Udacity was not the only MOOC provider transitioning from an offer of "education for all" to re-branding its offerings as upskilling opportunities for increased employability. Coursera has marketed its courses as pathways to better employment as well (Watters, 2015b). In such a manner, MOOC providers appeared to enter the niche of professional and workplace learning. Examinations of internal corporate policies also demonstrated that large corporations have integrated MOOCs into workplace learning practices. In the US, Microsoft, AT&T, and Comcast have started to offer MOOCs for internal employee training as part of their workplace professional development (Watters, 2015b.). According to a 2015 online survey of 222 US-based human resource, corporate learning, and talent development professionals, 44% are producing internal MOOCs and curating external ones, 31% are creating internal MOOCs, and 22% are using externally offered courses only (Meister, 2016). It would appear that MOOCs started to settle into the niche area of non-formal education and professional learning.

The number of higher education institutions utilising MOOCs suggests that most institutions that had plans to deliver them have already begun doing so. To explain, Allen, Seaman, Poulin and Straut (2016) reported that in the US, MOOC-delivering universities represent a small segment of higher education. The percentage of US universities delivering MOOCs increased from 2.6% in 2012 to 11.3% in 2015. Although the segment has grown, it seems that growth has occurred mainly within the group of institutions previously planning to deliver MOOCs. In contrast, the number of institutions without plans to deliver MOOCs remained steady. Hollands and Tirthali (2014) offered several reasons for university involvement with MOOCs. According to their survey, the universities reported offering MOOCs to extend the role of higher education in access to education. Other reasons included public relations strategies (building or maintaining a brand) and profit-related motives (lower cost or profit).

Institutions also mentioned running MOOCs for self-improvement, to enhance learning experiences for their on-campus students, as well as to conduct innovative teaching practices and cutting-edge research.

In 2016, the single largest group (58.7%) of US-based institutions reported that they had no plans to offer MOOCs (Allen et al., 2016), with the lack of interest attributed to the growing belief that MOOCs are not financially viable; that is, they do not bring profit. The degree of MOOC adoption, at least in the US context, is nearing completion. The next stage of MOOC development will likely have an increased focus on quality assurance and accreditation of prior learning to validate learning outcomes.

2.1.5 Latest Trends for MOOCs

Although MOOC providers represent a small segment of the higher education sector, their global outreach is considerable, and the number of students enrolled is high. According to Class Central, an online portal collecting information about MOOCs and MOOC providers, in 2015, students enrolling in at least one MOOC topped 35 million, up from 2014's 17 million (Shah, 2015). Allen and Seaman (2016) asserted that "the number of students touched by a MOOC can easily match that of those taking DE courses" (p. 38). As well as the massive student uptake, MOOC platforms have been adopted all around the globe; MOOC-delivering companies now span countries, cultures, and languages. For instance, FutureLearn is UK-based, Iversity targets German-speaking students, MiriadaX provides Spanish-speaking MOOCs, and France Université Numérique (FUN) is a French government-funded MOOC platform. Edraak, an Arabic language MOOC platform with nearly one million learners, is yet another strong example, as well as China's XuetangX that already exceeded six million users in 2016 (Shah, 2016). Latin American countries have also entered the field with two platforms of their own – Telescopio in Guatemala and Veduca in Brazil – with some 62% of the countries in the region involved (Perez-Sanagustin, Maldonaldo, & Morales, 2016). Altogether, these numbers demonstrate a vast uptake that has been established globally in a short amount of time.

A common critique of the MOOC demographic is that those enrolling and succeeding already possess the necessary learning skills and knowledge (from past higher education degrees) and often come from well-off backgrounds (Ebben & Murphy, 2014; Hansen & Reich, 2015). For example, a study by the University of Pennsylvania demonstrated

that 80% of MOOC completers had already completed a college degree (Zhenghao et al., 2015). Critics maintain that the educated few are engaging in these learning opportunities, thereby widening disparities and inequality gaps in learning and education. In an interview with Watters (2015a), Cauldfield argued that "the numbers alone are not indicative of democratised access to education, as the real challenge is if the access to the perceived benefits of education, i.e. completion, is democratized" (cited in Watters, 2015a). Koller, former Coursera CEO, addressed the argument that the MOOCs' value is limited to the already privileged by noting that 25% of MOOC students without degrees represent some 2.5 million people (Knowledge@Wharton, 2015). It is also worth noting that not all those who already have college degrees come from OECD countries. Not all degrees earned around the world carry the necessary weight or prestige required for competitive employability. Moreover, those suffering from disadvantage or limited access to information and education have trouble engaging in higher education. In support of these views, Coursera's analysis (Zhenghao et al., 2015) revealed that MOOC-takers from developing countries with lower socioeconomic power, or with lower education levels are the ones likely to regard MOOC learning as beneficial in post-course evaluations. Such examples demonstrate the value of MOOCs through increased access to information for thousands of learners who otherwise may not have access to the same quality of educational offerings. It should be noted, however, that some of the benefits derived from MOOCs are becoming somewhat overshadowed by monetisation schemes. Providers have begun introducing more aggressive payment mechanisms, such as Coursera's charges for taking in-course assessments alongside already existing charges for certificates.

Positioning MOOCs as an online non-formal education provision highlights their potential role within as well as beyond formal education systems. MOOCs emerged within the historical context of online education as an educational provision for adult learners. The age of the MOOC learner has been reported to range from 18 to over 60. The dominant age group represents the working population of 20-40 year-olds (University of Edinburgh, 2013; Grainger, 2013; Harrison, 2013). Myers and colleagues (2014) claimed that within lifelong learning systems designed by policy makers, education for these age groups falls within the domain of adult learning, a "purposeful and directed learning undertaken by adults, either alone or in groups, to increase knowledge and skills, and/or change behaviours, values, or beliefs" (p. 3).

They described learning opportunities for adults as, for example, incidental learning, informal learning, non-formal learning, and formal learning. Czerniewicz and colleagues (Czerniewicz, Deacon, Fife, Small, & Walji, 2015; Walji, Deacon, Small, & Czerniewicz, 2016) situated MOOC provisions in the domain of non-formal education; and Myers and colleagues (2014) defined non-formal as:

> … structured learning that includes activities such as: participation in courses that are not part of an educational program; workshops; seminars; private lessons; and guided/organized workplace training … that do not usually follow the "ladder" system … characteristic of formal learning (p. 2).

From this perspective, MOOCs represent planned learning activities structured by teachers or students, which do not carry the formal accreditation that enables learners to re-engage with lifelong learning systems. Although in MOOCs there is an option to receive a certificate of completion, MOOC certificates are rarely endorsed by institutions to the same academic standards as institutional transcripts. There is also no formal recognition of certificates from the labour market, unless a MOOC is offered in collaboration with a recruiting company or a company sponsoring ongoing professional development. This lack of formal acknowledgment of MOOC-certification suggests that their current role is in offering adults education access and professional learning at a lower cost, alongside the promise of personal growth and fulfilment.

Several attempts to award credits for MOOCs were reported in late 2016. In one such announcement, a consortium of two European and two Australian universities announced they would start providing course credit for successful completion of MOOCs they deliver (Mulder, 2016). The initiative, to begin in 2017, has been described as the "virtual Erasmus exchange" in reference to the established exchange programs between the universities. These credits appear to be valid only among the consortium's institutions. FutureLearn also announced its partnership with Deakin University in December 2016. In 2017, the two institutions will offer a range of postgraduate degrees falling under the Australian Qualification Framework (Tucker, 2016). Awards of university-issued credits for some MOOCs may represent the formalisation of MOOC provisions. This may also signpost a push to more intricate systems associated with recognition of prior learning. The fate of these initiatives is still unknown; however, they may well mark a new page in the evolution of MOOCs.

2.1.6 Section Summary

Section 2.1 described the emergence of MOOCs and positioned them as a form of non-formal online education. In the past several years, MOOCs have catered to millions of learners. Dozens of universities worldwide, private industries and governments have invested in the development of MOOCs to offer new opportunities for learning and skill development to learners at little or no cost. The following section examines the research investigating the quality of the MOOC learning experience. The rapid development of MOOCs has necessitated the exploration of efficient processes and practices that can improve the quality of the learning experience in response to increasing concerns about low retention rates and ineffective scaling of interactions. To overcome these challenges, researchers have engaged in the analysis of learner activity traces. However, much of the research has been framed under the logic governing the provision of formal education.

2.2 Formal Perspectives on Non-Formal Courses

Within a relatively short period of time, individuals and public and private education institutions have adopted MOOCs in significant numbers. The scale of adoption has contributed to a surge in research attempting to understand, critique, explain or evaluate MOOCs. Since 2008, numerous articles, opinion-eds and institutional reports have been written. Major educational technology journals have also dedicated special issues to researching MOOCs (e.g. *British Journal of Educational Technology, 46*(3) (Ng'ambi & Bozalek, 2015)); *The International Review of Research in Open and Distributed Learning, 15*(5), (Siemens, 2014); *Distance Education, 35*(2), (Jona & Naidu, 2014); *Journal of Computer-Assisted Learning, 32*(3), (Kirschner & Kester, 2016). From 2014 to 2016, at least three ongoing conferences have been established to focus on MOOCs and learning at scale. Collectively, this has resulted in numerous annual proceedings dedicated to research on MOOCs and other large-scale education offerings (e.g. 'European MOOC Stakeholder Summit'[1] , 2014, 2015, 2016, 2017;

[1] http://emoocs.eu/

'Learning with MOOCs' Conference[2] , 2014, 2015, 2016, 2017; 'Learning at Scale' Conference , 2013, 2014, 2015, 2016, 2017)[3].

The first review of MOOC research stemming from the period 2008-2013 comprised some 45 articles that were related mostly to connectivist courses (Liyanagunawardena, Adams, & Williams, 2013). More recently, Veletsianos and Shepherdson (2016) analysed 183 papers for their review of research in 2013-2015. To select a dataset for the latest systematic review of research measuring learning in MOOCs, Joksimović et al. (2017) read through more than 1000 papers with keywords focused on learning outcomes, as well as engagement, retention, interactions, behaviour, participation, completion and similar. The growing sample of these systematic reviews undertaken within short intervals indicates the rapid increase of research activity in this area.

The booming amount of research has in part, resulted from the availability of and access to the student interaction data collected via the MOOC platforms. Availability of trace log data has enabled research groups to start exploring the collected "terrabytes of data about what students clicked" (Reich, 2015, p. 34). Despite the vast quantity of papers, Reich (2015) noted their lack of implications for teaching and learning: "Few MOOC studies make robust claims about student learning, and fewer claim that particular instructional moves caused improved learning" (p. 34). In a similar manner, Rafaghelli and colleagues (2015) described the methodological approaches employed in MOOC analysis as being in the initial phases of development.

This section offers a critical account of the MOOC research on learner interaction. From the literature review, it is asserted that the MOOC research agenda was governed by the issues typical to the formal education paradigm. The review demonstrates that the studies focused predominantly on learner performance and completion. It is argued that these issues are of relevance to formal university courses. To this end, Section 2.2.1 shows that MOOC learner success has been examined through approaches that consider MOOC contexts to be homologous to university contexts. Studies converged on reporting enrolment numbers and the types of activities maintaining learner engagement until completion. Section 2.2.2 illustrates how in-course forum interactions

[2] http://www.learningwithmoocs2016.org/

[3] http://circlcenter.org/cyberevent/learning-scale-2017/

indicative of social learning activity have also been interpreted as complementary to performance and completion. To offer an alternative research agenda, the section notes empirical findings suggestive of the specific MOOC interaction patterns. Section 2.2.3 introduces research on learner activity patterns, and Section 2.2.4 highlights the empirical studies bringing out the differences between MOOCs and formal courses. The section argues that the patterns of interpersonal interactions in MOOC forums are distinct from those observed in more formal class settings, and their idiosyncrasies need to be accounted for.

2.2.1 Learner Success in MOOCs

In the release of the first institutional reports, student dropout rates in MOOCs were pronounced as one of the major challenges to be overcome (Koller, Ng, Do, & Chen, 2013). MOOC retention rates were reported to be less than 6.5% of enrolled participants (Jordan, 2014). In a manner akin to on-campus university courses and programs, retention rates were calculated against the number of students enrolled. The critique of MOOC retention rates was unsurprising. Retention rates are frequently used to measure the societal value ascribed to higher education (Mullaney, 2014). In a bid to enhance learner retention, early MOOC research focused on the factors that underpin learner engagement and success (Gašević, Kovanović, Joksimović, & Siemens, 2014) as defined by completion and retention (Veletsianos & Shepherdson, 2016). The interest in completion and retention was the determining interest factor driving the scholarly inquiries into learner behaviours, motivation, performance, participation, course design, learning strategies, and interaction with peers and learning resources. In this context, the success of a learner was defined through the same metrics and lens that were used for formal education settings; namely, the percentage of completing students and individual performance in assessment activities.

The continued dominance of the formal educational mindset in MOOC research has been well demonstrated through a systematic literature review by Joksimović et al. (2017). The authors found that academic achievement in the form of a final exam or an accumulated course grade was the most common measure of learning in MOOC research. Joksimović et al. (2017) cited the following studies to support their claim: Bergner, Kerr & Pritchard (2015), Coffrin, Corrin, de Barba, and Kennedy (2014), Crossley et al. (2015) , Gillani and Eynon (Gillani & Eynon, 2014), Kennedy, Coffrin,

de Barba, and Corrin (2015) , Koedinger, Kim, Jia, McLaughlin, and Bier (2015), Ramesh, Goldwasser, Huang, Daume, and Getoor (2014), Tucker, Pursel, and Divinsky (2014), Wang, Yang, Wen, Koedinger, and Rosé (2015). Course outcome has also been defined by certification and completion, for example "no certificate", "normal certificate" and "certificate with distinction" (e.g. Brooks, Thompson, & Teasley, 2015), as well as "completing some exams" and "completing all exams without passing the course" (Engle, Mankoff, & Carbrey, 2015). Learner persistence was another commonly accepted approach to measuring student learning outcomes. For instance, various studies included quiz grades as points of reference for persistent behaviour (Zhu et al., 2016), as well as engagement with content and assessment, or the discussion forum. Ye and colleagues (2015) defined a dropout as accessing fewer than 10% of the lectures and performing no further activities or assessment activities. Vu and others (2015) predicted dropout events as a consequence of stop in engagement with learning events that span course activities, including forums as well as quiz grades. Whitehill, Williams, Lopez, Coleman, and Reich (2015), defined a "stop-out" as a student not earning a certificate and taking no action between a certain point in time and the time of the issuance of certificates. Altogether, the studies noted above imply that the learner who does not persist receives fewer learning outcomes, as would be the case in formal online settings. However, in a MOOC setting, the outcome intended by the learner may in fact be undertaken by sampling parts of the course. Fundamentally, the studies described learner activity with the focus on completion and performance, not necessarily learning. Such an approach is typical of a university setting, but does not fully address the attributes of open courses where completion may well be outside a learner's intention.

2.2.2 Interpersonal Interactions and Performance

As summarised in the previous section, research into the MOOC learning outcomes has defined learning success as completion or persistence. These aspects of learning are inherent within formal education approaches. Similarly, research that analyses learning through interpersonal in-course interactions has linked participation in social learning activities with learner performance. Interpersonal interactions in platform-based MOOCs normally take place on discussion forums that record student posting activity. Studies have been conducted that investigated the volume of posts or number of student connections in relation to learner grades (Blom, Verna, Li, & Dillenbourg, 2005;

Breslow & Pritchard, 2013; Goldberg et al., 2015). Such analyses were conducted mainly at the level of the individual and viewed social learning activity (i.e. forum participation) as complementary to academic performance (grades). For example, Kizilcec and colleagues (2013) demonstrated that completing learners who attempted a majority of assessments were also more active on forums. The majority of studies undertaken demonstrated that, inasmuch as there is a relationship between forum activity and grades, there are also students who participate in forum activity without completing the course.

The MOOC studies associating grades with a learner's social activity resemble the contested attempts to research computer-supported collaborative learning in formal education settings. In a seminal paper, Dillenbourg and colleagues (1996) argued that research questions asking if collaborative activity benefits learning are reminiscent of the behaviourist "effect" paradigm, which addressed whether working together is better than working alone. They reviewed existing findings in formal research and concluded that the evidence was diverse. A number of studies demonstrated that collaboration enhanced performance in certain specific subject- or skill-domains. They also cited other results that were not as optimistic, presenting a spectrum of outcomes from no learning gains, to partial gains for one of the partners, to a regression in academic performance. Although some insights can be gained by looking into the association between social activity and grades, such an approach ignores the socio-cultural and relational aspects of learning (further discussed in Section 2.3).

Besides analysing the relationship between grades and the number of student posts, MOOC researchers have investigated the association between grades and more context-aware metrics of social engagement. Particularly, individual positioning in a network of other learners was perceived to be indicative of student social engagement. This strand of research draws on the higher education retention studies that suggest that learners who are well connected in the social network are more likely to do better academically (Dawson, Macfadyen, & Lockyer, 2009; Gašević, Zouaq, & Janzen, 2013; Thomas, 2000; Vaquero & Cebrian, 2013). In line with this tradition of research, learner grades in MOOCs were compared in relation to the number of people an individual replied to in a forum or individual positioning in relation to other forum posters (Dowell et al., 2015; Jiang, Warschauer, Williams, O'Dowd, & Schenke, 2014; Joksimović, Dowell, et al., 2015). Similar to the learner success studies, these analyses were framed from the

perspective of formal course considerations, with fixed boundaries and set numbers of people participating, who, most importantly, have the implicit intention to complete. However, in a MOOC, not all forum participants are interested in completing the course or in ongoing assessments. Studies that account for all posted forum interactions from the networked perspectives ignore motivational differences in learner behaviour and the different quality of interactions. As a result, focusing on the relationship between completion and social learning activity limits the capacity of social network analyses to fully explain the MOOC context.

2.2.3 Classifying MOOC Learners

Research classifying learner participation patterns has offered evidence that individuals engage with learning opportunities in MOOCs in different ways. By analysing traces of learner activity, researchers classified MOOC learners into sub-populations based on the characteristics of their engagement with the course. For instance, Hill (2013) categorised MOOC students as: "lurkers", who only enrol; "active" students, who fully engage with course materials, forum and quizzes; "passive" students, who consume the content without engaging in social learning activity; and "drop-ins", students whose behaviour was active only for some part of the course. Kizilcec, Piech, and Schneider (2013) analysed learner log data across the entire course and described learners as belonging to one of four groups: "completing", "auditing", "disengaging", and "sampling". Ferguson and Clow (2015) replicated the profiling of learner behaviours conducted by Kizilcec, Piech and Schneider (2013), deriving more clusters of learners: "samplers", "strong starters", "returners", "mid-way drop-outs", "nearly there", "late completers", and "keen completers". Their profiling targeted the identification of learners in need of support. Ferguson and Clow (2015) concluded that the four clusters where learners engaged but failed to complete were "strong starters", "returners", "mid-way drop-outs", and "nearly there". Other typologies distinguished between "no-shows", "drop-ins", "latecomers", "observers", "drop-in latecomers", "engaged", and "non-engaged" (Alario-Hoyos, Pérez-Sanagustín, Cormier, & Delgado-Kloos, 2014), "auditors", "active" and the "qualified" (Coffrin et al., 2014), "active with normal participation", "active with distinction", and those who "failed", as well as "active viewers" and "passive viewers", "wiki-users", "drop-outs" and "completers" (Sharma, Jermann, & Dillenbourg, 2015). The overall focus of engagement patterns is not on what had been learnt, or whether or not the individuals had satisfied their learning

intentions. Rather, they capture when the learners started using a MOOC, how long they continued to use it, and what kinds of MOOC resources they used. Despite the granular differences, the basic distinctions were drawn between learner behaviour associated with course completion, and the varieties of course completion.

One common finding across the classification was that individuals taking a MOOC tended to use different strategies to achieve what appeared to be individual learning goals. Some learners intended to sample, some wanted to complete it and some may have decided to return and complete it long after the course has started. These different motivations reflected the very diverse patterns of behaviour. Analysis across the MOOC learner classification studies suggested that the individuals exhibited varying preferences as to when, how and in which combination they watched lectures, used the forum, or completed assignments, if at all. From this, it is clear that learner classification studies in MOOCs make an important contribution: that learner activity in open settings varies, as does the use of learning resources.

2.2.4 Recognising the Differences Between MOOCs and Formal Courses

As demonstrated through learner classification studies (see Section 2.2.3), individual patterns of engagement with MOOC resources have been diverse, and are not always oriented towards persistence and completion. In light of these differences between the expected behaviour in a formal course and a range of divergent learner pathways in a MOOC, DeBoer and colleagues (2014) urged re-operationalisation of the language borrowed from educational settings. The authors argued that such language, when used to describe MOOCs, did not capture the specificities of this new learning environment. According to DeBoer and colleagues, individuals taking a MOOC do not fall under the conventional concept of a "student". Instead, individuals signing up for a MOOC should be differentiated into "registrants" for those who enrol, and "users" or "learners" for those who interact with the course. In an important empirical study that made progress with these distinctions, DeBoer and colleagues (2013) put forward a compelling argument against applying the conventional concepts borrowed from education research, such as those related to learner behaviour (e.g. enrolment, participation, curriculum, and achievement) to the MOOC context. The authors explained that in a MOOC, "enrolment" refers only to the expression of interest in learning, as opposed to the way the term is used in a formal course, which refers to all

students committed to completing a full class experience (Liyanagunawardena, 2014). For example, individuals signing up after a MOOC start date may not all be as interested in completing the full "class" experience. DeBoer and colleagues (2013) presented a similar critique of the concept of participation. In formal education, participation may be measured by attendance, whereas in a MOOC, it manifests through a range of observed behaviours, from using the content to "raising your hand" by posting in forums. In their study, DeBoer et al. (2013) emphasised that the diversity of behaviours shows that participation may carry different meanings for different people. They further noted that patterns of participation vary over time. In addition, the authors argued that referencing curriculum as set syllabi, textbooks, lesson plans, or content standards may not be relevant for MOOC contexts. In a MOOC, curriculum relates to the learning opportunities that individuals choose to take up rather than a systematised and structured experience that is finalised by a teacher. As a final point, DeBoer and colleagues critiqued the use of the term "achievement". In MOOCs, achievement can no longer be regarded as successful participation in all of a course's assessments. Overall, these distinctions demonstrated that formal education mindsets were embedded within the very language used to describe learning environments, including that of MOOCs. Moreover, this study by DeBoer and colleagues remains as a highly influential piece of work for MOOC research. The study brought to the fore that it is the individual's motivations in a MOOC, not the teacher's intention, that defines how much activity a learner opts for in the aspects of curriculum, achievement, enrolment and participation.

Motivations and the use of resources. Research into learner behaviour supports the premise that MOOC learners engage with the course resources that are specific to their individual learning needs and interests. The differences between learner choices are largely explained by their differing motivations to learn. For instance, through interviews with MOOC participants, Eynon and colleagues (2014) identified four underlying motivations for engaging with MOOCs: (1) just-in-time learning; (2) just-in-case learning; (3) lifelong learning; and (4) education access. These four orientations to learning in a MOOC reflect the different resources that learners would focus their engagement on, depending on their underlying motivations. From their personal agendas, individual learners would opt to engage with some aspects of the in-course

resources and/or other MOOC-takers. Brief descriptions of the orientations are as follows:

- Just-in-time learning defines learners who intend to take a MOOC to solve a particular problem. The learners expect to develop knowledge for specific decision-making processes. People within this category are generally not interested in the more social aspects of the course. Oriented towards individual learning, just-in-time learners consume mostly video lectures and prescribed readings.
- Just-in-case learning is associated with professional profiling. These learners expect to develop skills for future employment prospects. People within this category treat the social interactions as a source of information and use mostly video lectures.
- Lifelong learning is motivated by the learner's aspiration for formal accreditation. They expect to receive a traditional educational experience. Within this category, people treat social interactions as an opportunity for knowledge construction and networking. Participants within this group also use lectures and readings, and identify study groups.
- Educational access as a learning motivation characterises those individuals who take MOOCs simply for the sake of learning. People who are motivated to take a MOOC because they just want to learn expect to experience the global outreach and connectivity. Such learners enjoy the experience of learning with, and from, the crowd. To them, the crowd is a source of knowledge. These learners use mostly forums, as well as provided video content.

By distinguishing between learner motivations, Eynon and colleagues (2014) drew attention to the fact that only some learners are interested in engaging with other learners. Furthermore, the extent of this engagement with others can vary. For instance, those interested in learning for learning's sake appeared to have had more commitment to interpersonal interactions and little interest in certification. On the other hand, some of those interested in formal credentialing could have opted for some short interpersonal exchanges. Those sampling a specific piece of content could have opted out of interactions with other learners completely. These differences contrasted with formal education, where engagement with other learners is often an integral and required part of an accredited course. These discrepancies further strengthened the evidence that distinguished MOOCs from formal courses.

Several analyses of interpersonal interactions in a MOOC setting similarly demonstrated the uneven learner engagement in social learning activity. The study by Vu and colleagues (2015) found that the relationship between learning performance and social interactions may not be bi-directional. That is, some forum-active learners did not perform better in a course. The study by Dowell and colleagues (2015) also showed that forum participation in non-formal settings such as MOOCs goes beyond academic performance. The authors quantified social activity through the concept of social capital; that is, the number of individuals one replied to on the MOOC forum. They further demonstrated that socially engaged MOOC-takers with higher grades and socially engaged participants with higher social capital were not the same individuals. These individuals differed in the quality of the discourse they produced in forums: those with higher grades tended to be more focused and narrow in their discussions. In other words, although high learning performance was correlated positively with active social interactions, active social interactions were not necessarily associated with performance.

Temporal Patterns of Interpersonal Interactions. In addition to overall motivations and learner agenda for social activity, the temporal patterns of forum use in a MOOC setting were also found to be distinct. Yang and colleagues (2014) analysed the association between the time learners started posting in a forum and the persistence in their forum posting activity. Each group of students who started posting in the new week of the course was considered a new cohort. The results showed that cohorts starting forum activity early were more dedicated and more likely to have higher forum engagement. In contrast, the cohorts that joined the forum activity in later weeks appeared to be less active and posted at a lower rate than learners starting earlier. The authors explained these observed differences in activity through motivational differences between the cohorts. In addition, the study adds further significant observation as it revealed the uneven pattern and variations in the starting time of posting activity. Although learners disengage from forum activity, new learners join to replace the pool of forum participants over the course duration. These temporal patterns are strikingly different from those observed in formal online education, where a fixed group of learners interact with one another in a sustained manner.

Social learning activity in a MOOC setting appeared to be unlike that observed in formal courses. Only a portion of all MOOC-takers intend to use forums for learning. In addition, those who use forums for learning may not be interested in assessment or

course completion. Therefore, the motivation for using a forum varies among individuals. As a result, MOOC forum participation can be considered an experience occurring in a group of people with different agendas. Moreover, the membership of this group constantly changes, and the scale of this dynamic was found to be higher than in conventional formal courses. These fluid boundaries of MOOC forums calls into question the socio-emotional quality of relatedness associated with closely-knit small groups, and how this can occur in open online learning environments.

2.2.5 Section Summary

Section 2.2 has argued that, research into MOOCs has been interpreted largely through the concepts associated with formal education. Learner success has been examined in relation to completion, retention, persistence and performance in a given course, rather than in relation to individual motivation to learn. In a similar vein, individual learners' patterns of engagement have been grouped to capture the learners who are successful in formal education terms; that is, those who complete the course with reasonable academic achievements. Studies have examined the volume of activity as complementary to learner success in terms of course completion. In attempts to describe these patterns of participation, researchers recognised that language borrowed from educational settings, such as "enrolment", "student" and "curriculum", does not fully describe the specificities of MOOCs. Insights into the relationship between in-course activities such as posts and different learner motivations suggests that MOOC-takers engage with course resources and with one another in a variety of patterns, and for more diverse purposes than they do in a university course.

Moreover, empirical evidence suggests that the patterns of interpersonal interactions in MOOC forums are distinct from those observed in more formal class settings (DeBoer et al., 2014; Eynon, 2014; Yang et al., 2014). First, the open course affords learners an opportunity to act upon their divergent learner needs, commitments and choices. Behaviour that is driven by these varying intentions gives rise to diverse and fluctuating levels of learner activity. For those individuals interested in seeking interpersonal interactions for their learning, this dynamic creates a social environment with constantly shifting boundaries for the socially engaged learner. These new dynamics of a MOOC's social environment may have implications for how learning can be facilitated.

The next section explains that DE pedagogy and contemporary learning theories imply the presence of relational processes between learners. These relational processes unfold through sustained interpersonal interactions. The relatedness of individuals in the learning process enhances the quality of the learning experience and offers learners cognitive and emotional benefits. Much of formal education theories and models rests upon the assumed aggregation of interpersonal interactions over a course's duration, which further evolves through an increased quality of relatedness between individuals. Given the discrepancies between formal and MOOC interactional patterns, the applicability of formal education models to teaching in MOOCs can be contested.

2.3 The Role of Interpersonal Interactions

Section 2.2 demonstrated that research has approached MOOCs as if they were more formal university offerings. For instance, learner interactions with content, peers and instructors have been reviewed in relation to course completion and assessment performance (e.g. He, Bailey, Rubinstein, & Zhang, 2015; Jiang, Fitzhugh, & Warschauer, 2014; Jiang, Warschauer, et al., 2014; Ramesh, Goldwasser, Huang, Daum, & Getoor, 2014). However, empirical studies offer evidence that MOOCs are distinctly different from formal online courses: the extent of engagement with content, peers and instructors in MOOCs is based on learner intention and motivation (Eynon, 2014; Reich, 2014). In other words, completing the course or participating in forum activities may not be a part of a learner's intention, neither it is not a MOOC requirement either.

Since formal educational approaches are limited in explaining individual learning activity in a MOOC, investigating interpersonal interactions in MOOC settings through a formal lens does not address the idiosyncrasies that exist between formal and non-formal educational environments. Moreover, as this section will emphasise, interpersonal, learner-to-learner and learner-to-teacher, interactions are central to contemporary theories of learning and models of instruction that are applied in formal online settings. In particular, Section 2.3.1 demonstrates the significance of interpersonal interactions within the analytical lenses examining learning. The section maintains that contemporary learning theories often assign a central role to the interpersonal interactions, as well as interpersonal relationships and the social contexts emergent from them; in addition, that the formation of interpersonal relationships

through interactions have been associated with learner cognitive and emotional benefits. Section 2.3.2 will further contend that interpersonal interactions are significant for learning within DE models. The section will explain how formal education models such as the community of inquiry (Rourke & Kanuka, 2009), knowledge forum (Scardamalia & Bereiter, 1996), or phases of knowledge construction (Gunawardena, Lowe, & Anderson, 1997), among others, often embed a particular notion of how interpersonal interactions unfold. The role assigned to interpersonal interactions by theories and models of online learning is critical, as MOOC settings may not afford a similar development of interpersonal interactions. In fact, as discussed in the subsequent Section 2.4, the properties of how interpersonal interactions unfold in MOOCs has been largely unexamined, thus, calling into question the use of formal educational approaches.

2.3.1 Interpersonal Interactions and Learning

This subsection demonstrates the central role of interpersonal interactions within learning theories. Moving from cognitivism to constructivist approaches and transactional approaches to learning, this section presents the importance and growing role assigned to interpersonal interactions within learning theories. It also contends that the socio-emotional aspects of interpersonal interactions (their relatedness) have a positive impact on learning, while emphasising that social contexts embedded within socio-constructivist and socio-cultural theories contain certain relational qualities of interactions and of the emergent social contexts.

Cognitivism. The cognitivist paradigm centres on the individual, with limited focus on the social aspects of learning. In the early traditions of the cognitivist approaches to learning, knowledge was considered to be acquired by an individual (Sfard, 1998), from a particular knowledge source. Cognitivism addresses what Bruner (2009) described as a "computationalist" nature of mind. He argued further that computationalism is centred on storing and organising finite well-formed information units. In that sense, cognitivists have referred to learning as the acquisition of mental schemes, i.e. accurate representations of information to be learnt (Koschmann, 1996). From such a perspective, interpersonal interactions serve as a bridge between a knowledge source and the mind of a learner; that is, through interpersonal interactions, a learner can reach an instructor (human or computer-based), or vice versa. Although interpersonal

interactions are part of learning processes in cognitivism, their role is limited: a teacher or an intelligent tutor can pose problems and provide feedback to help a student construct "the proper representation of a problem space" (Koschmann, 1996). However, a process of knowledge transfer remains largely decontextualised from the social environment where it unfolds.

Constructivist Approaches to Learning. Building on behaviourist and cognitivist paradigms, constructivist and interactional epistemologies situate learner–content interactions within the context of interpersonal interactions (Suthers, 2006). Design for learning from the constructivist perspective integrates ideas by Piaget (1954), Vygotsky (1980/1978) and Dewey (1933, 1958), regardless of the philosophical differences between these learning theorists (Anderson & Kanuka, 1999; Koschmann, 2001, 2002). Anderson and Kanuka (1999) summarised the contributions this group of theoreticians made to pedagogical practices. In essence, knowledge was no longer considered to be as a product that can be transferred from teacher to learner. Theorists stressed the importance of an individual's experience with a task, content and context – all interacting and contributing to a learning process. In short, the constructivist perspective on learning implies that an individual has an active role in constructing knowledge, which may take place through scaffolding; that is, learner support for conceptual understandings that enables a learner to move further within the zone of proximal development (Pea, 2004). Alternatively, a peer, rather than a superior, can engage in a constructive dialogue to aid in knowledge construction. The notions of scaffolding, zone of proximal development, and knowledge construction through dialogue implies a need for two-way learner–learner interactions. Other features embedded within the socio-constructivist pedagogy, include: (1) the importance of prior knowledge into which new knowledge structures are integrated; (2) the significance of the context in how these knowledge structures develop; (3) the activeness of a learner as opposed to receiving knowledge through a prior way of thinking; (4) the role of language in mediating learning and social experiences; and (5) the need to assist individuals to regulate their own learning (Anderson & Dron, 2012; Anderson & Kanuka, 1999). Within constructivist and interactional epistemologies, interpersonal interactions complement, or motivate cognitive processes through, for example, dialogical engagement with peers, or collaborative engagement with various scaffolds.

The theorised impact of interpersonal interactions on learning was upheld through the reviews of evidence by Webb (1982), Bransford, Brown and Concking (2000), and Dillenbourg, Baker, Blaye and O'Malley (1996), among others. They offered solid evidence that peer interactions can enhance the individual's learning. In particular, learning through interpersonal interactions could lead to better understandings achieved by verbalising a problem, seeking or receiving help, or co-constructing a solution. The role of interpersonal exchanges became not so much to correct, as in cognitivism, but to enhance one's knowledge and deepen one's cognitive engagement.

Socio-cultural Approaches to Learning. In the next generation of learning theories, interpersonal interactions became essential to learning as a part of a context where learning becomes situated. Bruner (1996) explained that such socio-cultural approaches did not negate prior computationalist approaches; however, besides computationalism, learning could also be centred on how human beings communicate and transform meaning in cultural (i.e. social) settings. Bruner explained further that the evolution of the mind is linked to the development of a way of life, where "reality" is represented by a symbolism shared by members of a cultural community in which the technical–social way of life is both organised and construed in terms of that symbolism (ibid., p. 160). The individual expression of culture is, therefore, inherent in assigning meanings across settings and situations. To know what the activities are about, they need to be interpreted in relation to their cultural context. Here, the origin of individual understandings and the value assigned to them is embedded within culture. Bruner also noted that the existence of "private meanings" is of no importance; what matters is whether or not meanings serve as a basis for cultural exchange (ibid., p. 161). Because of this relationship between knowing in the mind and co-constructing in a shared space, the socio-cultural perspective interprets knowing and communicating as interdependent entities.

Socio-cultural theories of learning view interpersonal interactions as comprising group-level learning where knowledge and meaning are created jointly through interaction (Suthers, 2006). According to Sfard (1998), through such a lens, learning can be perceived as participation. The mechanism for learning shifts from the recipient or constructor of knowledge to a participant in the practices of a community. The result of participation comes to be understood not only in terms of a change in knowledge, but through the construction of a new identity (J. Brown & Duguid, 2000; Lave & Wenger,

1991). As such, interpersonal interactions are considered as integral to socio-cultural approaches to learning. Interpersonal interactions serve as a means of participation and communication – an essential part of shared practices. Socio-cultural approaches (e.g. Engeström, 2005) also assign significance to the tools and situated activities as these create value for situated, meaningful learner–learner and learner–artefact interactions. The socio-cultural view of learning assumes the continuous evolution of context from interpersonal interactions.

Relational Quality of Interactions. The rise of the social, through both social constructivism and socio-cultural theories, also highlighted the relevance of socio-emotional aspects to learning. Manifested in learner–learner interactions, interpersonal aspects of learning tapped into personal and emotional domains. At the core of this perspective is an understanding that, over time, interpersonal relationships acquire a quality associated with an emotional connection between two people. Interpersonal interactions, described by the quality of relatedness (i.e. underpinned by emotional connection) are seen as supporting achievement (Meyer & Turner, 2002). In addition, this quality of relatedness is also seen to vary in terms of degree of affect, which results from a history of interactions between individuals.

Through social interaction, learners establish interpersonal relationships that impact on the quality of learning experiences. For instance, relationships can enable individuals to overcome difficulties associated with specific academic tasks and challenges. By activating their relationships, individuals can receive instrumental help, but also emotional support and companionship (for a review see Martin & Dowson, 2009). Connel and Wellborn (1991) synthesised evidence from empirical studies showing that positive emotional attachment to peers, teachers, and so on promotes positive feelings of self-worth and self-esteem. A heightened sense of belonging to a learning group – a feeling developed as a result of an individual's striving to form and maintain positive interpersonal relationships (Baumeister & Leary, 1995, p. 497) – was also found to be associated with higher student satisfaction from educational experiences (e.g. Shackelford & Maxwell, 2012; Shea, Sau Li, & Pickett, 2006). In short, research supports the view that interactions resulting in relatedness (underpinned by certain affective quality) indirectly facilitate and hinder learner motivation, engagement and achievement.

The quality of the relatedness within a group has also been shown to have an impact on how the collective interacts and learns. The web of relational interactions may evolve into group-level relational processes. Martin and Dowson (2009) summarised the way relationships affect achievement motivation as directly influencing motivational beliefs and emotions. Through ongoing social interactions, individuals socialise into attitudes and values inherent in the environment to which they relate. In higher education contexts, that translates in learning how to function effectively in a study environment, or how to become a professional in a particular field. Beliefs influenced by the environment can regulate individual behaviour, in stronger goal-striving and self-regulation. For instance, according to Ryan and Deci (2000) students can internalise some of their teacher's beliefs by both forming habits of behaviour in particular situations and developing an identity of being within an academic situation. In other words, based on interpersonal relationships, from a psychological perspective, learning came to extend beyond knowledge exchanges or construction, towards socialisation of values through a particular learning and/or professional environment.

Transactional Approaches to Learning. In recent years, DE pedagogies have been inspired by socio-material approaches to learning. As emerging technologies could afford learner–learner and learner–artefact interactions, learning theories needed to account for the elements of technology in learning processes. In highlighting this need, Andrews and Haythornthwaite (2007) urged for the development of a new generation of learning theories that could explain the relationship between technology, the individual and the social. The sociotechnical perspective is considered to be capable of addressing these issues.

A sociotechnical perspective considers that the social system of learners and of the technical system of technology interact with one another in producing affordances for learning (Jones, 2015). Hence, the relationship between the social and technological is treated as interdependent and inseparable. The sociotechnical perspective rejects the deterministic approach, which claims that either the social or the technological drives the process of learning (Creanor & Walker, 2010). By treating the social and technological elements of learning systems as equal, no prior assumption is held as to the importance of one aspect over another.

The sociotechnical paradigm is also socio-cultural in nature. Socio-material networked approaches allow for the integration of constructivist and cognitivist approaches into the learning process. In socio-material approaches, on the one hand, a learner's autonomy and agency is situated within a created web of relationships and connections to other individuals, material objects and ideas. On the other hand, the socio-contextual aspects are considered to either constrain or enable learner activity. The integration of individual agency and the group-level structure within a networked socio-material system offers possibilities to address different levels at which learning may be taking place. Communication and connecting different actors, tools and ideas through interpersonal interactions are essential means of creating the network structure and the networked context.

The role of interpersonal relationships in underpinning interactions is also central to sociotechnical approaches, as the understanding learning in sociotechnical systems is socio-cultural in orientation. However, a networked perspective on learning distinguishes between the different types of relationships. Socio-constructivist settings focus on the homogeneity of relations – as learners interact, they move towards establishing strong interpersonal relationships and a sense of unity of purpose. Networked learning researchers emphasise, however, that the strength of relations formed between individuals within a particular group are likely to differ. For example, de Laat (2006) described collective learning as taking place in networks, communities and teams, each of which can be represented through a networked structure. Such distinctions allow for one to differentiate between weaker and stronger relations through the use of "network" as a metaphor, which de Laat (ibid.) used to reference a social structure that has no leadership, is based on weak relationships, is driven towards problem-solving activities and lacks face-to-face contact. In so doing, de Laat (ibid.) built on Brown and Duguid's (J. Brown & Duguid, 2000) previous work, whereby the close-knit concept of communities of practice (Wenger, 1998) was brought into the organisational context, ascribed to loosely coupled relationships. To emphasise the differences that interact continuously and in large groups, for instance in organisations, where individuals do not necessarily meet, the researchers coined term "networks of practice" (J. Brown & Duguid, 2000, 2001; Faraj & Johnson, 2011). In an educational context, Jones and Esnault (2004) juxtaposed the networked context to socio-constructivist understandings of a community. According to Jones (2004), a networked

context did not privilege the strong relationships that imply closeness and unity of purpose as much as a "community" would. Networked learning therefore allows for the inclusion and consideration of relationships that differ in strength; that is, both closely-knit and loosely coupled groups.

Connectivism is another approach to learning that considers sociotechnical systems. It provides for a similar understanding of the relational context that emerges through learning networks. In particular, the connectivist view on learning infrastructure can be interpreted as allowing for different ties in a learning network (Siemens, 2013). Learning infrastructure in connectivism describes core learners, as well as peripheral learners and experts. Core learners interacting with core content can be understood as linked within a more closely-knit community. Peripheral learners and incoming experts can also be linked to the same network as core learners, but in a loosely coupled way. The content of learning within learning networks is co-created by all participants. To summarise, within sociotechnical approaches, the notions of network and of networked contexts imply that a learning context is created through interpersonal interactions. Further, a network created from interactions comprises heterogeneous interpersonal relationships.

Summary. This section has demonstrated that interpersonal interactions are integral to the theories of learning. In particular, both contemporary constructivist and sociotechnical approaches emphasise that strong interpersonal relationships that may evolve between individuals over a period of time, within a shared practice, and with sustained commitment to a joint activity. The next section will highlight further how the assumptions about interpersonal interactions and relationships arising from these theoretical approaches are present within the practices of DE and online learning.

2.3.2 Interpersonal Interactions in Distance Education

The relational aspects of interpersonal interactions, as well as their role in knowledge construction processes, described in Section 2.3.1 have placed them at the heart of DE pedagogies. According to Moore (1989), three types of interactions constitute learning at distance: learner–learner, learner–teacher, and learner–content. Learner–content interactions, which are central to the process of learning, represent "the process of intellectually interacting with content that results in changes in the learner's understanding, the learner's perspective, or the cognitive structures of the learner's

mind" (p.2). Interpersonal interactions (learner–learner and learner–teacher) bridge the transactional distance. Transactional distance refers to the psychological and communication space in DE that are imposed by constraints of space and time. A dialogue between teachers and learners shortens the transactional distance. In that, interpersonal interactions have been foundational to teaching and learning at distance.

Numerous studies and meta-analyses have supported the importance of learner-learner and learner-teacher interactions in formal online learning settings, and its impact on learner achievement and satisfaction (Bernard et al., 2009; Borokhovski et al., 2012; Joksimović, Gašević, Loughin, Kovanović, & Hatala, 2015; Ravenna, Foster, & Bishop, 2012). For instance, the meta-analysis by Bernard and colleagues (2009) offered solid meta-analytic evidence that student-content and student-student interactions had a higher effect on academic performance than student-teacher interactions. Furthermore, Borokhovski and colleagues (2012) revealed that when it comes to supporting student collaboration, intentionally designed student-student interactions outperform interactions that result from contextual affordances. In addition, Ravenna and colleagues (2012) linked effective student interactions with course structure, guided discussions, as well as instructors' participation and design to foster a sense of community. Supporting this, Dawson (2006) demonstrated an association between interpersonal interactions and a sense of belonging. These findings point towards the need to design learning in ways that facilitate peer and teacher interactions and enable the benefits associated with them.

Assumptions about the evolving relational qualities of interpersonal interactions are implicit within online learning theories and models. In particular, theories and models for teaching online embed the importance of the developing the social (i.e. relational) context from interpersonal interactions that are sustained over time. The focus of well-known approaches to online learning, such as community of inquiry (Garrison & Akyol, 2013), knowledge forum (Scardamalia & Bereiter, 1996), and phases of knowledge construction (Gunawardena et al., 1997), is on helping learners reap the cognitive benefits of learning within a social context. Kreijns and colleagues (2003) emphasised that besides the focus on mastery of content, the social environment plays a significant role within formal online education models and theories. They referred to the foundational work of online education to urge for relationship building, sharing a sense of community and a common goal (Kreijns et al., 2003, p. 342) – the socio-emotional

processes help create a safe environment where learners are free to make mistakes that enable them to learn. According to Kreijns and colleagues (2003), the socio-emotional aspects of learning are those of group formation and dynamics, i.e. "processes that have to do with getting to know each other, committing to social relationships, developing trust and belonging, and building a sense of on-line community" (p.342).

In online education research, the socio-emotional dimension of interpersonal interactions is present within the constructs of social presence (Gunawardena, 1995), social ties, learning community, sense of trust, virtual or distributed community of practice, network of practice, sense of belonging and identity. In their premises of such research, these constructs refer to elements of community. Although the definitions of what community means vary, they are all expressed in terms of locality and defined by a sharing of common interests (Dawson, 2006, pp. 13–14). Community members often experience the sense of community through membership, a fulfilment of needs, influence, and shared emotional connection (McMillan & Chavis, 1986). Relationships underpinned by these experiences are often referred to as strong ties, evidenced by time spent in the relationship and its depth, or emotional intensity. Strong ties are normally characterised by high levels of trust (Krackhardt, 1998), expected reciprocation of services (Wellman, 1979), self-disclosure, multiple means of maintaining them (Haythornthwaite, 2001), and resilience, even when it seems reasonable to discontinue the relationship (Granovetter, 1973). Continuity of interactions cuts across these varying features of community.

Hrastinski's (2009) approach to online participation exemplifies how socio-emotional aspects of interpersonal interactions become embedded within online learning approaches. He theorised online learning as participation. In particular, he referred to participation as the process of "taking part and maintaining relationships with others" (p. 78), in an adaptation of Rovai's (2002) elements of online community. Rovai defined community as a "mutual interdependence among members, sense of belonging, connectedness, spirit, trust, interactivity, common expectations, shared values and goals, and overlapping histories among members" (para. 10). In building on Rovai's work, Hrastinski has also embedded these properties of participation into his own theory. Furthermore, Hrastinski's characterisation of online learning as participation is linked strongly to the interpersonal social processes that might unfold: he specifically incorporates different types of communal relations that underpin participation. They

can be both negative or conflictual, as well as positive and cooperative (Wenger, 1989, cited in Hrastinksi, 2009). Throughout his theory, Hrastinski drew parallels with the concepts of a sense of community, learning communities (Carlen & Jobring, 2005), and knowledge building communities (Scardamalia & Bereiter, 1996), which all assume that growth of interpersonal relations facilitate cognitive engagements for learning over time.

The essential role of the socio-emotional dimension in interpersonal interactions for learning has also been noted in the competences required of an online educator. The areas of an online educator's competence includes the ability to facilitate a group as a part of the teaching presence, alongside instructional design and direct instruction competencies (Anderson, Rourke, Garrison, & Archer, 2001). Facilitation surfaces as a core area of teaching competence in varying approaches to online pedagogy, along with skills in learning design and support of knowledge construction (de Laat, Lally, Lipponen, & Simons, 2007b). In contrast to learning design and assistance in understanding, facilitation addresses the formation of socio-emotional processes known to support and benefit learning processes indirectly. As a practical field, facilitation provides methods for supporting group processes and can be defined as "the design and management of structures and processes that help a group do its work and minimize the common problems people have when working together" (Justice & Jamieson, 2012, p. 5). Process skills can be realised through: (1) welcoming; (2) establishing ground rules; (3) creating a community; (4) managing communication; (5) modelling social behaviour; and (6) establishing own identity (Goodyear, Salmon, Spector, Steeples, & Tickner, 2001). These aspects of facilitation address collective-level outcomes of learning, such as the existence of democratic participation, a culture of learning and respect, and a climate of trust. If a teacher succeeds in creating such socio-emotional conditions, learners may develop interpersonal trust and belonging strengthening the potential for learning.

2.3.3 Section Summary

Section 2.3 reviewed learning theories to demonstrate that interpersonal interactions have significant yet different functions in a learning process, ranging from the correction of learner mental schemas, to triggering better understanding of subject matter, to having a richer and deeper learning experience in a safe environment.

Interpersonal interactions may also acquire relational qualities related to socio-emotional processes; these have been shown to help individuals self-regulate their studying and internalise group values. Distance education pedagogy has taken notice of these empirically supported benefits of learning by integrating interpersonal interactions into its models of learning and teaching. Socio-emotional aspects of learning in evolving social contexts are also present in socio-constructivist and socio-cultural approaches to learning; and continuity and an accumulated intensity of interpersonal interactions are implicit in the development of learning communities. Consequently, in order to help learners develop a social context conducive for learning, educators are required to be skilled in facilitating socio-emotional aspects of interpersonal interactions. This highlights the role of a particular pattern of interpersonal interactions embedded within formal educational models and theories. However, as argued in Section 2.2.4, the patterns of social interactions in MOOCs may follow different dynamics. As will be demonstrated in the next section, little insight has yet been gained into the relational qualities of the patterns that emerge from interpersonal interactions in MOOC settings.

2.4 MOOC Research on Interpersonal Interactions

Section 2.3 has shown the growing importance that learning theories and online education models assign to interpersonal interactions. Section 2.3 also addressed the theory that interpersonal relationships and that social context support and enhance individual learning outcomes (e.g. Martin & Dowson, 2009; Woo & Reeves, 2007). However, as will be further highlighted in this section, the properties of social context development in MOOC settings have not been sufficiently addressed by MOOC research. In particular, studies have examined aspects of social context in MOOCs without accounting for the specificities of this educational offering, for example diverse motivations for participation, including social learning. Consequently, the section will identify areas of research that lack evidence for the properties of the social context in MOOCs.

This section argues that the existing evidence around social context development in MOOCs is insufficient. To examine the evolution of social contexts in MOOCs, the section identifies the three major research propositions shaping this Doctoral thesis. Section 2.4.1 explains that research on interpersonal relationships in MOOC forums is

scarce. Studies have presented evidence that learners perceive the volume of interactions as overwhelming and chaotic, thereby calling into question whether meaningful student relationships can actually be formed. In this context, the present research construes MOOC forums as a form of "learning in crowds", where learners require well-developed skills to navigate (Milligan, 2015) communication, with loosely connected individuals (Gillani & Eynon, 2014) dropping in and out of discussions in a highly asynchronous manner (Yang, Sinha, Adamson, & Rosé, 2013). However, the actual potential for interpersonal relationships within such a context remains unclear, forming the first research proposition in the thesis.

The remainder of the Section 2.4 examines the limitations associated with analysing MOOC forums from the perspective of socio-constructivist theories (Section 2.4.2) and transactional, (i.e. networked) approaches to theories (Section 2.4.3). because of the lack of theorisation as to how the social context is enacted in MOOC settings, there is limited scope to interpret the findings derived from both socio-constructivist and networked approaches. Simply put, socio-constructivist studies assume that interpersonal relationships evolve between all participants in social learning without considering the intermittent activity and massive scale of forum participation in MOOCs. Therefore, the insights such studies can offer are constrained. Similarly, networked studies lack consistent theoretical justification to represent interactions as networks of learners. As a result, the findings produced through network analytical approaches are conceptually diverse, and cannot be integrated coherently into a single body of knowledge that informs the potential and nature of the context emerging from interpersonal interactions.

The chapter concludes with three research propositions forming the inquiry for the thesis. These propositions address the gaps in the research on the properties of social contexts in MOOCs, such as the potential to form interpersonal relationships through interpersonal interactions, the relational quality of interpersonal interactions, and their networked structure.

2.4.1 Can Relationships Form in MOOCs?

The class size in face-to-face settings and in DE environments has been associated with higher student outcomes and among the factors impacting the development of learning communities (Mathis, 2017; Rovai, 2002; Tomei, 2006; Vrasidas & McIsaac, 1999). In

formal DE, the recommended group size was previously noted as 17 students (Vrasidas & McIsaac, 1999), with an average class size of 22 (Orellana, 2006). In particular, the size of a learning group has been found to affect the quality of interpersonal interactions, including the comfort of learners, the degree of familiarity they can establish with each other, and their sense of privacy (Tu, 2002; Tu & McIsaac, 2002). Rovai (2002), for instance, suggested that the maximum teacher to student ratio conducive to the development of interpersonal relationships should be 1:30. The class sizes recommended for formal online education are much lower than the median MOOC student enrolment of 43,000 learners with its 6.5% completion rate (Jordan, 2014). Given the immense discrepancy in class size between a formal online course and a MOOC, the concern that learners will not be able to interact and establish relationships effectively (Miyazoe & Anderson, 2013) was legitimate.

Among the reasons contributing to the significance of class size in DE is its impact on the volume of interactions. That is, larger groups may generate high volumes of interactions that impact student perceptions. Tu (2002; Tu & McIsaac, 2002) demonstrated the critical importance of perceived familiarity among learners in a distance education class: individuals became more informal and more willing to self-disclose, when more familiar with each other, and over time, this was followed by an increased sense of trust and affectivity. Yet, at early course stages, in a distance education class, university students "became confused and frustrated, having difficulty determining 'who' was talking to 'whom' about 'what'… motivated to 'withdraw' or 'just observe'" (Knox, 2014, p. 143). Similarly, MOOC learners reported anxiety about the anonymity of participants, caused by the size of the posting group. Knox (2016) captured this confusion in the following MOOC participant statement: "the course feels like an ocean so I am making myself post this comment as a way of dipping one toe into the water" (p. 193). These experiences of being overwhelmed were attributed to both the volume of learners and the flow of the messages they generated. Forum posts may appear to be too abundant and/or too fast for the learner to react to them; or, on the contrary, not responded to, leaving the poster with a sense of being abandoned by fellow MOOC-takers and course instructors. Knox (2014) highlighted this effect through the following participant observation: "this course was a waterfall, and I only felt a few drops", interpreted by the researcher as an expression of "an additional sense of loss at being unable to interact with all the material" (p. 168). Kop

and colleagues (2011) described a similar anxiety associated with learners' inability to comprehend the talking crowd was also reflected within earlier voices around connectivist MOOCs. The authors questioned in particular whether a pedagogy of massiveness could at all allow for a learner to find a voice and establish presence. As observed through the narratives of learner experiences, the scale, volume and flow of information in MOOCs appeared to limit the evolution of interpersonal interactions.

Research into platform-based forums where MOOC learners interact with one another resonated with those learners' personal narratives of confusion. Similarly, studies of forum activity in MOOC settings suggested that scaled interpersonal interactions offered individuals limited potential to establish interpersonal relationships. For example, Gilliani and colleagues (Gillani & Eynon, 2014; Gillani, Osborne, Roberts, Eynon, & Hjorth, 2014; Gillani, Yasseri, Eynon, & Hjorth, 2014) posited that most connections formed between learners through discussion thread co-participation in MOOCs are not indicative of strong underlying relationships. The authors reached such a conclusion because the overall structure of MOOC forums in Coursera MOOCs, which they studied, was sparse and appeared to be held together by a small number of individuals. They concluded that the modularity of discussions' structures prevented a truly global dialogue between all MOOC learners (Gillani, Yasseri, et al., 2014). Furthermore, the authors stated that "these participants assemble and disperse as crowds, not communities, of learners" (Gillani & Eynon, 2014). They observed that groups appeared around short-lived discussions; that network structure was modular; and that participation was intermittent and ephemeral, with low similarity in group membership between different time points. The dynamics researchers observed in MOOC forums was the opposite of communities: "students started off with high-volume participation in on-line discussions, and over time, their apparent commitment to these conversations tailed off" (ibid., p. 23). Complementing this is a perspective by Milligan (Milligan, 2015; Milligan & Griffin, 2016) who suggested that navigating MOOC forums' scale and specificity of participation is a skill that only few individuals are competent in. Milligan (2015) contended that only a few individuals have advanced skills in learning from the social interactions taking place in MOOCs, further advancing the notion of MOOC forums as crowd-sourced learning.

Because of the sparsity of interpersonal interactions between learners, research also experimented with integrating synchronous communication into MOOC forums. One

example of such integration took the form of a collaborative dialogue tool, Bazaar, which allowed participants to connect with one another in a MOOC forum chatroom in an edX course (Rosé et al., 2015). Although researchers noted a 50% increase in attrition rate for the learners who used Bazaar, the impact of the tool on the formation of interpersonal relationships remained unclear. Contradicting evidence of the effect of a chatroom on a MOOC forum has been offered by Coetzee and colleagues (2014; Coetzee & Hearst, 2014) who experimented with a moderated chatroom within edX forums, and investigated the effects of the chatroom on grades, retention, forum participation and students' sense of community. In contrast, to Rosé et al. (2015), Coetzee and colleagues. (2014) reported no significant effect of chatroom use on any of these variables.

Research Proposition 1. MOOC research has presented MOOC forums as spaces for crowd-source learning that require competence in order to navigate learning opportunities and sparsely connected discussions. Course size and a lack of cohort boundaries cast doubt as to the capacity for interpersonal interactions to occur in a MOOC, in acquiring a quality of relatedness. Participation appears intermittent, suggesting low opportunities for familiarity between learners. Yet, only a few studies have directly examined the potential for relationship formation in MOOC forums. Given the scarcity of research on the nature of interpersonal interactions in MOOCs, there is a need to analyse the strength and types of relationships formed in such open forums. The first step towards uncovering the nature of the social context is to examine the types of participation occurring in MOOC forums, particularly in relation to affordances for relationship formation in MOOCs. In addressing this research gap, Chapter 4 explores forum log data from four MOOC offerings in order to delineate different types of forum users. Building on this work, Chapter 5 focuses on comparing the opportunities for relationship formation between the varying types of forum participants identified through the analyses undertaken in Chapter 4.

2.4.2 Are There Communities?

Socio-cultural and socio-constructivist theoretical lenses are suitable for the investigation of social contexts in MOOCs, particularly in order to observe whether or not the evolving interpersonal interactions are comparable to those in a community. For example, in formal education settings the construct of social presence (defined as a

perceived sense of somebody being present and real in an asynchronous online course) has been construed as an indication that interpersonal relationships and the formation of a group climate (Akyol & Garrison, 2008; Gunawardena, 1995; Tu, 2002). In MOOCs, however, few studies examined learner perceptions of social presence (Cheung, 2014; Kilgore & Lowenthal, 2015). Kilgore and Lowenthal (2015), for instance, examined perceptions of social presence for MOOC learners at the end of a MOOC. They reported that the learners, on average, established a below average perception of social presence, as measured through the questionnaire for the community of inquiry (Rourke, Anderson, Garrison, & Archer, 2001). Kilgore and Lowenthal (2015) concluded, however, that MOOC participants "were able to experience social presence first hand and that social presence can be established in large online courses" (p. 398). However, because researchers tended to treat all forum users as if they all started on the same date and interacted continuously, the insights from the analyses are unclear. In other words, all MOOC learners were analysed through a social presence theoretical lens that assumed a continuity of interactions between learners. Such an assumption, however, does not apply in MOOCs. Poquet et al. (2017) demonstrated that learners who have never posted in forums, as well as those who posted only once, also reported the perception of social presence. The authors showed that the scores between these groups were significantly different, skewing the average. In other words, Poquet et al.'s analysis challenged the blunt re-application of the socio-constructivist instrument measuring social presence without methodological considerations for the open participation patters in MOOC settings. Yet, neither Cheung (2014) nor Kilgore and Lowenthal (2015) offered any indication as to whether the forum participants they had analysed have had consistent presence in MOOC activity or continuity of interactions with one another; therefore, making sense out of their findings became a challenge.

Similarly, Kellogg and colleagues (2014) applied a socio-constructivist lens to examine knowledge construction levels in forums. They analysed the content of learner posts qualitatively and evaluated this in relation to the levels of knowledge construction within the network of learners. They noted that forum conversations were typically at a low level in terms of co-construction of knowledge. By analysing the content of forum discussions, they found that only 7% of all conversations went beyond the negotiation and co-construction of knowledge phases (Gunawardena et al., 1997). Kellogg and colleagues (2014) did not distinguish between learners who posted only once in a forum

and others who may have had continuous interactions throughout a given course. Yet, the theoretical lens they applied assumed that all individuals participating in the network experienced progressive interactions with one another over an extended period of time. As such, it is unclear how to appropriately interpret the low level of knowledge construction or social presence collected from a one-time drop-in forum participant. In essence, researchers had applied frameworks from more formal educational settings to learning in open environments without testing whether or not the theoretical assumptions could hold.

Research Proposition 2. The use of theoretical lenses with an embedded assumption of interaction continuity carries limited in insight when applied to MOOCs without any adjustment to the differences associated with the openness of a MOOC context. To compare the nature of interpersonal interactions in MOOCs to the context of groups formed in a formal educational experience, it is necessary to examine discussion threads shared between forum participants in MOOCs. Understanding the nature of these discussions could determine the types of group-level processes that could take place in MOOCs. Chapter 6 contributes with qualitative analyses of content driving the interactions between forum posters which provide opportunities for relationship formation. Chapter 7 further analyses selected discussions thematically, to draw interpretations about the nature of relatedness underpinning interactions.

2.4.3 What Network Structure Explains MOOC Forum Activity?

Analysis of the network structure of the interactions comprising MOOC forums has been widely used in MOOC studies (see Gašević et al., 2014; e.g. Gillani, Yasseri, et al., 2014; Hecking, Hoppe, & Harrer, 2015; Jiang, Fitzhugh, et al., 2014; Joksimović et al., 2015, among others). Such interest and the adoption of the social network analysis as an analytical tool is due to its capacity to examine large volumes of data, as well as its connection to the theories of networked learning (Haythornthwaite & de Laat, 2012). As reviewed in Section 2.3.2, approaching learning as occurring in a network is an idea that has been around in online education for some time. Pioneering work by Wellman (Hiltz & Wellman, 1997; Wellman, 1997, 2001; Wellman et al., 2003; Wellman & Gulia, 1999) and Haythornthwaite (Haythornthwaite, 1996, 1999, 2000, 2001, 2002, 2005; Haythornthwaite & de Laat, 2012) extended the application of social network analysis to online contexts and e-learning settings. These researchers demonstrated that

networks mediated by technology can maintain relational qualities that are associated with interpersonal and community relations in face-to-face settings.

In contemporary approaches to learning, the term "network" can be used both in reference to the quality of social relations and simply to represent group structure. On the one hand, as detailed in Section 2.3.6, a networked context implies that the relationships constituting it can be of different strengths. That is, a networked context evolves both from the relationships described by the unity of purpose or interpersonal bonds and from the weaker connections generating information flow in the network. At the same time, the network perspective can be used simply to represent the structure of relations (not relationships) in any social entity. Analysis of the structure of interpersonal relations can explain the power dynamics or roles taken up by individuals in a group (Poole, Hollingshead, McGrath, Moreland, & Rohrbaugh, 2004). Such an approach has been used in the online education domain and in the area of computer-supported collaborative learning, where small groups have been analysed through a network lens (Goggins, Galyen, & Laffey, 2010; Suthers, 2015). The differences between the distinct uses of the term "networks" were well-captured by Wenger-Traynor and Wenger-Traynor (2011). They wrote that, from the perspective of social learning, community and network are simply different aspects of social structure: the network focus is on information flow, whereas the community focus is on the establishment of identity through a shared domain of practice. Accordingly, all communities can be networks (in reference to the structural representation of a community), but not all networks can be communities (in reference to relational quality of interactions typical for communities).

In MOOC research, analysis of MOOC networks has also been conducted by approaching networks as representations of social relations, without a particular reference as to the quality of the ties. An analysis of learner networks in MOOCs has been conducted using social network analysis because this methodology lends itself to empirical observations within the theoretical premises of the sociotechnical approaches to learning. Social network analysis has become applied extensively in analyses of computer-mediated communications (CMC). Haythornthwaite and colleagues (Haythornthwaite, 1996, 2005, Haythornthwaite & de Laat, 2010, 2012; Haythornthwaite, de Laat, & Schreurs, 2016) recommended using it to analyse learning networks in e-learning settings. The set of methods offered within social network

analyses was adopted widely in the analysis of interpersonal interactions in MOOC forums. Its popularity can be explained in part by its capacity to simplify patterns of forum communication that occur on an incomprehensible scale in MOOC settings.

Unfortunately, the diversity in how researchers are conceptualising information and communication flow as network structures, impedes synthesising findings about collective learning activity in open online education settings. De Laat and Schreurs (2016) cautioned, "SNA [social network analysis] is a very flexible method, but it requires a solid theoretical framework to enable the interpretation of findings", warning that otherwise it may result in flawed conclusions (p. 266). This is exemplified in analyses of network clusters in MOOCs. Clusters within networks of interpersonal interactions are identified by algorithmic detections. These clusters are then labelled as 'communities', the term used for densely connected clusters in graph theory. For example, Brown and colleagues (R. Brown, Lynch, Eagle, et al., 2015; R. Brown, Lynch, Wang, et al., 2015) detected communities based on sub-graph detection algorithms that identify learners who have communicated with one another more than with individuals outside a partitioned group. However, their study does not elucidate on the type of connections underpinning these sub-graphs. The problem arises when the authors continued interpreting these clusters in social scientific terms: the study posited that "membership" in these "communities" is correlated with course performance. From the social scientific perspective, the justification of both membership and communities requires further evidence, limiting the value of these findings.

Other social network analysis research of MOOC forums addresses individual-level outcomes and the structure of MOOC networks. The difficulty in making use of these existing analyses is in the lack of theoretical framing that shapes what is being represented by the network structure, and why. In social network research into MOOCs, approaches differ in how to interpret the post-reply structures. For example, researchers examining Twitter networks in MOOCs analysed communication represented through person-to-person information flow (Gruzd & Haythornthwaite, 2013; Skrypnyk, Joksimović, Kovanović, Gašević, & Dawson, 2014). Twitter communication interactions are possibly the least ambiguous to interpret as they are, by design, identical with the "name networks" (i.e. a network tie is defined by actor A explicitly referring to actor B by name) (Gruzd & Haythornthwaite, 2008) where there is less doubt as to 'who spoke to whom'.

Similarly, post-reply communication structures appear to have been replicated in studies of platform-based MOOCs delivered via edX and Coursera. However, the approaches adopted in order to analyse the forum data differ. For instance, Joksimović et al., (2016) used a conventional directed post-reply structure, commonly used in online education forums: "if author A2 replied to a message posted by author A1, we would add a directed edge A2→A1. Further, if A3 posted a comment on A2's post, we would include A3→A2 edge as well" (p. 318). In contrast, Brown and colleagues (R. Brown, Lynch, Eagle, et al., 2015; R. Brown, Lynch, Wang, et al., 2015) used a post-reply network structure based on communication going from one person to many prior posters. Gillani and Eynon (2014) adopted another alternate approach in a Coursera MOOC, representing communication through an edge that connected "two learners simply if they co-posted in at least one discussion thread". Because there is no consistency in the way researchers conceptualised communication structures, it is literally impossible to synthesise interpretations of graph analytical metrics employed in the social network analysis of MOOCs; the researchers measure what may appear as conceptually different types/structure of interactions.

Research Proposition 3. Networked approaches to learning are contemporary approaches to understanding learning in sociotechnical systems; however, to make use of quantitative methodologies for structural analyses of networks in MOOCs, social network analysis needs to be applied to theoretically defined entities. Because of the lack of explanation of the social context in MOOCs, network analyses of open courses have been largely atheoretical. In order to offer a theoretically grounded analysis of the network structure, Chapter 8 presents a study that employs the statistical network modelling of a sub-group of forum users conceptualised as an identity-based community that emerged in a MOOC forum. The theoretical framing is derived through the empirical analyses in Chapters 5-7, and informs the ways in which a network is constructed and modelled.

2.4.4 Section Summary

The size and scale of interpersonal interactions in MOOCs have raised doubts about the overall effectiveness of learning at scale. Reported narratives of learner experiences further support these concerns. They have reported feelings of fear, confusion and being overwhelmed by the sheer volume and flow of information (Cao, 2014; Filvà,

Guerrero, & Forment, 2014; Knox, 2014; Kop et al., 2011; Veletsianos, 2013). Yet, despite the impact of interpersonal interactions on learner perceptions, research has not directly addressed the question of whether interpersonal relationships can form in MOOC settings. Research on interpersonal interactions has not sufficiently explained, described or captured the evolution of social contexts in MOOC settings. Without an explicit theoretically driven examination of the social context, studies have concluded that MOOC forums are akin to a crowd of learners (Gillani & Eynon, 2014). The term "community" has been applied, devoid of its conceptual roots, most notably in relation to algorithmically derived sub-graphs representing structural elements of forum communication rather than the quality of relatedness underpinning them (R. Brown, Lynch, Wang, et al., 2015). Even when studies on MOOCs have employed a theoretical lens that appropriately incorporates the social context, the studies have lacked adjustments in the research design that would satisfy comparability issues between the formal and open (informal) contexts. Finally, the most promising strand of research on MOOC forums, which is based on the social network analysis of interpersonal interactions, does not use a theoretical premise for its structural representations. In sum, the social context that emerges from socio-emotional processes underpinning interpersonal interactions in open online scaled courses is yet to be adequately explained. This critique of the existing research highlights its current limitations. Such a critique offers a strong rationale for undertaking this Doctoral thesis.

Based on the limiting insights offered by MOOC research to date, this section put forward three propositions for this thesis' research agenda. First, to gain insights into the nature and evolution of a social context, the MOOCs' potential for dyadic interpersonal interactions needs to be understood. Second, in order to evaluate this potential in relation to the socio-emotional processes that may underpin it, the nature of discussions driving these interpersonal interactions needs to be examined. Based on the insights, a theoretical hypothesis about the nature and dynamics of MOOC forums can be made. Third, social network analysis can be used to test if the formulated hypothesis can explain the nature and evolution of the networks observed in MOOC forums. Chapter 3 presents methodological considerations relevant to the implementation of these research propositions, and also introduces the courses analysed in the thesis.

2.5 Chapter Summary

The literature review explained how MOOCs appeared, driven by the opportunities introduced by the technological advances and by the market forces propelling the roll-out of online education within the higher education sector. The chapter highlighted the limitations of existing MOOC research. In reacting to strong media critique of MOOCs, researchers have tended to address the perceived challenges in a manner that is more representative of formal university courses. As a result, MOOCs have been examined under the same theoretical and methodological approaches that have been applied to formal (structured) educational contexts. This chapter has presented a suite of empirical evidence that suggests that MOOC settings are different from those in formal education, and are a composite of a complex pattern of interpersonal interactions. Because of the reliance of formal educational theories and models on the somewhat predictable pattern of interpersonal interactions in formal education contexts, the chapter questions the applicability those models to the emergent social context in MOOCs. Without insight into the nature and evolution of the social context, a re-application of the current analytical theories cannot explain or contribute to improving experiences in MOOCs.

By reviewing existing studies of MOOC forums relevant to understanding social context development, the chapter delineated several research gaps: (1) the lack of understanding about opportunities for relationship formation; (2) a lack in the comparability of interpersonal interactions to the qualities of community-like structures; and (3) a lack in understandings of the network structure describing the social context. This thesis comprises of a series of empirical studies contributing to each of these research gaps. The rationale intrinsic to its methodology, framed within the domain of learning analytics, is explained further in Chapter 3.

3 Methodology

The previous chapters explained that current research and practice on platform-based MOOCs has been strongly influenced by prior work in formal educational settings. To date, research has largely focused on issues topical to higher education, such as student retention and completion. As such, the literature review presented an argument outlining the lack of research on the social context of interpersonal interactions in MOOCs. However, understanding the nature and development of interpersonal interactions in a MOOC-like setting can bring new insights into how students form relationships in both formal and informal contexts. The lifespan of learning and the ongoing evolution of higher education models will bring a further softening between such formal and informal education boundaries. Students will seek to increasingly integrate formal learning opportunities with those undertaken in non-formal and informal settings such as MOOCs, the workplace, and professional development and training.

This study affords new insights into student learning as MOOCs integrate the structured process of course design with an unstructured process of open scaled participation. The constant influx and change of participants disrupts the expected evolution of interpersonal relationships. The relational quality of interactions, i.e. an emergent emotional connection underpinning participation, is needed to shorten the distance between the learners in an online setting. Bridging psychological distance through interpersonal dialogue has been fundamental to online education pedagogy. Although the social context is an essential part of online pedagogy, research lacks explanations about the patterns and quality of interpersonal interactions in MOOCs. Consequently, the aim of the thesis is to contribute to the narrative about the social context in open online educational settings.

To gain insights into the social context in MOOCs, the thesis describes, explains and measures the patterns and quality of group-level interpersonal interactions in platform-based MOOCs. To fulfil these objectives, the literature review highlighted several research gaps. First, to understand the very potential for interpersonal relatedness, the study examines the opportunities for interpersonal relationships between forum posters of different types. Distinguishing between different types of posters is a necessity,

given the varying motivations of learners who join MOOC forums. Secondly, the study enquires into the nature of exchanges taking place between the posters who have more opportunities to develop relationships. Investigating and understanding the nature of exchanges between them can help draw parallels between the dynamics of social interactions which are also observed within groups formed in formal online education, and also help address the larger question of whether communities can be formed at scale through more frequent interpersonal interactions. Finally, given the lack of theorisation in networked approaches to MOOC analyses, the thesis analyses the network structures that quantitatively capture the social forms emergent from more frequent interpersonal exchanges. Being able to draw these indicators of network structure can help researchers and practitioners get faster access to the qualitative insights associated with this particular social form. These three research propositions, inquiring if relationships can form, if communities emerge, and if a network structure can capture them, form the foundation for empirical work within this thesis.

The research agenda of the thesis is achieved through a series of interdependent studies. This chapter describes the overarching research methodology framing the studies. The chapter introduces the thesis design, presents the data, and addresses the philosophical orientations behind this inquiry. A brief overview of the methods specific to the individual empirical studies is also provided. Further detailed discussions for each study can be found in the corresponding chapters. The first section (Section 3.1) positions the thesis in the learning analytics domain. Learning analytics is an interdisciplinary area of research driven by the need for practical, actionable research about learning and learning environments. The need for practicality of the research necessitated the adoption of the pragmatist research orientation and mixed method research design. Within pragmatism, a researcher moves between quantitative and qualitative analyses of the data, depending on the specific outcome of interest. Section 3.1 also explains the premises of pragmatism and why mixed method research design is a good fit for learning analytics studies. Specifically, learning analytics inquiries have been criticised for measuring what is easy to capture, rather than what is of value . A combination of data-driven and theory-driven quantitative and qualitative approaches allows for the exploration of the data, gradually narrowing the analyses towards the phenomenon of interest. To map out these relationships between the studies, Section 3.2 presents the research questions, and Section 3.3 demonstrates how mixed method

research design balances data-driven and theory-driven inquiries, addressed in the research. Section 3.4 introduces the four MOOCs examined in the thesis; courses delivered by the Delft University of Technology, 2013–2014, using the edX platform. Section 3.5 describes the associated ethical protocols for the project.

3.1 Researcher's Assumptions

3.1.1 Research Domain: Learning Analytics

This thesis is situated within the research domain of learning analytics (LA), a field which has gained momentum amidst the increasing adoption of online education and associated technologies (Siemens & Long, 2011). Technologies such as Learning Management Systems/Virtual Learning Environments (LMS/VLE) provide institutionally protected online spaces for learning, and also capture digital traces of learner activity. The extractions and analyses of these data have afforded new opportunities to link student recorded activity with observed learning outcomes (Macfadyen & Dawson, 2010). Traces of learner activity have been linked to learner performance and demographics in order to deepen the understanding of learning, and provide support to learners-at-risk (Arnold & Pistilli, 2012; Jayaprakash, Moody, Lauría, Regan, & Baron, 2014). Such research possibilities gave rise to the research agenda backing LA, which has the goal of measuring and analysing information about learning, learners and learning environments (Siemens & Baker, 2012). Although early LA studies focused on predictions of students at-risk of dropping out of the university studies, Gašević, Dawson and Siemens (2015) stressed that the focus of LA is on capturing and understanding learning processes. Essentially, LA links captured data with the constructs of interest associated with learning. The ability to communicate these links to interested individuals, who then act upon them, is an essential part of the LA cycle (Clow, 2013a). In short, insights driven from LA research provide feedback to both the theoretical knowledge-base and practical decision-making.

Its positioning within a socio-cultural system of stakeholders (Suthers & Verbert, 2013) and its "actionable" aspect for research findings (Greller & Drachsler, 2012) distinguish LA from its neighbouring fields, such as educational data mining (EDM). In contrast to LA, EDM is data-driven and focused on the decontextualised identification of patterns in data in relation to a mined construct of interest. In terms of the applied methods, EDM and LA overlap (R. Baker & Inventado, 2014). The methods developed within

EDM enrich LA research, and vice versa. The application of the variety of methods, often of a quantitative orientation, and then linking them to theoretical concepts requires a broad suit of analytical tools. To highlight this specificity of LA, Gašević, Dawson and Siemens (2015) described it as a "bricolage" field, that is, to gain insight into a range of phenomena, LA researchers use distinct and epistemologically varied disciplines, such as statistics, computer science, data science, natural language processing, social network analysis, as well as social science, psychology and education. Reconciling the underlying differences between the epistemological approaches behind LA methods has been an ongoing challenge for the discipline (Wise & Shaffer, 2015).

Disciplinary theories are often used to focus research efforts, and avoid the so-called "streetlight effect" in quantitative research (Ochoa & Worsley, 2016). Freedman (2010) summarised the problem as follows: "It is often extremely difficult or even impossible to cleanly measure what is really important, so scientists instead cleanly measure what they can, hoping it turns out to be relevant" (2010, p. 1). Similarly, Biesta (2009) expressed that when quantifying and measuring data collected from educational contexts one risks to fall victim to the so-called technical validity of measurements in education (p.2). Biesta argued that to ensure that "we measure what we intend to measure", value judgments about what is educationally desirable have to be put forward (2009, p.1). Consequently, measurement of elements related to learning need to take place in relation to the theoretical concept of value rather than what can be measured, or quantified.

To identify "where to cast the fishing nets", Wise and Schaffer (Wise & Shaffer, 2015, p. 5) urged for the use of theory-driven approaches and understandings of the situational context. The need for an increased contextual understanding supported through LA approaches is well illustrated by Gašević et al. (2016). The authors made an empirical case for the importance of situational context by showing how course data aggregated across an institution covered-up deficiencies in a predictive model for academic success. In essence, although the developed model had a high degree of accuracy across the institution it examined, its level of accuracy and hence relevance, was more variable and diminished at a more granular course level. The authors argued that the pedagogical context was an important ingredient for the development of future models and associated early warning systems targeted predicting student academic

success. A theory-driven inquiry would also aid in the generalisation of results to other contexts and populations, as well as assist researchers in identifying variables to include, what potential pitfalls to account for, what results to attend to, and importantly, how results should be interpreted (Wise & Shaffer, 2015). If a researcher cannot formulate a theory-driven inquiry, Wise and Schaeffer suggested that empirical exploratory analyses be used to formulate testable hypotheses and explanations of phenomena in learning.

With the consideration of recommended approaches, the thesis balances data-driven and theory-driven inquiry within its research design. First, the thesis included a series of successive studies that contributed to the formulation of a hypothesis. Chapters 4–7 adopted exploratory and theory-informed distinct methodologies to inform an understanding of the social context in MOOC forums. Second, once the hypothesis about the social context in MOOCs was formulated, a confirmatory part of the analysis was conducted (Chapter 8). Such research design demonstrates the close relationship between the inductive and deductive reasoning in LA research. The exploratory nature of the study, as well as the size of data in MOOC research necessitates the integration of inductive exploratory approaches at the earlier research stages. These were further supplemented with deductive inquiries of both quantitative and qualitative nature. This research design resulted from the practical need to thoroughly explore the data through all possible ways prior to constraining it to the theoretical lenses. This practical need could not have been reached without a mixed design approach. The methods used throughout the thesis, however, fell unto the different domains of what it means 'to know'. To address the discrepancies between the methodological epistemologies, and justify chosen research design, the following section elaborates on the pragmatist orientation adopted for this Doctoral thesis.

3.1.2 Research Orientation: Deweyan Pragmatism

This project is rooted in the pragmatist views of educational research and learning. Pragmatism strongly resonates with LA's agenda that seeks to collect, measure and analyse data about learners and their contexts, with the ultimate goal of establishing a feedback loop for practitioners of education. Similar to LA, pragmatism focuses on solving practical problems by drawing on a variety of tools. A focus on problem solving does not limit pragmatic research to a particular philosophy (Creswell, 2009),

and is guided by an overall assumption that qualitative and quantitative approaches combined will provide a deeper understanding of a research project than either approach alone (Creswell & Clark, 2011). Within pragmatism, a researcher can move between quantitative and qualitative analyses of the data, depending on the specific outcome of interest.

In particular, the study builds on John Dewey's pragmatist ideas about knowledge and knowing. In his philosophy, Dewey went beyond traditional oppositions of objectivism and subjectivism, addressing how the mind acquires knowledge of the world outside of it (Biesta & Burbules, 2003). Philosophers have opposing opinions as to whether the source of knowledge is within the subjective experience or within the objective world. The positivist or interpretivist worldviews then privilege particular methodologies. Although situated within the interpretivist paradigm, Deweyan pragmatism viewed the immaterial mind and the material matter as co-existing (ibid., p.10). The implication of such an epistemology is that a socially existing idea is not a match, or representation of an observable phenomenon, but that the two exist concurrently, and can be linked. Dewey proposed that the elements of human nature and the environment, the natural and the social co-exist and are in a "transaction", i.e. dynamic interdependency (Dewey & Bentley, 1949). An outcome from Dewey's transactional approach (Dewey & Bentley, 1949) is based on the realist assumptions that the construction of a human mind is located in this transaction (p.11). On the other hand, this transactional approach is constructivist, as we construct our own individual world. In other words, Dewey's view of knowing is that "it is both a construction and is based on reality" (p.12). Due to the constant transaction between what is and what humans construct of it, the world is constantly transformed through the process of communication. Consequently, knowing is always fallible, as the linkage between the observed and the socially-existing changes. Hence, we can't be certain if the patterns of action developed in the past will be appropriate for the problems to encounter (Dewey & Bentley, 1949, p. 13). Such socio-cultural views of knowing are linked to the positivist observations of the world, in service of a practical solution at-hand, and are well suited to the LA research context.

The duality of Dewey's nature of knowing and the changing form of knowledge poses an important implication for LA research. In the previous section, LA is referred to as a research area 'linking observed indicators with the constructs of interest'. The wording was intentionally chosen to signal that observable indicators of behaviour are not

necessarily identical to the construct of interest. When a researcher interprets observable behaviour, an analytical connection is made between the natural observation and the abstract idea. That connection is made by the researcher between what can be captured externally with an idea that exists in a given time period, consistent with the existing scientific thinking about that particular phenomenon. This linkage, as well as the idea can be critiqued, challenged and questioned. The focus on knowledge creation in LA as making the links between the natural and the abstract allows for both the empirical examination of the material phenomenon, as well as for interpretivist (e.g. what is the meaning), critical realist (e.g. why is this phenomenon this way) or post-structural (e.g. who is excluded from the meaning making process) inquiry into both the ideas and the constructed associations. By making the assumptions that link the natural with the abstract explicit, the research community can address, critique and challenge the explicit logic of the inquiry. Within a pragmatist view, enabling a dialogue around the links is as much a contribution to further knowledge creation around the observed phenomenon, as the empirical insights the research has obtained.

In summary, the study roots itself in a pragmatist orientation to research. In line with pragmatist reasoning (Morgan, 2007), this study acknowledges that both the framing of the problem and the research inquiry are embedded within the social, historical, political, disciplinary and other contexts. The study describes the observable activity on MOOC forums as the natural, realist part of knowing, and links the observable activity with existing thoughts and ideas from past lines of research. These links and ideas can be critiqued by the examination of the strategies of action informed by them, i.e. the link could be of no use to solving the problem. In line with the pragmatist philosophy, such an outcome of research would advance the body of knowledge.

3.2 Research Questions

The thesis aims to develop and expand our understanding of learning in open online education. The literature review demonstrated that the transfer of existing ways of thinking and evaluating to MOOC settings requires further empirical validation. That is, MOOCs differ from formal courses due to their non-formal nature, scale and open participation. The interpersonal interactions emergent in MOOCs can be seen to vary from expected patterns and quality of relatedness in formal settings. The potential to establish such quality of relatedness in MOOCs has been challenged (Daniel, 2012; Kop

et al., 2011). Yet, theories and models of learning and instruction imply a particular social context, emergent in educational settings (Kreijns et al., 2003; Kreijns, Kirschner, & Vermeulen, 2013). Therefore, this thesis has argued for the need to understand the nature and patterns of interpersonal interactions specifically in relation to the social context in MOOCs.

Hence, the thesis poses an overarching research question:

What are the properties of social context development in MOOC settings, and what indicators can capture it?

To address the overarching research question, the literature review (Chapter 2) highlighted three areas in need of empirical investigation, namely if interpersonal relationships could form and evolve into the communities, as well as what network structure would describe them. These three research areas provide structure to the sequential empirical studies of the thesis (Section 2.4). The sequential studies therefore inquire: if relationships can form, if groups in MOOC forums can evolve into communities, and if a particular network structure can describe the formation of communities.

Study 1 (Chapter 5). Due to the discrepancy between formal and open contexts, the thesis examines if relationships can form in MOOC settings. To tackle this research proposition, an exploratory data analysis (Chapter 4) delineates several types of forum users. Chapter 4 singles out several groups of posters that may have opportunities for relationship formation in MOOCs. Study 1 (Chapter 5) consequently compares the frequency of co-occurrence between different types of posters in order to address the first research sub-question:

What types of forum participation are associated with opportunities for relationship formation?

Study 2 (Chapter 6). The second research proposition questions if groups emerging from interpersonal interactions can evolve into communities. Study 2 therefore, analyses how posters from a MOOC's forum core use their participation for relationship formation. For this purpose, the content underpinning the interactions between these posters is investigated. The research question was framed as follows:

What is the role of on-task and off-task communication in the formation of a forum core?

Study 3 (Chapter 7). Upon the examination of the content underpinning the formation of a forum core, the study concluded that in some of the analysed MOOC courses, patterns of interactions could be interpreted as community development. Study 3 (Chapter 7) applied the community of practice theory (Wenger, 1998) to better understand the mechanisms behind the formation of communities. Hence the following research question was posed:

What were the forum core posters talking about within social-task, social non-task and mixed discussion threads?

Study 4 (Chapter 8). Interpretations of the discussion threads containing traces of socio-emotional processes illuminated the mechanisms of community development in a MOOC context. As a result, Study 4 (Chapter 8) addresses the third research proposition. The analyses examine if certain network forms (i.e. configurations), can represent underlying social processes of community development in MOOCs. Based on the qualitative insights from Chapters 6–7, Study 4 models a forum core's networks in order to explore the following research question:

To what extent are the participatory patterns observed in a forum core's networks typical of identity-based communities?

3.3 Design and Methods

Altogether, the empirical studies in this research describe, explain and capture the social context emerging from interpersonal interactions in MOOCs. For this purpose, the research employed a mixed methods sequential design (Figure 3.1). Figure 3.1 schematically presents the phases of inquiry: inductive and then deductive. The inductive phase consisted of Studies 1 and 2 that described and explained the social dynamics in a non-formal course. The deductive phase included Studies 3 and 4, theoretically driven and interpreted in relation to the findings from the previous two studies.

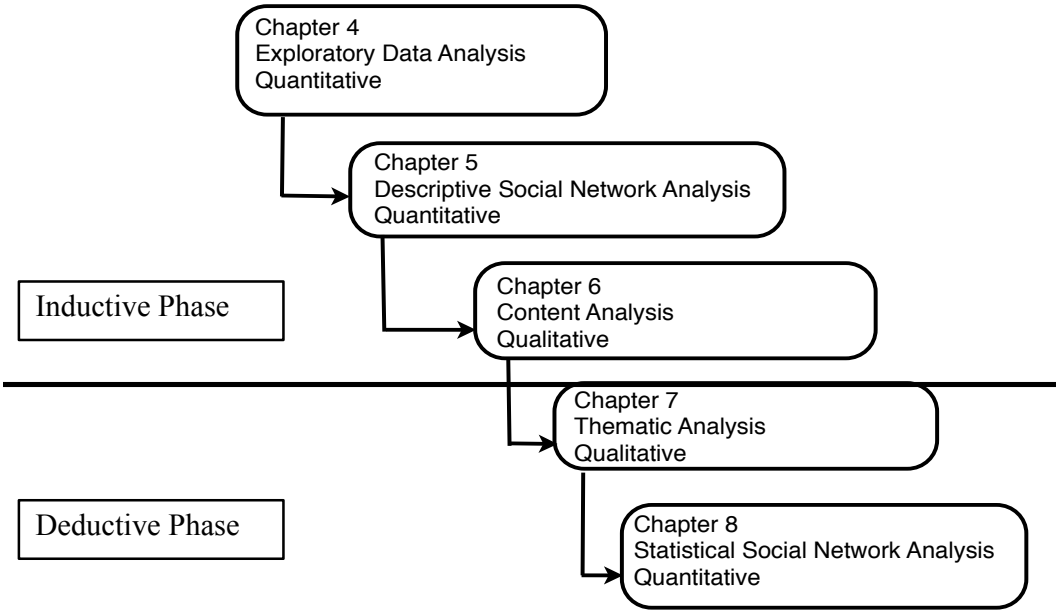


Figure 3.1 Schematic representation of the mixed method sequential design of the research.

Such a combination of inductive and deductive approaches required a mixed method design. Employing a combination of methods aligned with the adoption of a pragmatist research orientation that is also equally concerned with solving a problem, as much as it addresses research applicability. All analyses were framed through an initial data exploration, conducted at the level of all forum interactions (see Chapter 4). The initial exploratory data analysis delineated different indicators of student commitment to social activity. The study analysed posting and viewing logs for four groups of forum users, ranging 6507–14175 users per course (Table 3.1). The results helped understand the differences between user behaviours, and provided a rationale for delineating students into sub-groups that are compared in Study 1 (Chapter 5). Study 1 adopted a quantitative approach to analysing the frequency of co-occurrences across the different sub-groups identified in the exploratory data analysis. A social network analytical approach was used to compare co-occurrences within sub-groups of posters in four courses. The network size in the courses ranged from 1027 to 2287 posters (Table 3.1). A group with the highest co-occurrence between the posters was linked with having the highest opportunities for relationship formation. Study 2 (Chapter 6) and Study 3 (Chapter 7) involved the qualitative analysis of the discourse underpinning the network, which demonstrated the highest potential for co-occurrences (as identified in Study 1). As shown in Table 3.1, the number of discussion threads analysed in Chapter 6 ranged

from 889 to 1167 per course. The discussion threads were selected if they included contributions from at least two posters of a delineated sub-group. Chapter 7 (Study 3) focused its interpretative analyses on an even smaller subset of discussions, ranging from 91 to 357, depending on the course. In sum, analyses in Study 1, Study 2 and Study 3 narrowed down the data used to interpret the nature of social context development, as undertaken in Study 4. Based on the interpretation, a deductive hypothesis-driven analysis of the same sub-group was conducted at a network level in Study 4 (Chapter 8). Chapter 8 modelled four forum core networks, ranging in size from 161 to 341 users. The purpose of the analyses was to measure the elements of interest theorised upon the inductive phase of the thesis.

Table 3.1 Overview of data sources and methods

Empirical studies	Data	Method
Framing the analyses	Log data: 6507–14175 users	Exploratory data analysis
Study 1	N of posters: 1027–2287	Analysis of posters' co-occurrence networks
Study 2	N of discussions: 889–1167 N of posts: 3784–4281	Content analysis of discussion threads shared by forum core
Study 3	N of socio-emotional discussions: 91–357	Thematic analysis of the discussions of a socio-emotional nature
Study 4	N of forum core members: 161–341	Statistical social network analysis of forum core networks

3.4 Description of Analysed Courses

The four MOOCs examined comprised courses in engineering, computer science and data analysis. All MOOCs were delivered via the edX platform from the Delft University of Technology (TUD) in 2013-2014. MOOCs offered through TUD were a logical continuation of the University's prior involvement with open educational resources. The first two MOOCs offered by TUD were in Solar Energy and Water Treatment. These open courses ran in the autumn of 2013. As shown in Table 3.2, Water and Solar were both large courses with high enrolment numbers over 30,000 and almost 60,000 participants, respectively. They also had a large number of forum users:

3984 posters producing 15796 posts in Water, and 5949 posters producing 17537 posts in Solar. The other two MOOCs ran in the beginning of 2014. The MOOC in Functional Programming (FP), attracted 38029 registrants, and 1027 posters produced 7168 posts. The Excel MOOC focused on teaching the use of excel for data analyses. This MOOC had 33516 registrants, and 2287 posters produced 8348 posts (Table 3.2). All four MOOCs were representative of a typical DelftX MOOC – a platform-based STEM course with videos, quizzes and exams, with little or no design for social learning.

The researcher worked together with a TUD team on the evaluation of the first two TUD MOOCs. Based on a pre-existing working relationship with TUD, the researcher requested TUD data for the analyses of MOOC forums. DelftX staff picked courses that appeared comparable to other MOOCs ran by the institution, and also had high enrolment numbers and high forum activity.

Table 3.2 Overview of the courses

Course ID	Dates	Students		Duration, weeks	Posts, N	Posters, N
		Enrolled	Certified			
FP101x	10.15.2014–12.10.2014	38029	1968	8	7168	1027
ET3034Tux	16.09.13 – 8.11.13	57091	2730	8	17537	5949
CTB3365x	16.09.13 –25.11.13	34877	546	10	15796	3984
EX101x	07.04.2015–19.05.2015	33516	2191	8	8348	2287

The four MOOCs analysed in this thesis involved conventional video-based instruction that was facilitated mainly through in-course discussion forums. MOOC offerings were assessed via close-ended assignments such as multiple-choice quizzes, or submit-your-answer quizzes. As outlined in Table 3.2, the four MOOCs ran throughout a course of similar duration of eight weeks (except for Water course that ran for ten weeks). Course design was also similar across the MOOCs.

In regards to the discussion forums, the courses employed different strategies regarding moderation. Of the four MOOCs, FP101x (FP) had the highest staff and instructor-forum involvement. MOOC EX101x (Excel) also had staff, teaching assistants and instructors moderate its forum. Instructors in the FP and Excel MOOC forums demonstrated high teaching presence by regularly engaging in discussions with the

students. ET3034Tux (Solar) was moderated by a group of volunteer student assistants who were also students in the course, elected in the first weeks of the MOOC. Finally, CTB3365x (Water) was not moderated.

Similarly, the positioning of forums within the courseware slightly differed across the courses. For instance, Solar's forums were embedded under each video segment, quiz or test, but a learner could also access them all in a central forum, through a link on the course's page.

The following sections provide further details regarding the investigated MOOC offerings. The references to student demographics below have been obtained through institutional reports about the courses.

3.4.1 FP101x – Functional Programming MOOC

Introduction to Functional Programming, referred to as FP, was delivered by the Faculty of Computer Science. The eight-week course was based on the foundations of functional programing using the Haskell programing language. According to the course syllabus, the course assumed no prior knowledge of functional programing but at least one year of programing experience in a regular programing language, e.g. java, .NET, Javascript or PHP.

The instructor expected students to study in the course for around 4-6 hours per week. To receive a certificate of mastery, students were expected to earn the minimum overall grade of 60/100. The total grade was derived from 11 homework and seven lab assignments.

The course content was delivered via weekly video recordings. Some course readings were suggested; however, these readings were not prescribed – they were for student exploration only. The course's forum was the primary communication tool in the course, announced on the course's home page. The instructor was also active on Twitter, and due to the programing nature of the course, students also used Github. Five staff members assisted the instructor in forum facilitation, and all of them, including the instructor, were active in the forum.

The pedagogical design of the course replicated conventional STEM course. The weekly lectures required engagement with assessed quiz and a programming

assignment. The teacher and assistants offered a recorded tutorial that prepared for the assessment. No other pedagogical provisions were in place.

3.4.2 ET3034Tux – Solar Energy MOOC

The Solar Energy course, referred to as Solar, was one of the two first MOOCs TUD offered. The course was designed as a bachelor's level course and requirements stated that students must have basic-level knowledge of physics and mathematical skills such as competence in integration and differentiation. The advertised course workload for students was anticipated to be 8 hours per week. The course attracted much interest due to its practical orientation as the course also allowed for students to experiment with making their own pv-systems, i.e. solar panels.

To receive a certificate of completion, a student was required to achieve at least 53% of the final grade. The grading was a result of five assignments comprising 40% of the grade and 3 exams representing 60% of the grade (De Vries, Hennis, & Skrypnyk, 2013).

Course forums were attended to by six tutoring staff in the first two weeks, without heavy instructor involvement. While the course was in progression, four students were nominated to support forum activity as a form of community assistance. These co-called community teaching assistants (CTAs) were active throughout the full duration of the course. Thus, the primary means of forum moderation came from two staff members and the CTAs.

The Solar course followed a conventional pedagogical design observed in xMOOCs: weekly video lectures were supplemented with unassessed quizzes, and occasional assessed assignments and exams. The teacher had not engaged with the students in any way beyond offering the content. The course however included several assignments that had practical orientation. That is, the learners did not only engage in the theoretical concepts around solar power, but also built their own solar panels, posted pictures of them from their environments and shared their experiences around using solar energy.

3.4.3 CTB3365x – Water Treatment MOOC

The Introduction to Water Treatment MOOC, referred to as Water, was among the first MOOC offerings at TUD and taught by the Faculty of Civil Engineering. This MOOC

was described as an introductory course to water treatment technologies and was intended for students interested in sanitary engineering. It was advertised at the bachelor level and did not require pre-requisites. The course design included two long modules of four weeks. The modules were dedicated to drinking water and waste water. The course ran for ten weeks, with an expected workload of 10-12 hours per week for students.

The course assessment allocated 80% of the final grade on two exams for each of the modules. To pass the MOOC, learners needed to earn a minimum grade of 60%. Aside from the exams, the course also included ten assessed homework assignments, and over 150 self-check unassessed quizzes.

Four instructors were involved in the Water MOOC and, in contrast to the other MOOCs included in this research; there was no moderation of discussion forum activity. To communicate back to the students, instructors produced a weekly 5-minute video that summarised the feedback on some of the questions raised in the discussion forum.

3.4.4 EX101x – Taking Data to the Max MOOC (with Excel)

The final MOOC was an eight-week course on data analysis in Excel, referred to as the Excel MOOC. Although this course assumed some prior experience with working with spreadsheets the course was designed in a way that allowed beginners to also engage and work with the implemented learning activities. The MS Excel 2013 software was required for the course with additional software, introduced half-way into the course. The expected student workload was 4-6 hours per week. Students could earn a certificate of completion, if they achieved an overall score of at least 60% in the weekly exercises. The instructor and five supporting staff members, including student assistants, were involved in forum facilitation, along with three community assistants elected from the student body.

Water course followed the same pedagogical design as the other MOOCs analysed in the study. It included a weekly lecture, followed by unassessed quizzes, and major exams.

The design of the excel course revolved around practical tasks designed by the instructor. The instructor and the assisting team was actively responding to students'

posts both related to the assignments and simple information sharing. The information presented by the instructor to the students carried tutorial-style videos where the instructor demonstrated different data manipulations and analysis. The tasks were not theoretical but required solving particular problems with data. Besides its heavy reliance on the technical tool (MS Excel), overall pedagogical design of this MOOC resembled that of other courses.

3.5 Ethics

Ethical clearance for conducting this research has been received from the Ethics Committee of the University of South Australia (Application IDs: 0000036182 and 0000033878). MOOC data in this study have been collected by edX. When participants sign-up for a MOOC on edX, they sign the Terms and Conditions, as well as the Privacy Policy (see Appendix B). In this policy, edX notifies potential participants that if they join the platform, their personal data will be collected by the system and may be used by the international participating institutions delivering the course. It is specified that research publications involving these data will not include identified information about the participants. With this, the participants consent that there is no reasonable expectation of privacy when they take a massive open online course.

The researcher has also signed an agreement with TUD indicating that her conduct is governed by the ethical principles of the institution. All data collected by TUD throughout the MOOCs' delivery has received institutional clearance allowing for further research. The data were transferred from TUD to the researcher using a protected TUD Surf server. Data were stored on a local password-protected UniSA computer.

Analyses in the thesis did not use any identifiable data, and when student quotes are presented, their nicknames are removed. The courses are no longer available for public access, except to those who were registered for them at the time of the delivery.

3.6 Chapter Summary

This chapter introduced the major philosophical and methodological considerations undertaken in the thesis. The chapter commenced by positioning the thesis within the LA research domain. By highlighting the challenge of "measuring what is easy to

measure" in LA studies, Section 3.1.1 addressed the need to use theory to focus the research inquiry, but also not to ignore the exploratory potential for data-driven analyses. Section 3.1.2 discussed Deweyan pragmatism as a philosophy that allows for the incorporation of tools that emerged in different research paradigms, in order to achieve the goal of solving a posed problem. The section justified the mixed method research design of this study through the pragmatist orientation that embraces different paradigms and allows for the linkage of observed phenomenon with the abstract ideas that may help to act towards it. Section 3.2 listed the research questions, and Section 3.3 explained the study design moving from the inductive to deductive analyses. Finally, the data analysed in the thesis was presented in Section 3.4, along with the ethical considerations guiding the thesis' inquiry (Section 3.5).

4 Exploration of forum log data

This chapter presents the data analyses that served to frame the empirical studies. The chapter suggests that a learner's level of sustained forum access over the duration of a course can frame analyses of interpersonal relationships that may arise through participation. In this context, the exploratory data analysis provides an empirical rationale for establishing a working typology of forum use. Three sub-groups of learners contributing to a forum's social context are delineated. The subsequent chapter (Chapter 5) evaluates the opportunities for relationship formation of the three identified groups: occasional posters, regular posters, and forum core.

4.1 Introduction

Prior research in formal online education has established practices and processes leading towards active learner-to-learner and learner-to-teacher interactions. For instance, the use of meaningful tasks to prompt exchanges of ideas, teachers' timely feedback and checks for understanding, as well as fostering a sense of community, have been noted to increase the quality and quantity of interactions (Darabi, Liang, Suryavanshi, & Yurekli, 2013; Ravenna, Foster, & Bishop, 2012). However, Chapters 1–-2 argued that there remains limited evidence demonstrating how these techniques can, or should, be applied at scale, given the distinct characteristics of MOOCs. The features of open participation and learner diversity pertinent to this new educational context challenge the notion that practices from formal online education settings can be simply transferred and adopted to MOOCs.

In a formal online course, students enrol to receive credit and formal recognition of mastery, largely anchoring the cohort by a shared goal. Conversely, in a MOOC, students are driven by a variety of goals, from sampling course content, to being interested in a subject area, or seeking peer interactions (Eynon, 2014). In addition, MOOCs are much more asynchronous than conventional online education (Mullaney, 2014). In stark contrast to the compulsory start and finish times in formal online courses, individuals can join most MOOCs at any point in time throughout a course's duration. The scale, diversity of learners' backgrounds and intentions, and the open

participation aspects of MOOC environments affect how the learning processes unfold in these educational settings.

In particular, as argued in Chapter 2, interpersonal interactions at scale may not evolve into the same social context that is assumed within the models of online educational pedagogy (Section 2.3.2). Due to varying learner intentions for course engagement (Eynon, 2014), only a tiny fraction of all registrants uses forums (A. Anderson, Huttenlocher, Kleinberg, & Leskovec, 2014; Breslow & Pritchard, 2013). As Eynon et al. (2014) demonstrated, some learners are not at all interested in engaging with other individuals, some may use a forum to find what they need quickly, while some consider it as an essential part of their online experience. Learners who use MOOC forums also drop in and out of them (D. Yang, Wen, Kumar, Xing, & Rosé, 2014), resulting in the constant change of forum participants. Although these findings suggest that interactions in forum use may drastically differ from those in formal online education, there is a dearth of research investigating patterns of engagement specifically within MOOC forums. Hence, for the present Doctoral thesis there is no analytical framework that can suggest a delineation for MOOC forum activity and how this may be analysed in terms of meaningful patterns of participation.

To provide an empirical description of forum use, this chapter presents exploratory analyses of forum access logs in four edX MOOCs. To understand the extent of forum participation by different forum users, the volume and continuity of forum log activity and the proportional size of the different groups of forum users were analysed. Forum users were then differentiated based on the regularity of their posting and viewing activity. In addition, to account for the impact that the MOOC forum environment may have on learner motivation, a group of forum users receiving timely replies, i.e. forum core, was singled out from regular posters. Interactions between occasional posters, regular posters and forum core posters were deemed relevant for analyses of emergent social contexts in MOOC forums.

4.2 Framing Forum Participation

To frame the data analyses, this section reviews literature relevant for describing forum use. The section further emphasises the importance of observing forum posting and viewing activity in MOOCs through the lens of continuity of an individual's activity. For example, regular posting or viewing activity can demonstrate a higher level of

motivation and commitment to engaging in social aspects of learning. The section also discusses the relevance of the indicators of posting reciprocation and timeliness. These indicators afford new opportunities to assess the impact of a social environment on learners' social activity. The next section describes the data from the four MOOC forums through these established indicators of regularity, timeliness and reciprocation.

4.2.1 Forum Use in MOOCs

To date, research examining student engagement patterns in MOOCs have not extensively nor exclusively investigated forum use. Instead, available studies have analysed how enrolled students interact with the totality of course resources (including forums, among other activities). Through such analyses, typologies of participants were identified (Coffrin, Corrin, de Barba, & Kennedy, 2014; Ferguson & Clow, 2015; Kizilcec, Piech, & Schneider, 2013). Empirical data supporting these typologies suggest that individuals exhibit diverse preferences as to when, how and in which combination they watch lectures, use forums, or complete assignments, if at all. For example, Kizilcec et al. (2013) noted that completing learners, i.e. those attempting assessment, exhibit higher levels of forum activity, than for instance, learners who audit by watching lectures without attempting assessment. Alternatively, Ferguson and Clow's (2015) study demonstrated that learners who only sampled content, posted very few comments (6%–15%), "strong starters" who disengaged with the course early on appeared active in posting. Some 73% of "strong starters" posted an average of 13.7 comments. "Mid-way dropouts" had lower posting activity, in contrast to "keen completers", "late completers" and "nearly there" participants, who all had higher levels of posting activity. The analyses of learner typologies are insightful in demonstrating that learners engage in social activities to varying extents. However, focus in research was kept on the volume of activity, with only a mention of the temporal aspects of learning activity. The studies did not illuminate how such activity accumulates to a group level, and into a social context. Existing perspectives unto participation patterns do not highlight the aspects of forum use related to stability to learner interactions, related to the relational qualities of interactions.

This section contrasts empirical data about posting and viewing activity in MOOC forums with available interpretations from formal online education. Through this review, an additional layer of continuity of interactions is suggested in order to provide

an indicator of the level of commitment exhibited by learners. For example, learners may be driven by a strong social motivation to spend more time in forums, and as such, frequently return to view and post comments. Alternatively, learners may be motivated by a need to address a pressing question or problem in a given course and therefore only visit if a need arises. From such perspectives, MOOC forum activities can be framed as emergent from the intertwined participation flows, on a spectrum, of residents and visitors (White & Le Cornu, 2011). The section proposes an indicator of regularity, i.e. frequency of forum access, to capture this distinction.

Posting and Viewing Activity. In formal online education there has been a tendency to evaluate forum participation from the perspective of the amount of messages posted, since the activity of posting messages has been associated with engagement (Hew, Cheung, & Ng, 2010; Hrastinski, 2008; Hrastinski & Keller, 2007). According to computer-mediated communication (CMC) research, a high degree of peer interaction that drives learning socially, and over time, can assist in the development of a community of learners (Gunawardena, 1995; Rovai, 2002; Rovai & Barnum, 2007; Swan, 2002, 2004). At the same time, viewers have been described as "vicarious learners" (Sutton, 2001) that are also engaged in listening activities (Wise, Speer, Marbouti, & Hsiao, 2012). This take on viewing behaviour is a more positive framing of what has been previously described as lurking activity. The connotation associated with a learner lurking in a forum was, as the term suggests, perceived as negative and as diminishing to the educational value potentially derived from online learning communities (Rovai, 2002). In formal educational settings, the most frequently cited reason for students not posting to forums and exclusively engaging in lurking behaviour was "just reading/browsing is enough", followed by "still learning about the group" and "shy about posting" (Preece, Nonnecke, & Andrews, 2004, p. 208). A more neutral perspective on the differences between learners who predominantly post or predominantly read was captured by Muller et al. (2010), who framed posting and reading activities as differentiations between learners who produce and those who consume ideas.

However, a simplistic differentiation of forum users who post and those who only read is insufficient to understanding the interactions in MOOC forums, on two important levels. First, posting activity in MOOCs should not be directly associated with motivation or interest to participate in forum activity. That has been different in formal

online education, where posting is the most acknowledged form of forum use (Hrastinski, 2008), and all learners in a course may be required to participate as a part of assessment. However, in MOOCs, most enrolled students never use forums. For instance, in an analysis of the first edX MOOC "Circuit and Electronics" (6.002x), Breslow and Pritchard (2013) observed that only 3% of those enrolled actually accessed the MOOC's forum. Anderson et al. (2014) similarly reported that "numerically the forum is used by a small fraction of the full population of registered students" (sec. 4). Yet, both studies demonstrated that highly engaged MOOC learners participate in the constructive behaviour of contributing to MOOC forums. In the "Circuits and Electronics" MOOC, around 27% of learners with certificates of completion asked a question in the forum, while 40.6% offered answers, and 36% commented (Breslow & Pritchard, 2013, p. 22). Further, Anderson and others (2014) also showed that the typical activity for learners with higher academic achievement is that of replying to, and not of initiating posts. Although these studies demonstrate that many learners engaged with a course's posts in a forum, the extent of forum use has been unaddressed. In other words, it is unclear whether the majority of learners completing a given course and using forums, post often or once, since the focus of forum use remains mainly at the level of the number of postings made. As previously noted, posting behaviour in a MOOC is intermittent, and learners may have engaged in posting behaviour on one occasion producing a relatively higher volume of posts, and then never return to the platform. In other words, it cannot be assumed that making a forum post is associated with an individual's intention to continuously use the forum for learning.

Furthermore, as much as posting activity may be a one-time exploration of a forum, viewing activity in MOOCs may still be associated with learners' interest and motivation to use the forum, irrespective of their own posting activity. Studies suggest that 90% of forum use comprises reading threads without providing additional comment (Breslow & Pritchard, 2013, p. 15). Mustafaraj and Bu (2015) also demonstrated that posting and voting makes up only 3.3% of all forum use, with the remainder being reading and searching. Similarly, Bergner and colleagues' study identified 95% of forum access logs reveal viewing activity, not posting behaviour (Bergner, Kerr, & Pritchard, 2015). Sharma et al. (2015) further demonstrated that viewing activity describes individuals with different motivations towards learning and course completion. The study by Sharma et al. revealed that the average number of forum

views is significantly higher for students who completed the course with distinction than for those who completed it with normal certification, and those who failed the course. In short, research showed that viewing activity is typical for most forum users in MOOCs, and that learners who complete MOOCs also favour lurking. These findings resonate with formal online educational notions that viewing and some of the listening activity associated with it, comprises a large portion of forum use (Wise et al., 2012). Therefore, much like in formal online education, lurkers or viewers may be lacking skills, confidence, motivation or simply the time to participate (Nonnecke & Preece, 2001; Preece et al., 2004). Yet, despite the overall understanding that viewing behaviour is closely intertwined with posting behaviour in MOOC forums, there is a lack of temporal characteristics describing viewing activity. Similar to the empirical research gap in relation to posting activity, there is a lack of insight demonstrating the extent and continuity of forum use through viewing behaviour.

To conclude, the interest and motivation of learners in MOOCs cannot be distinguished merely on the basis of the type of activity learners engage in. Learners with an interest to draw on social activity may never post, and learners with an interest to engage in social activity may post but then disengage. To further understand how posting and viewing behaviour patterns aggregate at a group level, analyses of MOOC forums need to reach beyond the simple distinction between these activity types.

Motivation and commitment to use forums. In a MOOC forum, reaching beyond observations of posting and viewing behaviour necessitates a consideration of learner intentions underpinning forum use. Research has offered solid evidence that a learner's choices in social activity are grounded in their commitment and motivation. Eynon et al. (2014) demonstrated that certain learners are interested only in accessing a forum when deemed necessary. Other learners strive to engage in a global learning community, while others do not need to access forums at all. Milligan and Griffith (2016) described engagement with forums as dependent on choices made by individuals about learning in a particular situation. For instance, one singular individual may engage in a social activity expertly navigating the scale and volume of messages, consistently posting in ways that may allow for the forging of relationships. However, in another MOOC this same person may choose to act differently in navigating forums, such as withdraw from learning with others, or access the forum for a short question-and-answer conversation. Both studies (Eynon et al., 2014; Milligan & Griffith, 2016)

pointed out that behaviour observed in forums is an outcome derived out of a choice for the given learning resource, and an attitude accompanying this decision.

White and Le Cornu's (2011) continuum of online engagement in digital spaces lends itself nicely to understanding forum use in a MOOC through learner intentions rather than skills. They proposed a typology of online engagement based on distinct attitudes towards the use of digital technologies on the web. For them, learner presence and commitment to such presence are key to differentiating between user behaviours. Such an approach was proposed to contrast the biological and technological determinism of Prensky's (2001) "digital natives" concept. White and Le Cornu (2011) challenged the premise that (inherent) skill fully determines the use of technology. Instead, individuals adopt digital technology for learning, based on the underlying choice as to what they want to use it for. Such choice is situational: depending on context, the same person can approach a specific digital environment differently. Two opposite choices of using technology can be summarised as choosing 'to visit' when a tool is necessary, or to 'reside' and make a digital space one's own. Individuals with a visiting engagement attitude are likely to define a goal or task, and select a tool online to meet associated needs, and then once the task is completed, leave. Visitors leave little or no social visibility after their Internet use. In contrast, a resident's attitude to a digital environment is that of a place that does not differ much from other physical places they participate in. Resident mode implies a motivation or an attitude "to connect to, or to be with, other people", according to White and Le Cornu (2011); for "residents, the Web is a place to express opinions, a place in which relationships can be formed and extended" (sec.IV.2). Residents are "happy to go online simply to spend time with others and they are likely to consider that they 'belong' to a community which is located in the virtual … While they use 'tools' such as online banking and shopping systems they also use the Web to maintain and develop a digital identity" (ibid., sec.IV.2). The differences in approaching a digital tool within the visitors-residents metaphor can help capture the possible differences in commitment to social activity.

An understanding of forum participation requires a capacity to distinguish between learners with different intentions towards social activity. Empirical data suggested that varying commitment to forum use may underpin both posting or viewing activity. Driven by a motivation to 'reside' in a forum, learners may want to return, and possibly post; driven by a motivation to 'visit', learners may once turn-up to forums to seek

quick solutions to their needs. Along these lines of reasoning, the following section proposes that indicators of regularity could capture differences in learner intentions towards forum use.

4.2.2 Measuring Regularity of Interactions

Temporal aspects related to posting and viewing activity can help understand the differences in the patterns of forum use. Arguably, regularity of forum activity is the single most important aspect for capturing the differences in learner forum activity. That is, continuous use of a forum indicates that there is an ongoing commitment, interest or need to use the tool for learning. For posters, regularity of use also impacts the development of a social context, as repeated activity creates continuity of interactions between the posters. Individuals who continue to interact could then evolve into a group, and through a regularity of interactions, a shared history can be created. To capture the potential of continuous interactions in MOOC forums, forum user data representing log posting and viewing activity were analysed alongside regularity of participation.

Existing measures of regularity in MOOC forums were found to either yield data that are too granular or unfit for the undertaking of a forum analysis. The online education literature previously associated learner regularity with the development of online communities (Burnett, Bonnici, Miksa, & Kim, 2007; Swan, 2003b). The adopted measures to address persistence in interpersonal interactions included counts of posts made in subsequent weeks of a course, or weekly mean. In other words, formal online research has focused on both the volume and week-after-week continuity of postings. Such approaches are not directly applicable to MOOC forums, however. MOOCs include learners with extreme posting activity (Huang, Dasgupta, Ghosh, Manning, & Sanders, 2014). That is, measuring the volume of participation skews results and does not represent all learners' activity in a forum. Alternative measures of regularity for MOOC contexts have been devised by Boroujeni and colleagues (2016). In this instance, researchers quantified regularity in MOOCs (including forums) through analyses of intra-day, intra-week and intra-course frequencies of course access. Boroujeni and colleagues found that intra-week regularity, i.e. participation on the same day of the week, was the strongest predictor of performance in the course ($r = .76$, $p < .001$). While these measures are potent, they favour students with high self-

regulated learning skills and discipline, as well as high interest in course completion. In sum, a simple indicator that could embrace students who have varying posting habits still eludes research.

In order to measure the frequency and potential continuity of forum attendance, this study approached regularity as the total number of weeks a given learner used a forum. These weeks were not necessarily sequential, in order to capture students who may have posted in the beginning, middle or end of the course. By adopting this approach to measure regularity, the study broke free from a reliance on the amount of posts a learner undertakes. That is, a learner was considered as attending a forum for one week, regardless of how many posts were made in that timeframe. This approach does not capture the time of the day, or similarity in the time when a given student would use a forum, as in Boroujeni et al. (2016). Such low granularity measure for the continuity of forum interactions was considered more suitable, due to the exploratory nature of the thesis. This method does not privilege particular posting behaviours, such as self-organisation or volume of activity. The measure allows for both prolific students, as well as those who make few posts in a sustained manner – altogether considered to be a part of continuous interaction.

4.2.3 The Role of the Environment: Receiving Timely Replies

The chapter has thus far explained that sustained weekly use of a MOOC's forum has been treated as a measure of continuity, in order to distinguish between a learner's intended motivation. The study also acknowledged that a learner's initial motivation could be impacted by responses received from the environment. For example, students with a high motivation to use a forum actively may be discouraged if they do not receive a timely reply or any reply at all. And vice versa, a learner lacking the intention 'to reside' could also be encouraged to use a forum more often by receiving prompt and helpful replies. To account for the potential impact of a forum, the study established two simple indicators: if a learner received replies, and if the replies were 'timely'. This sub-section reviews considerations around these indicators.

Receiving Replies. Aside from continuity of interactions, another important element for the formation of a social context is the bi-directional quality of interactions. In essence, if a learner does not receive a reply then no inter-personal interaction takes place. This bi-directional nature of an online interaction has been conceptualised in the

foundational work on distance education. Moore (1989) theorised that dialogue, structure and learner autonomy interact to bridge the psychological distance among participants of an online setting. He noted that with an increase in dialogue, the perception of distance grows smaller. Anderson (1989) extended Moore's notion of dialogue from that between a teacher and a learner to include peer learners. The bi-directional nature of interactions is embedded within community development, captured through the elements of interactivity (Tu, 2002; Tu & McIsaac, 2002). Also, the bi-directional nature of learning interactions is implicit within the analyses of knowledge construction. Knowledge peer-construction processes assume the presence of two peers interacting with one another. This assumption is implicit within knowledge construction metrics. For instance, the cognitive engagement of a conversation can be understood through the observable indicator of the number of turns taken between learners (Wiley, 2002). In short, to address the importance of bi-directional interactions to the emergence of a social context, postings that were replied to were singled-out.

This study poses a simple measure to capture the reciprocity of interaction: if nobody replies to a person's post, her interaction is not considered as reciprocated and remains as one-directional. Posts that are replied to, or are a part of a discussion thread are considered reciprocated or reciprocating.

Immediacy in Interaction. The timing of a reply can also have an impact on the emergence of a social context, and the motivation of a learner. The dimension of time (timing of posts and timing of responses) is crucial for understanding the asynchronous forum environment, even more so in an open course. Hesse, Werner and Altman (1988) argued that temporal aspects of message posting can be both enabling and constraining. According to Dringus and Ellis (2010), fast interactions that continue create a momentum that captures the nature of online discussions through interaction duration, frequency and density. One of the conventional constructs addressing the impact of timing is that of immediacy, i.e. how fast learners receive a reply. In their seminal work, Short, Williams and Christie (1976) explained that immediacy is one of the basic temporal characteristics that impacts how people form impressions of one another. The concept of immediacy has been associated with the establishment of social presence, the perception of another person as real in the online environment (Lowenthal, 2012; Short et al., 1976; Tu & McIsaac, 2002). That being said, immediacy is a subjective construct, as the timing of a reply may be perceived by learners in positive, negative or

neutral ways. To be able to overcome the challenge of measuring immediacy in a MOOC context, this study adopted a measure of 'relevance' equal to its measure of 'timeliness'. That is, replies to posts that were made before a weekly change of the curriculum topic were considered both timely and relevant in this study.

This chapter poses the indicators of immediacy through a measure of "timeliness of reply". This measure was determined by the timeline of the content-releases in the MOOCs examined in this study. As new content and new assessment tasks were being put forward each week, (posted) replies to (posted) comments that arose within one week were deemed 'timely' to the initial poster. In contrast, if replies to questions posted in the first week of the course were given in week five, they were considered as no longer relevant, and therefore, untimely. The separation between weeks strictly aligned with content-releases associated with the MOOCs in the present study.

4.3 Results of the Data Analysis

This section presents different data analyses concerning forum use. Section 4.3.1 reports the counts of posting activity, per week, as well as shows the regularity of posting behaviour. It also highlights the reciprocal and timely elements of interactions. Data reviewed in this sub-section are relevant for the practical distinction made between regular and occasional posters. Based on this distinction, Section 4.3.2 suggests a working typology of forum users, and describes the size of each group linked to the various courses examined in this thesis. Section 4.3.3 offers an overview of the volume of forum use by each of the groups. The chapter concludes with a summary of insights from the exploratory data analysis, and discusses their relevance for further sections.

4.3.1 Describing Participation in Analysed Courses

Tables 4.1–4.6 present descriptive analyses of forum posting activity in the four analysed MOOCs. The analyses demonstrate that few individuals posted in the forums more than in any one or two weeks. This suggests that the posting activity of those who posted beyond any two weeks of a course can be considered as an indication of a higher learner commitment. Furthermore, the volume of postings stabilised over time and the fluctuation seemed to align with assessment deadlines and set-up. The analyses also demonstrate that returning learners, delineated based on the quality of their interactions, such as receiving replies in a timely manner, represented an even a smaller group than

learners whose participation was characterised as sustained. The counts and observations are presented below (Tables 4.1–4.6).

The overall pattern of the posting activity across the courses showed a usual MOOC decline in volume, i.e. the number of posts made per week decreased as the courses progressed (Table 4.1). The number of posters also decreased from week to week. Although the differences in the size of the posting cohorts decreased throughout a course's progression, their size somewhat stabilised around weeks 3–5 (Table 4.2). Stabilisation seemed to align with the timing of the courses' first major assessment. Given that much smaller numbers of people posted in more than two weeks in total (Table 4.3), it can be said that each week's posting activity comprised a relatively small group of people who had previously posted, and a certain group of people who posted for the first time (Tables 4.1 and 4.3). Such was characteristic for all weeks of forum activity.

Table 4.1 Number of forum posts, per week, in four edX MOOCs

Week number	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10
FP	1331	914	1084	747	960	828	653	646	n/a	n/a
Solar	7841	2603	1909	1210	920	968	1063	1023	n/a	n/a
Water	5544	2681	2500	1262	1137	684	502	566	291	629
Excel	2820	1602	1191	801	658	626	650	660	n/a	n/a

Table 4.2 Number of participants posting, per week, in four edX MOOCs

Week number	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10
FP	392	292	219	181	177	174	177	154	n/a	n/a
Solar	4245	1360	887	588	355	382	426	302	n/a	n/a
Water	2675	1370	996	490	378	265	182	172	113	173
Excel	1264	647	517	297	236	249	198	220	n/a	n/a

Table 4.3 Number of posters, according to the total number of weeks they posted in forums (consecutive or non-consecutive), in four edX MOOCs

Num-ber of weeks	1 week	2 weeks	3 weeks	4 weeks	5 weeks	6 weeks	7 weeks	8 weeks	9 weeks	10 weeks
FP	670	175	80	43	24	14	3	13	n/a	n/a
Solar	4483	884	304	131	69	45	21	12	n/a	n/a
Water	2695	689	268	126	68	56	34	20	21	17
Excel	1579	382	168	80	41	15	10	12	n/a	n/a

Reciprocation of interaction occurred more frequently in courses with higher moderation (Table 4.4). Table 4.4 presents the amount of un-replied posts in each MOOC. In facilitated FP and Excel, the number of un-replied messages was low. In contrast, in the first weeks of unmoderated courses, the amount of un-replied posts was relatively high, such as in the courses' second week, ranging from 332 to 668 in the unmoderated Solar and Water.

Table 4.4 Number of un-replied posts, per week, in four edX MOOCs

Week number	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10
FP	4	0	0	0	2	0	0	1	n/a	n/a
Solar	2269	668	432	208	80	968	27	16	n/a	n/a
Water	702	332	208	59	41	21	10	10	7	6
Excel	1446	156	79	17	14	16	6	16	n/a	n/a

The longer one participated in a forum (i.e. the higher the total number of weeks), the more likely was posting activity bi-directional nature (Table 4.5). For example, in the forum of the unmoderated Water course, 2695 people made at least one post in any one week of the course (Table 4.3). Only 1731 of these people received at least one reply to their posting activity (Table 4.5). Similar discrepancies are seen across the courses for learners who only 'attended' forums for one week, for all courses, except for the highly moderated FP course. The gap between those who returned to a forum and were reciprocated, and those who weren't, closes with the increase of frequency in 'weekly attendance' (the total number of weeks an individual posted). For example, learners who 'attended' forums for any four randomly chosen weeks within the MOOCs: in FP out of 43 learners, all 43 received responses to at least one message every time they

attended; in Solar 121 people out of 131 received responses to at least one message every time they attended; in Water this occurred for 124 out of 126 people, and in Excel, all 80 people who posted for any four weeks in the course received replies. For learners who sustained their forum activity over a longer period of weeks, the reciprocation gap seems to have become smaller. That is, the longer one participated, the more their interactions could be characterised as bi-directional.

Table 4.5 Number of posters, according to the total number of weeks they posted in forums, and received replies (at any point, by the end of the course)

Number of weeks	1 week	2 weeks	3 weeks	4 weeks	5 weeks	6 weeks	7 weeks	8 weeks	9 weeks	10 weeks
FP	670	175	80	43	24	14	3	13	n/a	n/a
Solar	2058	631	265	121	69	45	20	12	n/a	n/a
Water	1731	605	246	124	68	56	34	20	21	17
Excel	1033	355	165	80	41	15	10	12	n/a	n/a

The quality of timeliness, i.e. receiving a reply to one's post within the same thematic week of a MOOC, seems to have increased with a continuity of interactions (Table 4.6). Comparing Table 4.5 and Table 4.6 demonstrates that despite continuous, bi-directional interactions, many posters did not receive timely replies. The highly moderated FP course was an exception: for FP posters who posted in any 4 weeks, 40 out of 43 of them were replied to within the same week. However, in the other courses, differences are larger: only 50 out of 121 participants who posted in any four weeks in Solar received replies in the same week as their posts. The gap is 76 out of 124 in Water, and 43 out of 80 in Excel.

Table 4.6 Number of posters, according to the total number of weeks they posted in forums, and received replies in the same thematic week of their posts

Number of weeks	1 week	2 weeks	3 weeks	4 weeks	5 weeks	6 weeks	7 weeks	8 weeks	9 weeks	10 weeks
FP	667	164	72	40	21	13	2	13	n/a	n/a
Solar	1903	246	93	50	30	10	6	7	n/a	n/a
Water	2052	455	155	76	35	30	13	13	8	11
Excel	1142	277	120	43	31	9	5	12	n/a	n/a

The fluctuation of activity partially resembles formal online courses. Course activity appears to have fluctuated alongside the courses' design and assessment – a pattern

observed in formal courses (Thomas, 2002). In FP, a new assignment was released weekly, and completing all assignments was required for the final grade. Thus, a drop in posting activity took place around weeks 6-7. This drop in activity corresponded with the assessment profile. That is, learners who did not receive a sufficient grade for passing the assignments may have disengaged if they were assessment-oriented. A similar pattern was found within the Excel course, with a comparable assessment strategy. The Solar and Water courses had several exams learners were required to complete in order to obtain a final grade. In both courses, a sharp decline marked the times where the first exams took place: in Water, in the middle of the course, and in Solar, around week-3. In other words, overall forum activity appears to have been in sync with the timing of assessment – a pattern that replicates formal course trends.

4.3.2 Typology of Forum Use

As discussed in the review of literature of Section 4.1, MOOC forums are used by learners who engage in distinctly different posting and viewing activity. Although studies have made some reference as to the volume of posting activity, the dynamics of forum use over a course's progression have not been addressed. The total number of weeks an individual engages in posting activity, i.e. frequency of posting, can be linked to learner motivation to engage in social learning processes. To differentiate between learner posting behaviour, this study distinguishes between students using forums (posting and/or viewing) for more than any two weeks, from the rest of forum users. The distinction was informed by the data analyses in Section 4.3.1. Simply posting throughout a period of any two weeks was deemed as insufficient for establishing the continuity necessary for the evolution of a social context. Posting throughout any four weeks was considered as too stringent, as the time period implies a commitment to post throughout approximately half the duration of a course. As such, three weeks (not necessarily continuous) was set as the minimum presence. Data analyses have also indicated that the frequency of forum use for over two weeks (not necessarily continuous) was correlated with receiving replies.

In addition to a differentiation between individuals engaging in various types of forum use (posting and/or viewing) with distinct frequency, this study's working typology distinguishes a sub-group of regular posters whose motivations may be affected by the quality of interactions they experience. More specifically, a regular posters' group

includes a sub-group of forum core posters, i.e. learners who posted in MOOC forums at least once per thematic week of a course, in any three or more weeks, and received timely replies.

Table 4.7 Typology of forum users

A LEARNER WHO		**b) POSTED IN THE FORUM**		
		Regularly (≥ 3 weeks)	*Occasionally (posted 1–2 weeks)*	*Never*
a) VIEWED THE FORUM	*Regularly (≥ 3 weeks throughout the course's duration)*	Group 1 REGULAR POSTER REGULAR VIEWER (FORUM CORE: regular posters receiving timely replies)	Group 2 OCCASIONAL POSTER REGULAR VIEWER	Group 4 NON-POSTER REGULAR VIEWER
	Occasionally (1–2 weeks throughout the course's duration)	--	Group 3 OCCASIONAL POSTER OCCASIONAL VIEWER	Group 5 NON-POSTER OCCASIONAL VIEWER

Distinctions between forum users are summarised within a working typology presented in Table 4.7, and applied to count the size of each of the groups present in the forums, according to the four courses examined. The size of each group of forum users below is described in relation to the total number of posters. The overall number of forum users across the four courses ranged from 6507 individuals in FP, 14175 individuals in Solar, 6862 individuals in Water, and 9005 individuals in Excel.

Regular Posters (Table 4.7, Group 1) – participants who posted at least once in any three weeks of their course. This group includes the forum core – participants who posted at least one post per week, in any three weeks of their course and also received replies within the same week of their posts. It is implicit that the viewing activity of regular posters was also regular; that is, they also read in forums for at least three weeks of the course, i.e. in the weeks they posted. Learners who posted sparingly throughout the course as well as hyper-posting individuals who posted often and a lot of messages are both examples of regular participants. The relative size of regular posters across the

four courses is 4.9% (*SD*=2.4%), and includes a forum core that constitutes 2.9% (*SD*=1.3%) of all forum users.

Occasional Posters, Regular Viewers (Table 4.7, Group 2) – participants who accessed the forums for at least any three weeks of their course, but did not post in every week they logged in. That is, despite their regular use of forums, their posting activity could not be characterised as continuous. Learners with a commitment to social learning but who also have lower confidence levels or are lagging behind in coursework or understanding are examples of this type of forum user. The relative size of this group of learners across the courses is 8.7% (*SD*=0.6%) of all forum users.

Occasional Posters, Occasional Viewers (Table 4.7, Group 3) – participants who accessed (viewed) the forums in less than any three weeks, and who also posted in less than any three weeks of the course. A participant who logs into the system in week-1 of a course and says hello, either to one person or to a group of people, exemplifies this type of forum user. The size of this group across the fours analysed MOOCs is 21% (*SD*=13%) in relation to all forum users.

Non-Posters, Regular Viewers (Table 4.7, Group 4) – participants who did not post at all but accessed their MOOC's platform regularly, in at least any three weeks of their course. A learner with a higher commitment to social learning, who due to a lack of confidence or a lack of time, joins a forum, and systematically reads its content exemplifies such a forum user. The relative size of this group is 11.03% (*SD*=7%) as compared to the overall size of MOOC participants.

Non-Posters, Occasional Viewers (Table 4.7, Group 5) – participants who did not post at all but accessed their course's platform in any one or two weeks of the course. A registered participant who logs into a MOOC's platform yet decides not to post or use the MOOC's forum in any way exemplifies this type of user. The size of this group is 53% (*SD*= 9%) relative to the overall number of forum users, across all four MOOCs examined.

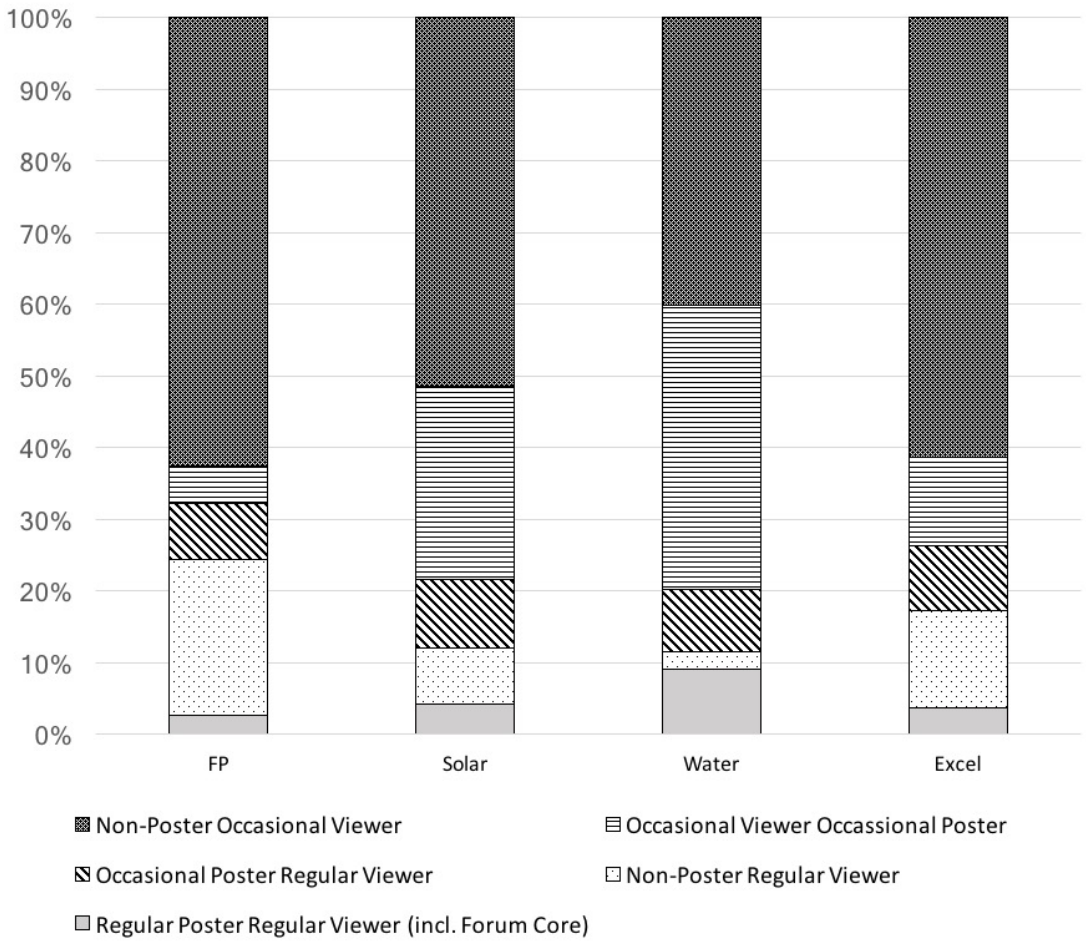


Figure 4.1 Relative break-down of main forum user types, per course

Figure 4.1 summarises the differences in forum use across the four courses examined. The figure demonstrates that the ratios of the groups of forum use differed across all four MOOCs. Courses with forum moderation by staff and teacher assistants (FP and Excel) demonstrate the highest ratio of learners who regularly accessed MOOC forums without posting. In contrast, the unmoderated Water course appears to have the highest ratio of learners who posted, as well as the highest ratio of learners who posted regularly. These indicators of forum use suggest that course design and forum moderation play a role in how a MOOC's forum is used. This insight regarding forum use will be further referenced in contextualising subsequent analyses, and will become relevant for the thesis conclusions.

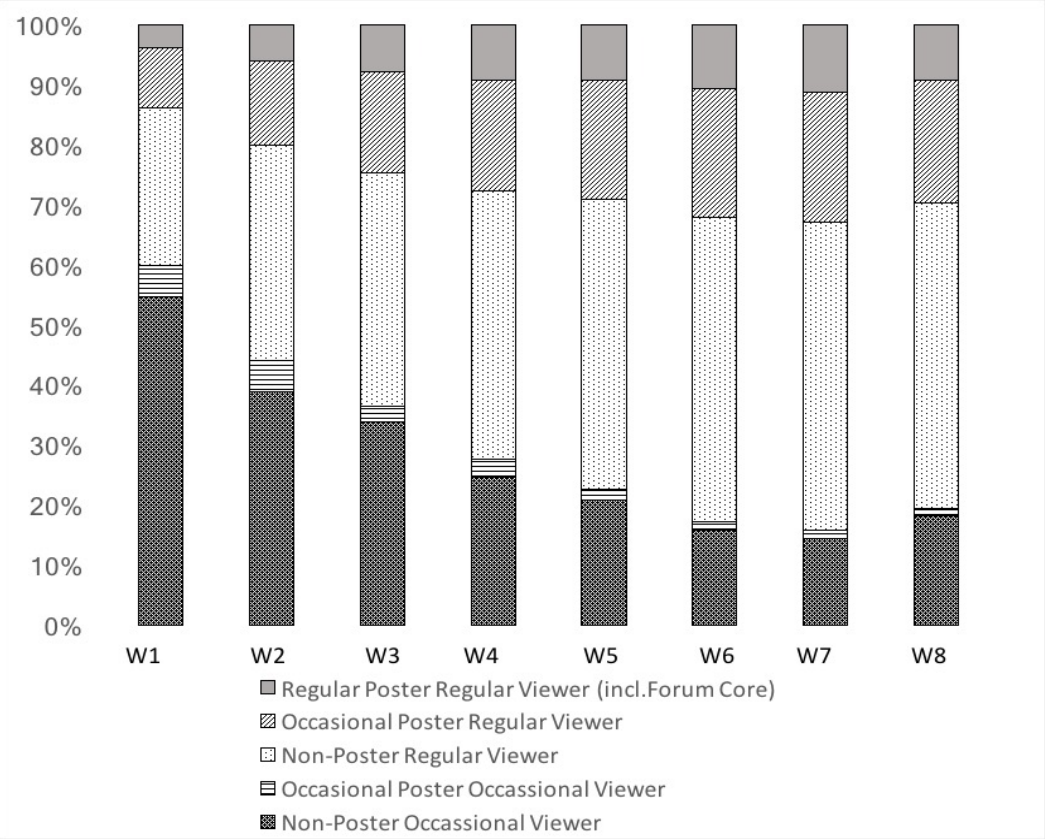
4.3.3 How Much is the Forum Used by Different Types of Forum Participants?

Weekly forum access was quantified in order to understand patterns of forum use across MOOC participants. Calculations were drawn from student log data, and the results indicate similar behaviour patterns across the courses. Figures 4.2 and 4.3 represent the volume of student access to forums and the number of people generating this volume of activity. Figure 4.2 refers to an example course – FP – and Figure 4.3 refers to Solar as an example.

Analyses of the clickstream activity demonstrate that regular posters (mean relative size 4.9%, *SD*=2.4%) produced the highest volume of forum activity, i.e. 10000-14000 clicks per week, across the various courses. Surprisingly, the number of users who did not post and occasionally visited (53% of all users, *SD*=9%) comprise around a quarter of the individuals using the forums from week to week. Yet, their clickstream activity was not as high as that of other sub-groups, suggesting a lack of exploratory activity for them, and probably simple searches for specific information. The results from this study demonstrate that this group undertakes more random visits to forums throughout the duration of a course.

Participants who accessed the MOOC forum and posted in any one or two weeks of their course (21% of all forum users), generated the highest volume of activity in week 1, and then became the least active of participants in the forums. The activity of this group markedly contrasted the sustained activity of those who, at minimum, occasionally (1-2 weeks of the course) accessed forums to read postings and information about their course.

a) Ratio of individuals in each of the forum use categories of the working typology (Table 4.7) accessing the FP forum every week



b) Number of clicks generated by each category of FP forum users (stacked); categories of forum users as presented in the working typology (Table 4.7)

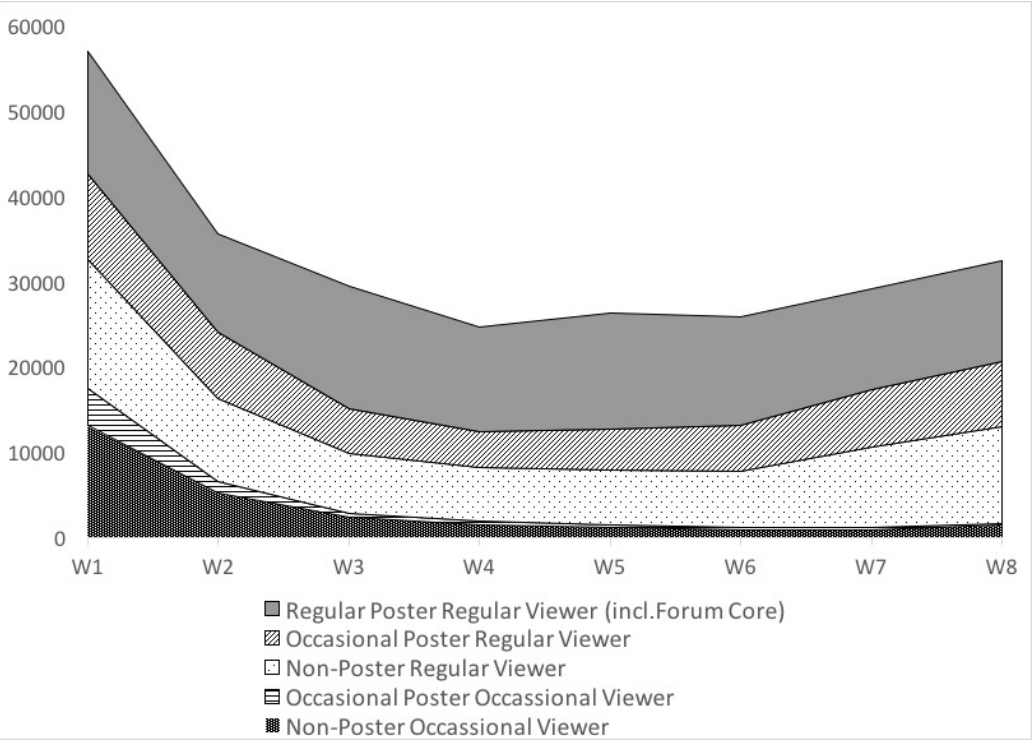
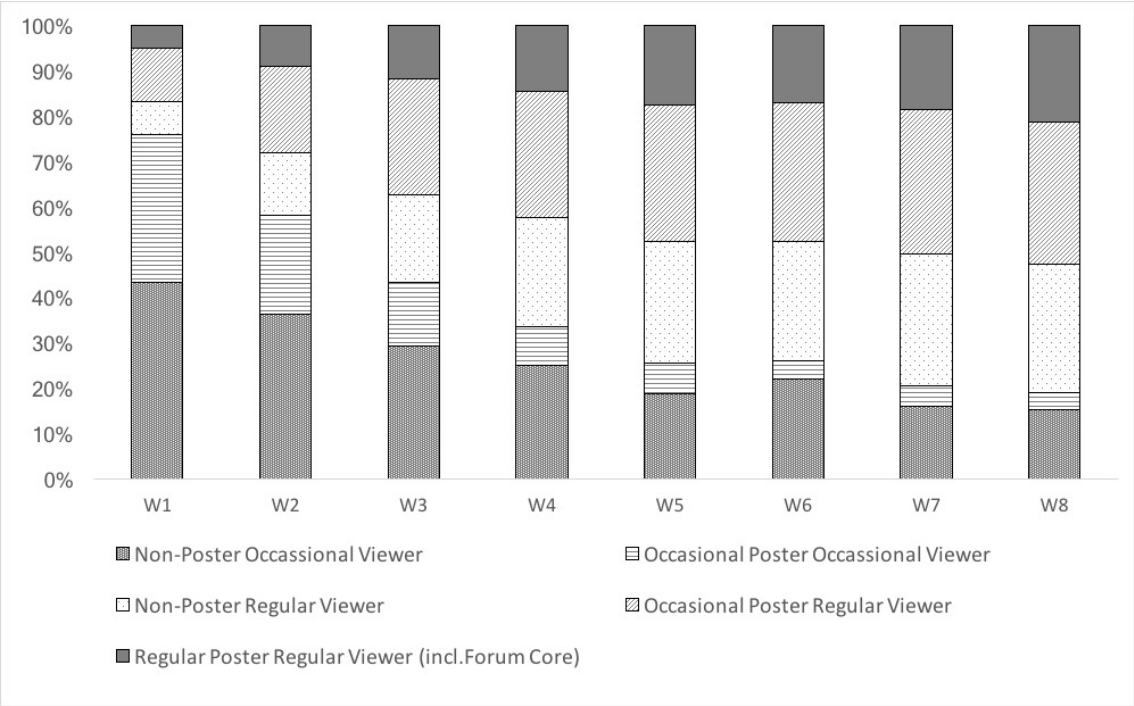


Figure 4.2 Forum access by different forum users (Table 4.7) in FP

a) Ratio of individuals in each of the forum use categories of the working typology (Table 4.7) accessing the Solar forum every week



b) Number of clicks generated by each category of Solar forum users (stacked); categories of forum users as presented in the working typology (Table 4.7)

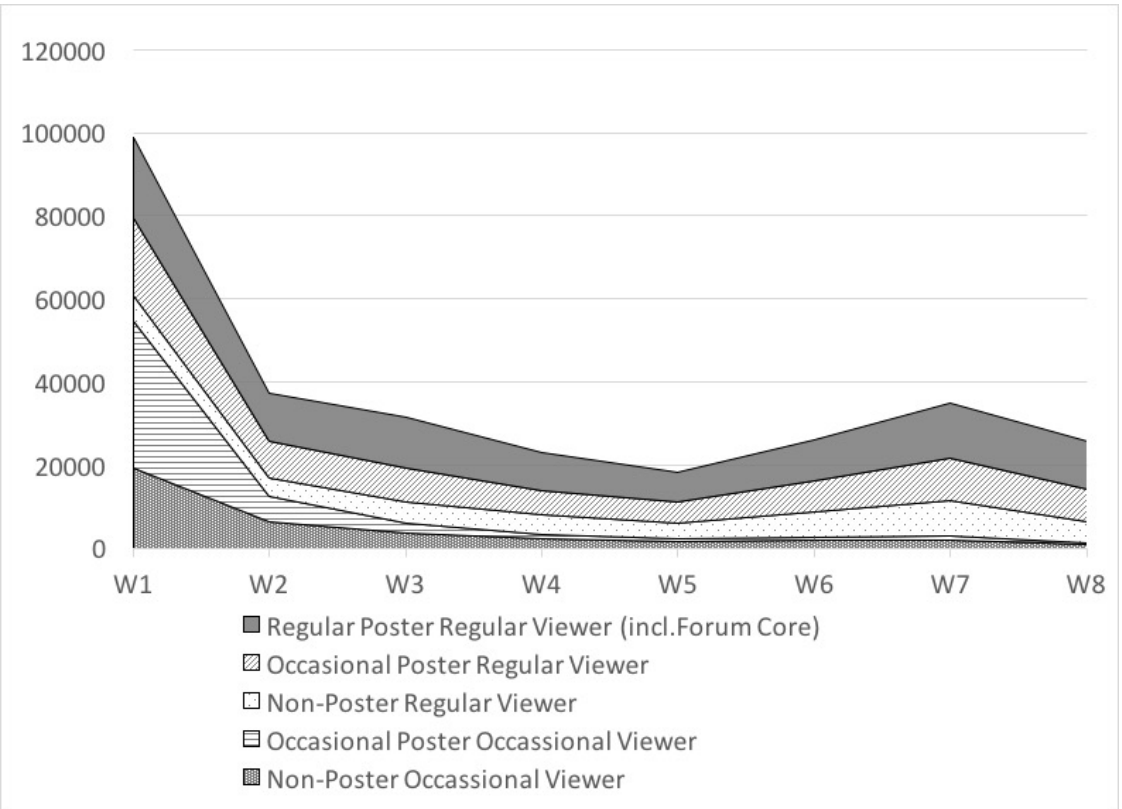


Figure 4.3 Forum access by different forum users (Table 4.7) in Solar

4.4 Discussion

To frame further empirical studies in this thesis, Chapter 4 presented the exploration of forum log data. Such exploratory analysis was necessary due to the lack of research on the patterns of participation activity limited to MOOC forums, as well as addressing the temporal quality of participation. Therefore, exploratory analysis of the regularity of forum activity was conducted. Besides delineating forum users on the regularity of their forum activity, a group of individuals receiving timely replies (forum core) was singled out from regular posters. Descriptive analysis addressed the volume and continuity of forum activity and the proportional size of these different groups of MOOC forums users. The results suggest to focus further analysis in this thesis on the interactions by the forum core, regular posters and occasional posters since that they contribute to the creation of the social context.

Using indicators of continuity, an arbitrary distinction was made between learners whose forum use is less frequent than in any two weeks (occasional) and MOOC participants whose activity is as frequent as in any three weeks, or more (regular). Based on this distinction, all MOOC forum users were grouped through a combination of posting and viewing activity:

- regular posters with regular viewing activity;
- occasional posters with regular viewing activity;
- non-posters with regular viewing activity;
- occasional posters with occasional viewing activity; and
- non-posters with occasional viewing activity.

Besides the creation of the working typology of forum use, exploratory data analysis examined weekly volume of forum use by each of the groups of forum participants. Analysis of forum clickstream activity revealed that most of the activity on the forum is generated by 5% of the forum users represented by the regular posters.

In relation to the topic of this thesis, the chapter suggests that a learner's level of sustained forum access over the duration of a course can frame analyses of interpersonal relationships that may arise through participation. Results demonstrate that, on average, 5% of forum posters return to make forum contributions for any three or more weeks. In the selected courses that applies to a range of 150-300 who people may experience

continuity of group development through forum interactions. That is, indicators of scale, continuity and pace of posting activity suggest that learners who are defined by repeated participation and/or receive timely reciprocated replies, are a moderately sized group that may as well garner interpersonal relations over time. At the same time, influenced by the scale of forum activity and perceived difficulty in navigating it, the potential for interpersonal relationship between the regular forum posters may equal that between the occasional posters. Therefore, to examine if forum participation affords the formation of interpersonal relationships, the thesis will focus on the evaluation of the opportunities for relationship formation of the three identified groups of posters: occasional posters, regular posters, and forum core. The non-posters will be excluded from the analysis as they did not participate in the creation of the social context.

The forum access of these groups of users, as addressed in this Chapter, highlighted the overall patterns of MOOC forum participation. The findings generally support previously reported "funnel of participation" observed within the overall use of MOOC resources (Clow, 2013b; Coffrin, Corrin, de Barba, & Kennedy, 2014; Kizilcec, Piech, & Schneider, 2013). In a funnel-like manner, the overall volume of participation in a MOOC forum decreases from week to week as the course progresses. However, this chapter's analysis suggests that such a "funnel of participation" describes the activity only of those forum users who occasionally posted and occasionally viewed the forum. It is plausible that these individuals largely overlap with the group of "strong starters" by Ferguson and Clow (2015): representing around 14% of all MOOC registrants, the strong starts disengage by week 3 but their posting activity is an average 13.7 posts per person. However, the pattern of the decrease of the volume of activity does not apply to the posters who participate in the forum regularly (three or more weeks). Figures 4.2 and 4.3 demonstrate that the volume of activity by those who access the forum regularly, in fact, increases by the third quarter of the course. Such increase is even higher for those learners who make posts in only one or two weeks, whereas the number of those posting in three or more weeks remains rather stable. Hence, the volume of forum participation activity can be described as decreasing for those viewing the forum occasionally, and increasing for those who view it regularly.

It should be noted that the perceived stability of the forum volume for those posting regularly does not imply that this group has the same composition from week to week. Table 4.3 summarized the total number of weeks that learners posted in each of the

analysed courses. The results suggest that some individuals participate more frequently than others, and a larger fraction of regular forum posters contributes to the forum for a lower total of weeks. For instance, in FP 80 people posted in any three weeks, while 40 people posted in any four weeks, and 24 people posters in any five weeks. Given that the sequence of these weeks may differ significantly from person to person, the activity of regular forum participants could be compared to "tides" or "waves" that keep returning at varying time intervals throughout the course progression. Therefore, the composition of the group posting on the forum regularly can only be comparable across time windows larger than one week, and are defined by course stages, such as assessment deadlines, rather than on a weekly basis.

The findings allow to draw some parallels with the learner clusters established in the previous MOOC research. As previously noted, forum users with occasional posting and occasional viewing activity seem to overlap with the "strong starters". Individuals who post and view the forum regularly (during three or more weeks) could be compared to the cluster of "keen completers" (Ferguson and Clow, 2015). The "keen completers" represent some 7%–13% of all MOOC registrants. According to Ferguson and Clow, two-thirds of the "keen completers" cluster contribute an average of 20.8–24.4 posts, and in some cases that indicator could be even higher. The size of the regularly participating forum posters reported in this chapter was an average of 5%, and also with a stable activity, alike the "keen completers". However, some of the "late completers" and "nearly there" as well as "mid-way dropouts" (Ferguson and Clow, 2015) could also be members of either regular posters regular viewers group, or occasional posters regular viewers group, as they exhibit course activity throughout the course duration, as well as may engage in posting. Furthermore, any of these groups could, in fact, also be engaging in regular viewing activity without posting, or occasional posting activity with occasional viewing. Ferguson and Clow (2015) demonstrate that all the learner clusters engage with MOOC resources every week, hence, they may participate in a forum activity dependent on their personal informational need and orientation towards learning through forums.

Forum activity appears to exhibit features found in formal online courses and open Internet communities. Similarly to the dynamics of contributing to open online forums, a small fraction of the users generated largest part of the content (Ochoa & Duval, 2008; OECD 2007; van Mierlo, 2014). Overall low frequency of appearance to contribute to

the forum appears comparable to contributions within open discussion forums that are driven by production cycles (temporal boundaries that are performance-driven) (Hecking, Göhnert, Zeini, & Hoppe, 2013). Yet, the activity within the forum suggests that the design of the course (its planned and formal element) structures the peaks of activity, and its stabilisation over time.

Analysis offered some insight into the viewing activity of the forum users. In particular, the results of exploratory analysis showed that number of individuals who regularly access the forum but do not post increases with the progression of the course. This finding can, in part, be explained through the research by Rosé and colleagues (2014). The authors empirically demonstrated that it is more challenging for learners to join the forum discussion if they join the forum at later course stages. Another possible explanation could be that the learners who do not get replies in earlier weeks (Table 4.4) may get discouraged from further participation. In particular, D. Yang and colleagues (2015) suggested that unresolved confusion expressed through a forum post is likely to result in forum disengagement activity. Besides these extrapolations, the findings highlight that the motivations of those learners with lower posting activity but regular forum use remain unclear. However, deeper analysis of the activity by the non-posters is beyond the focus of the thesis, whereas further Chapters are focused on individuals who exhibited posting activity.

4.5 Conclusion

This chapter analysed forum log data around the indicators of regularity of posting and viewing activity. For the purposes of this study, further analyses focus on the comparisons between three groups of posters: occasional posters (posting in the forum in any one or two weeks), regular posters (posting in the forum in any three weeks or more but not experiencing timely replies) and the forum core (regular posters who experienced timely replies).

The research propositions outlined in Chapter 2 are further examined through the analysis of these three posting groups. Chapter 5 investigates the potential for opportunities for relationship formation reflected through the extent to which social interactions between these various posters could result in co-occurrence between two individuals. In such instances a higher co-occurrence indicates the potential for a personal connection to be established.

5 Potential for relationship formation in edX forums

This chapter builds off the earlier exploratory analyses to compare the level of co-occurrence between all posters, regular posters and forum core posters in this study's four MOOCs. A comparison of co-occurrences across the various forum user groups allows for the identification of specific learners that demonstrate a potential for relationship formation. The nature of the discussions underpinning the co-occurrences is then examined in Chapter 6.

5.1 Introduction

The ability to technically scale the delivery of courses to an unprecedented number of learners facilitated the rise of MOOCs. As discussed in the literature review, due to their similarity with the online education paradigm, MOOCs have often been examined as formal online courses where analyses have been focused on the completion and retention of MOOC learners. However, the idiosyncrasies between MOOCs' contexts and formal contexts, such as open participation, scaled learner diversity and divergence of learner intentions, call into question the extent to which the two settings are comparable. In particular, the social context based on interpersonal relationships of familiarity and trust emergent from sustained frequent learner interactions in formal courses may not evolve in a MOOC setting.

The literature review (Chapter 2, Section 2.4.1) discussed the importance of investigating how individuals establish relationships through the use of MOOC forums. In comparison to formal educational settings, interpersonal interactions in MOOC forums take place among an increased number of learners and volume of messages. Such a setting may impede a learner's opportunity and willingness to form interpersonal relationships with peers. Thus far, the literature has largely supported this concern, regarding the ability of a learner to form relationships in MOOC settings (see Section 2.4.1). In essence, as the size of a cohort increases there is a subsequent negative impact on a learner's capacity to form an affective connection with a peer. As Knox (2014) and Veletsianos (2013) further added – the sheer volume of users, and hence the flow of messages, can also create experiences of anxiety and discomfort for individual learners. However, there is a current lack of empirical evidence that can adequately

describe the potential for two people to co-occur and eventually form an interpersonal relationship in MOOCs.

To analyse the potential for relationship formation in a MOOC setting, this chapter examines co-occurrences between different forum users. To differentiate between the types of forum use, the chapter builds on the typology of forum use developed in Chapter 4 (Section 4.3.2). Social network analysis is employed in order to examine co-occurrence networks of the forum core, regular posters and all posters. Methodological choices behind the construction of the networks, definition of nodes, ties and network boundaries are presented in Section 5.2. The section offers detailed explanations of the social network analysis measures of network centralisation, node-level and dyadic descriptive metrics, such as edge weights and degree, and interprets them in relation to the educational context of a MOOC forum. Section 5.3 presents the results of the descriptive social network analysis of the structure of co-occurrence networks and their change over time. Upon descriptive examination of the structure, the results of permutated analyses of variance (ANOVA) are reported. Statistical analyses applied to the network data demonstrate the differences in co-occurrences between the groups. The findings reveal the potential for a forum core to establish a majority of interactions within a regular posters' group as well as amongst the broader cohort of learners (all posters). The findings are positioned in relation to research on MOOCs and networked learning.

5.2 Research Questions

To examine the potential for co-occurrence between learners whose motivation and quality of interactions differed, the following overarching research question (RQ) has been posed:

What types of forum participation are associated with opportunities for relationship formation?

The research question was further de-limited into the following:

RQ1: What insights can be gained about a forum core co-occurrence network in relation to regular posters and all posters?

RQ2: How does the frequency of co-occurrence change over time for forum core posters, regular posters and all posters?

RQ3: Are co-occurrence frequencies for a forum core, regular posters and all posters significantly different?

5.3 Methods

To analyse the opportunities for co-occurrence between participants of varying commitment levels to forum participation, three sets of networks were constructed for each of the studied MOOCs. The networks were constructed based on the data recorded from a learner's posting activity on the edX forums for the courses examined. Each constructed network represented the co-occurrences of a particular group of posters: forum core, regular, and all posters. The networks were not mutually exclusive. That is, the forum core's interactions were a part of the regular posters' network, and regular posters' interactions comprised a part of the all posters' network. Descriptive social network analysis was used to understand the network structure. Permutation tests were used to draw generalisations about descriptive metrics derived during the social network analysis. The following section outlines the process for network construction and analysis.

5.3.1 Network Construction

This section details the methodological choices pertaining to the type of networks, definitions of nodes and ties, and explains the network boundaries adopted in this research. Brandes et al. (2013) asserted that "a network model should be viewed explicitly as yielding a network representation of something […This process entails] a specification of how the phenomenon … is abstracted to the network […and how] this conceptual network is represented in data" (p.3). Further, Robins (2015) emphasised that for a network to be defined, a researcher has to specify three elements: a tie, an edge, and a network boundary. According to Laumann, Marsden and Prensky (1989), errors in defining the network boundaries in network analysis reach beyond the consequences of slightly biased estimates of population means, proportions, or inefficiency in statistical estimation. Flawed boundary specification may lead to "the fundamental misrepresentation of the process under study, since …errors of omitting one actor may distort the overall configuration of actors in the system and render the entire analysis meaningless" (p. 19). Accordingly, this section offers an elaborate explanation of the definition of ties, actors and boundaries of the analysed co-

occurrence networks, as well as how forum log data were used to construct networks representing posters' co-occurrence.

Co-Affiliation Networks. Co-occurrence networks belong to a class of affiliation networks (Borgatti & Halgin, 2011). These are network graphs constructed from two independent sets of nodes: actors and events in which they participated. Commonly, co-participation in events is based on membership or engagement in a common activity. Examples include attendance to club events by southern women (Davis, Gardner, & Gardner, 1941), or relations between directors of company boards in Norway (Opsahl, 2013), or collaboration on scientific projects (Newman, 2001).

Co-participation in events could also indicate a potential for the development of social ties (Borgatti & Halgin, 2011). In such cases, co-affiliation data were collected to understand relations between a delineated set of actors, not between actors and events. In the context of the present study, MOOC forum postings that are part of one discussion thread constituted a common event that, if repeated, could result in an increased sense of familiarity and perceived presence between pairs of posters (dyads). Based on co-participation in events, two mode graphs are collapsed to only include actors, where a tie is representative of a shared event.

Defining the Nodes. Individuals who participated in online learning in a MOOC are included as actors in a network. Learners can participate in the course's forum in various ways (Hrastinski, 2009): they can read messages posted by others, they can write their own, reply to other people's messages, as well as endorse each other's content by voting. However, only participation that can be perceived by other learners would contribute to the development of interpersonal relationships. Therefore, for the analysis, all posters who made at least one post were included as network actors. Additionally, since interpersonal relationships could be formed between learners and staff, all posters (staff and students) were included in the analyses. The process of voting (like/dislike) on discussion posts were removed as the technical user design of the platform did not allow for the researcher to infer dyadic relationships between a voter and the voted. The edX platform was designed to use voting activity as a means to moderate the quality of group-level discussions. That is, learners could vote on a thread, but not individual content. Further, if a vote was given, posters in the discussion could not readily see who endorsed them, thereby reducing the opportunity for

reciprocal responses and interactions. Hence, the ability to align voting activity with user-to-user relationship construction was not feasible.

For each MOOC, three sets of networks were constructed, each defined by a specific sub-set of forum participants. Table 5.1 presents the number of posters across the 4 MOOCs. The first network comprised the forum core. This sub-population included participants that made a post within any three or more weeks of the course, and who also received replies within the same thematic week as each of their posts. Posters who made comments to their own postings but didn't receive replies to their posts from others within the same week of their postings were not considered. The second network consisted of regular posters. Specifically, this included all posters who made a post in any three or more weeks of the course, regardless of if or when they were reciprocated. The final network was developed to encompass all posters. Essentially, this network contained each participant that made at least one post in the MOOC forum.

Table 5.1 Number of posters in analysed courses

Course Id	Number of Posters		
	Forum Core	Regular Posters	All Posters
FP	161	177	1027
Solar	255	582	5949
Water	341	610	3994
Excel	220	326	2287

Defining the Ties. Networks were constructed from the information about a poster ("id"), a discussion thread into which the contribution was made ("thread id") and a time stamp as to when the post was made ("date"). This information was extracted from mongo files that encapsulated forum content, exemplified in Table 5.2.

Co-participation in a conversation with the same "thread id" ($oid) constituted a tie between two forum participants (parent-child relationship structure). The forum output did not include information as to whether the posters specifically interacted with one another's content (i.e. a learner read and then posted a direct reply or a learner simply replied to a message that may or may not have related to the specific topic at hand). It is not always evident from reading forum content as to whom the messages are being addressed within a thread (e.g. a reply to a comment or parent thread). Thus, a limitation of the approach to the network construction is that it captures opportunities to co-occur rather than the discrete co-occurrence as perceived by pairs of posters (dyads).

As a result, a tie resulting from co-occurrence in a discussion thread represents an opportunity for interaction. These opportunities as a whole demonstrate a network's potential for the formation of interpersonal relationships.

Table 5.2 Example of one row from a mongo forum file output (over 16000 rows, extracted information is bolded, student information removed)

{ **"_id**" : { "$oid" : "523f622b242b43c5040000f5" }, "_type" : "CommentThread", "abuse_flaggers" : [], "anonymous" : false, "anonymous_to_peers" : false, "at_position_list" : [], "**author_id**" : "STUDENT_INDIVIDUAL_ID", "author_username" : "STUDENT_INDIVIDUAL_NICKNAME", **"body"** : "Here you can post your questions about homework assignment HW 2.5.\n\nWe really appreciate students helping other students out, but please try not to reveal the answer in doing so.", "closed" : false, "comment_count" : 123, "commentable_id" : "i4x-DelftX-ET3034-course-Solar_Energy", "course_id" : "DelftX/ET3034TUx/2013_Fall", "**created_at**" : { "$date" : 1379885611054 }, "historical_abuse_flaggers" : [], "last_activity_at" : { "$date" : 1386820859337 }, "pinned" : true, "tags_array" : [], "title" : "[OFFICIAL] Week 2: HW Assignment 2.5", "updated_at" : { "$date" : 1381086741412 }, "visible" : true, "votes" : { "count" : 8, "down" : [], "down_count" : 0, "point" : 8, "up" : ["ID_OF_VOTER1", " ID_OF_VOTER2", " ID_OF_VOTER3", " ID_OF_VOTER4", " ID_OF_VOTER5", " ID_OF_VOTER6", " ID_OF_VOTER7", " ID_OF_VOTER8"], "up_count" : 8 } }

An additional consideration when establishing the tie definition relates to the forum posting structure initiated by the staff of courses. In two of the analysed MOOCs, forum moderators changed the structure of the course's forum mid-way throughout. The implication was that the same specifications within the log data represented quite different ways of forum usage (see Appendix C for examples). Typically, forum posting activity was differentiated by commencing a new discussion post (1st-order post), adding a comment (2nd-order post) or a reply (3rd-order post). Usually 1st-order posts set the problem statement. Then, 2nd-order posts constitute replies to 1st-order posts. Finally, 3rd-order comments represent detours or sub-conversations that may occur at a micro-level between individual interactions.

However, in two of the analysed courses, at times, the moderators used 1st-order comments to create sub-forums, resulting in participants adopting 2nd-order comment functionality to create discussion threads themselves. This was applied to some 35 discussion threads. Otherwise, in line with the more conventional forum structures, 2nd-order functionality was generally used to reply directly to 1st-order comments. The 35 divergent discussion threads were manually identified and turned into exceptions within the automated network construction process. As a result, for these conversations,

co-participation in 2nd-order comments defined the presence of ties between the posters, while for the rest of the course the 1st-order posts defined the ties, even between those posters who co-occurred in separate 2nd order discussions nested under the same 1st-order posts. Examples are provided in Appendix C.

Similar exceptions were made for discussion threads where learners greeted each other. Early research demonstrated that many MOOC courses have these discussions in the early weeks of the course that are produced at rates that are not feasible for students and teaching staff to read through (Brinton et al., 2013), resulting in feelings of information overload. While this volume of introductory posting causes much noise on the discussion forums, the conversations are however, often described as best practices in formal online education. The process for social introduction in formal education settings are meant to establish common ground between learners (Goodyear, Salmon, Spector, Steeples, & Tickner, 2001). Thus, instead of discarding these noisy conversations, they were reconstructed to reflect the personal impressions that may have taken place through them. In those conversations it may appear that everybody greeted everybody. However, often 1st-order comments simply represented an invitation to say hello, presenting a sub-forum for introduction. In these cases, 2nd- and 3rd-order comments were more resemblant of conversations between participants. An example is also included in Appendix C.

Finally, since co-occurrence as an activity has no direction, the ties were considered as undirected. The ties were valued, and tie weights represented the number of individual discussion threads learners co-occurred through. That is, if pairs of posters (dyads) co-occurred in one discussion only, the weight of the ties equalled one, regardless of the number of turn-taking within that particular discussion thread.

5.3.2 Temporal Network Boundaries

All cross-sectional networks of posters that had been formed by the end of the course, were restricted to the first and the last date of content delivery. For instance, forum interactions that occurred prior to the official course commencement were discarded from the analyses. The same protocol was applied to interactions that took place after the last lecture was released. It has been common across the four analysed courses that the final assignment or final exam deadlines, as well as result announcements would extend from two to four weeks past the end of the course. All four analysed MOOCs

had been archived and posts were made a year after end-dates. All interactions that took place more than one week past the release of the last lecture or homework were not considered in the analysis. Since communication between actors can continue well past a course's lectures due to exams or assignments, any established closing date for the forum analysis would be arbitrary. The discrete temporal limits applied in this study made the comparison across the four courses feasible.

In addition to the analysis of the cross-sectional networks of posters, the accumulation of ties over each course's duration were analysed. Each course was broken into stages defined by assignment (assessment) deadlines, necessary for receiving course certification. These stages differed across the four courses due to differences in assessment design, described in Table 5.3, further referred to as Stage 1, Stage 2, etc.

Table 5.3 Temporal boundaries for longitudinal analysis of network development

Course	Stages	Time Boundaries	Stage Duration	Assessment Design Notes
Solar	Stage 0	2013–09–16 00:01 – 2013–09–22 23:59	First week	Stages 1–3 were thematic; each concluded with an exam; taking only exams regardless of weekly assignments was sufficient for passing the course.
	Stage 1	2013–09–23 00:01 – 2013–10–06 23:59	2 weeks	
	Stage 2	2013–10–07 00:01 – 2013–10–27 23:59	3 weeks	
	Stage 3	2013–10–28 00:01 – 2013–11–11 23:59	2 weeks	
Water	Stage 0	2013–09–16 00:01 – 2013–09–22 23:59	First week	Stages 1–2 were thematic, each concluded with an exam; taking only exams regardless of weekly assignments was sufficient for passing the course.
	Stage 1	2013–09–23 00:01 – 2013–10–20 23:59	3 weeks	
	Stage 2	2013–10–21 00:01 – 2013–11–24 23:59	4 weeks	
FP	Stage 1	2014–10–15 00:01 – 2014–11–12 08:00	First 4 weeks	Stage 1 was defined by the first deadline on weekly assignments for the first four weeks being in week 4; all subsequent homework tasks were due by the end of each following week; all homework assignments equally contributed to the grade; the deadlines were set at 08:00 UTC, weekly.
	Stage 2	2014–11–12 08:01 – 2014–11–19 08:00	1 week	
	Stage 3	2014–11–19 08:01 – 2014–11–26 08:00	1 week	
	Stage 4	2014–11–26 08:01 – 2014–12–03 08:00	1 week	
	Stage 5	2014–12–03 08:01 – 2014–12–10 08:00	1 week	

Course	Stages	Time Boundaries	Stage Duration	Assessment Design Notes
Excel	Stage 0	2015–04–07 08:01– 2015–04–14 07:59	First week	Stage 0 was marked by the first week of the course; Stage 1 was defined by the first deadline for homework tasks of the first three weeks; all subsequent weekly assignments were due by the end of each week; all homework assignments contributed equally to the grade.
	Stage 1	2015–04–14 08:01 – 2015–04–28 07:59	2 weeks	
	Stage 2	2015–04–28 08:00 – 2015–05–05 07:59	1 week	
	Stage 3	2015–05–05 08:00 – 2015–05–12 07:59	1 week	
	Stage 4	2015–05-12 08:00 – 2015–05-19 07:59	1 week	
	Stage 5	2015–05-19 08:00 – 2015–05-26 07:59	1 week	
	Stage 6	2015–05-26 08:00 – 2015–06-02 07:59	1 week	

5.3.3 Descriptive Network Analysis

Co-occurrence patterns in networks of the entire posting cohort, regular posters and forum core posters were compared by analysing size, density and network centralisation. Individual node attributes were evaluated using degree and edge weight distributions. The following section explains the network measures employed in the study.

Size and Density. Size and density are basic network measures (Robins, 2015). Size represents the number of nodes and edges between them, whereas density indicates the ratio of how many ties exist within a network, out of all ties that could have been possible (Knoke & Yang, 2008). For instance, when all participants communicate with each and every other member in a course the resultant network density equals 1. According to Kadushin (2012), the density of a network is associated with the degree of cohesion within the analysed group. That is, the more connections there are between nodes, the faster the exchanged content spreads, reaching most nodes in the network. High density may be interpreted as a structure conducive for the transmission of ideas, rumours or disease.

In an e-learning context, density can be understood as the ratio of activated connections. According to Haythornthwaite (2002), a latent tie is "a tie for which a connection is available technically but that has not yet been activated by social interaction" (p. 389). MOOC participants reading in forums, therefore, have latent ties to one another, and the

extent to which they have created connections is then represented by the density of posting activity.

Degree and Edge Weight. Metrics such as size and density represent the state of learner connections at the collective level. Degree centrality indicates individual-level connections, as it represents the number of other actors one interacted with. Degree is often further de-limited into in- and out-degree. That is, for directed networks, the in-degree could represent actor popularity, as the flow of goods, services or information comes in from other network nodes. Out-degree is associated with the expansiveness of one's activity as the goods exchanged are directed towards other network actors. As previously noted – given the nature of the forum interface – the present study employed an undirected network. Consequently, only degree centrality measure was considered. In this analysis, degree is bi-directional and represents the number of people a poster co-occurred with (e.g. actor A co-occurred with actor B, implying that actor B also co-occurred with actor A). Degree distribution, i.e. the number of nodes with each given degree (Robins, 2015, p. 23) for each network includes minimum, mean, median, and maximum number of individual co-occurrences.

The number of times two people co-occurred in a conversation was used to represent a value for each tie. This weight of ties indicated the frequency of dyadic co-occurrences. Edge weight distributions thus demonstrated the spread of co-occurrence frequencies: from minimum, to most likely, to maximum. Although means are reported, they are not informative for both network degree and edge weight distributions. Network distribution, particularly in social network and online networks tend to follow a power law structure: with few people having a very high number of connections, and many people having a low number of connections (e.g. Adamic & Huberman, 2000). Hence, a relatively few number of people having a very high number of connections can skew the mean. Thus, for the descriptive overview of edge and degree distribution and their interpretation in the present study, median was considered a more appropriate calculation.

Graph Centralisation. Individual degree centrality represents the number of connections an actor has. The distribution of individual degrees offers an overall view on how many people had how many connections in the network. To understand the overall network structure resulting from these individual-level measures, graph

centralisation was used (Monge & Contractor, 2003, p. 44). This network-level measure encapsulated the information about each network's interconnectedness.

The application of graph centralisation measure can be explained as follows. Considering a lecture as representing a communication network, a teacher would be at the centre, and would be the only actor connected to all other actors in the network (students attending the lecture). The number of her connections would be exclusively high, while the number of connections between all the other nodes would be low. To understand the extent of uniqueness a powerful position may hold in a network, Freeman (1978) introduced centralisation measures that evaluate the differences between individual positions of actors in a network. The graph centralisation of such a network would be 1, representing stark differences in the number of connections between all actors. Were the teacher to be excluded from the communication network for the sake of the present example, and that the imaginary classroom would be represented by sets of group work in triads, such a network would have all actors with the same degree of 3 – they would be equally homogeneous. The graph centralisation of such a network would be low, indicating distributed nature; that is, many nodes sharing similar structural positions, rather than one or a few nodes being very central. It should be noted that in either case, although the measures represent a structural position of power, the meaning of that power needs to be interpreted in relation to the context. In other words, interpretations of network centralisation are relational to the contextual opportunities to which the central actors are exposed, as well as the constraints.

Centrality Measures. The number of nodes an individual is connected to can be used to represent a node's level of importance in a network. As such, the degree of centrality is helpful for identifying where the locus of power and influence resides in a particular network structure. However, other centrality measures can be employed to illustrate advantageous structural network positions. These include betweenness and closeness centrality measures. These measures are applied at an individual level. Betweenness centrality evaluates individual position in relation to brokering between otherwise disconnected clusters and nodes in a network. For instance, if two social groups that do not interact with one another establish one shared member, then this actor's structural position is considered exceptional and represents a high-brokering role between disparate groups (high betweenness centrality). In essence, an actor representing high betweenness is exposed to what is being shared through the network ties in both groups.

Closeness centrality evaluates the number of steps it takes from one person to reach another. That is, an actor's closeness centrality indicates the speed of information flow. Graph-level measures of betweenness and closeness centralisations inform about the variation of node-level metrics of betweenness and closeness centrality. For a graph-level measure, closeness in the number of connections that exist between participants, linking them directly, denotes a network's "compactness" (Wasserman & Faust, 1994).

5.3.4 Statistical Analysis of Network Measures

A series of permutation tests were applied (1) to analyse the variance within the edge weight distributions (analysis of variance (ANOVA) permutation tests) and (2) the effect of time on the aggregation of degree and edge weight across different forum users (permutated linear mixed models). Permutation of these statistical tests was conducted to address the interdependence of network metrics. Reasons for that are described below.

Common statistical procedures for the comparison of different groups or mixed models do not normally involve permutation (i.e. random sampling from the analysed dataset through multiple iterations). However, a standard statistical estimation is irrelevant in the analyses of network data. This is because the network data violate the assumption of independence between the analysed variables. In human networks, network metrics are assumed to be interdependent, particularly in relation to network closure effects. This is exemplified by the following: if A is friends with B, and B is friends with C, then A and C are likely to eventually be connected by a tie resulting from the fact that they both know somebody in common. In other words, both the degree and the edge weight between network nodes are interdependent measures (as derived from inter-dependent social processes). Such implicit dependencies within network data prevent the application of common statistical tests directly.

Robins (2015) has cautioned network researchers from applying statistical tests on network data referring to it as "trying to fit existing standard network techniques into an analytic world for which they were not originally intended" (p. 182). However, the particular use of permutation tests in networks that may include both dependent and independent data are acceptable (see Chapter 18 in Hanneman, 2005). Likewise, the use of bootstrapping for purposes similar to the one in this study, i.e. inter-group

comparison of measures derived from descriptive network analysis, has been recommended by network statisticians (Grunspan, Wiggins, & Goodreau, 2014).

5.4 Results

This section reports the results of the analyses of the networks established through co-occurrences between forum posters. Section 5.4.1 presents a range of descriptive social network analysis findings to describe structural features of poster co-occurrences. More specifically, the section describes the network-level structure of forum co-occurrences. It offers an overview of network sizes, interconnectedness and graph centralisation. It then describes the edge and degree distributions, representing the number of posters one co-occurred with, as well as the frequency of these encounters. Section 5.4.2 presents analyses of the changes in degree and edge distributions over time, in order to highlight how the co-occurrence between two people took place throughout the courses examined. Finally, Section 5.4.3 describes conducted statistical analyses that (1) compared the effect of time on the individual degree and edge weight of posters of different types, and (2) compared edge weights between different pairs of posters. These statistical analyses support the evidence gleaned from descriptive social network analysis that forum core posters in fact seize the highest frequency of co-occurrence with one another, as compared to other network actors, , and that their interactions evolve over time in a stronger manner that those of the other participation types.

5.4.1 Network Analysis of Co-occurrence Networks

Structure of Co-occurrence Networks. The first research question inquired about the differences between the structure of a forum core's co-occurrence network and networks of regular posters and all posters in a MOOC forum. Analyses of the forum core's, regular posters' and all posters' co-occurrence networks demonstrated that these networks share a similar structure (Table 5.4). The similarity in the networks' structures is evidenced in the graph centralisation, and metrics for the degree distributions. Table 5.4. lists the basic metrics across the networks of the various groupings, in the four analysed courses. The reported metrics include the number of nodes and ties in the networks, density, path length, centralisation, betweenness and closeness at the graph level, as well as degree and edge distributions. While the networks share some structural similarities, the size, density, and path length measures

also indicate that the forum core posters' networks had a more interconnected structure, indicating higher affordances for co-occurrence.

Networks representing the entire cohort appeared loosely coupled, highly decentralised. The overall network was characterised by most individuals being located in smaller interconnected parts of the network, linked by several individuals co-participating in what would otherwise be represented as disparate clusters. The all posters' networks had the lowest density compared to the other two groups of posters' networks. This finding is not surprising given the network sizes of several thousand people. Similarly, the average number of steps comprising shortest paths between all node pairs (average path length) is highest in the all posters' networks. The diameter, i.e. the longest shortest number of steps between two participants within a given network structure, was highest for the all posters' networks. These metrics signify that a large size of posters' networks have a disparate structure, representing low co-occurrence for the entirety of that given forum's participants.

The same metrics for the regular posters' group and forum core's group indicated that those learners had a higher co-occurrence if approached as a sub-set of forum users. The higher interconnectedness in relation to participant co-occurrence for the forum core and regular posters was observed through a higher ratio of connections made (density), shorter distance between nodes (path length), and smaller diameter. This may seem less surprising given that the number of nodes in these networks is reduced. Those participating regularly comprised, on average, 17% of the posting network *(SD*=1.8), with the forum core group's comprising an average of 11% of all posters *(SD*=3.3). However, it should be noted that these 'smaller' networks do not exist independently from their respective all posters' networks. Both the forum core (11%) and remainder of participants (89%) of the network were interacting in the same space and in the same period of time, as well as with one another. In other words, the distinction made between these groups is purely analytical, while in practice, the group delineations do not exist as all the posters interact with one another within one collective of forum posters on the MOOC forum.

Table 5.4 Descriptive statistics of co-occurrence networks in four edX MOOCs

	All Posters				Regular Posters				Forum Core Posters			
	FP	SOLAR	WATER	EXCEL	FP	SOLAR	WATER	EXCEL	FP	SOLAR	WATER	EXCEL
Nodes	1022	3221	2922	1817	177	495	569	323	161	254	340	220
Edges	8901	29366	20290	10020	2272	3765	5047	2323	2098	2430	3849	1744
Density	0.02	0.05	0.005	0.006	0.14	0.03	0.03	0.045	0.16	0.07	0.07	0.07
Path Length	2.3	3.3	3.5	2.7	1.9	2.4	2.7	2.1	1.8	2.1	2.3	1.9
Degree Centralisation	0.7	0.2	0.09	0.36	0.8	0.5	0.3	0.7	0.8	0.6	0.4	0.7
Betweenness	0.4	0.2	0.1	0.3	0.2	0.2	0.09	0.39	0.2	0.2	0.09	0.35
Closeness	0.01	0.006	0.0007	0.002	0.9	0.5	0.1	0.6	0.8	0.6	0.1	0.7
Degree, min	1	1	1	1	2	1	1	1	2	1	1	1
Degree, Q1	4	1	2	2	11	4	4	4	11	6	7	6
Degree, median	8	4	5	4	18	8	10	9	19	14	15	9.5
Degree, Q3	19	21	16.75	10	30	18	21	15	33	23	32	16
Degree, max	708	662	298	679	171	258	186	250	157	178	159	186
Edge, min	1	1	1	1	1	1	1	1	1	1	1	1
Edge, Q1	1	1	1	1	1	1	1	1	1	1	1	1
Edges, median	1	1	1	1	2	1	1	1	2	2	1	2
Edges, Q3	3	1	1	2	5	3	2	3	6	3	2	3
Edges, max	735	156	136	116	735	156	136	116	735	156	136	116

Degree centrality compares the number of connections an individual has and quantifies the discrepancies between them. Degree centralisation scores reaching towards the maximum of 1 are indicative of a core-periphery structure (Scott, 2013). Figure 5. illustrates that all posters' networks (dark black bars on the left for each course) have lower degree centralisation. Considering that over 80% of the nodes in the all posters' networks in this study represented occasional participants, such degree centralisation indicates that the number of people an actor co-occurred with was similar for the majority of participants. This is well represented in the Water MOOC where the all posters' centralisation was 0.1. Similarly, for the Solar course, the degree centralisation was 0.2.

The analyses revealed that regular posters and forum core posters have higher degree centralisation. This finding indicates that the sub-networks (regular and forum core posters) have a more distinct core-periphery structure. Given the networks' power-law distribution and low betweenness value (Table 5.4), one can understand that those who participate in most discussions, i.e. having high-brokering power in terms of the conversations they are exposed to, are also connected to the greatest number of people (degree). In addition to being brokers between conversations, these individuals also shorten the distance (path length) between regular participants

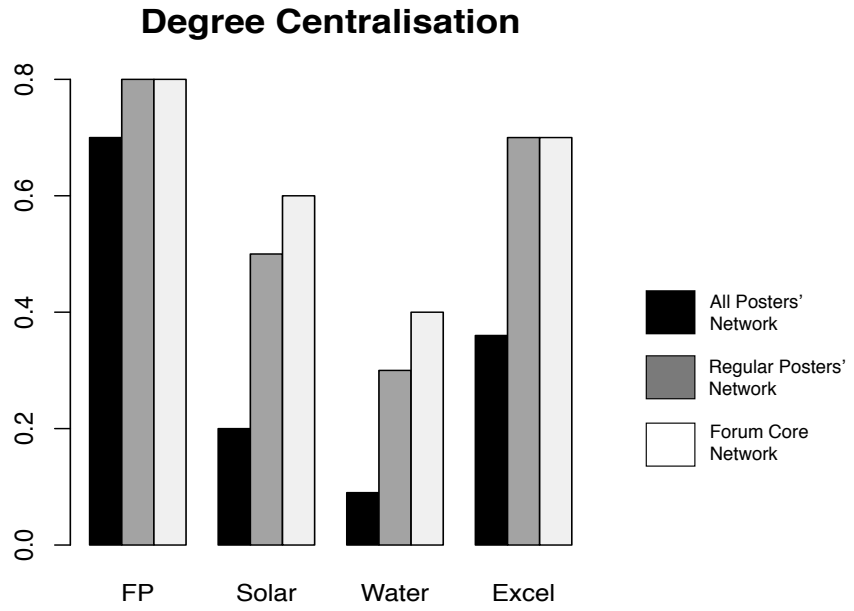


Figure 5.1 Centralisation of the networks of the entire cohort, regular posters and forum core, across four MOOCs. The highest possible value on y-axis is '1'. Light grey bars (right) represent forum core measures, grey bars (middle) represent regular posters, black bars (left) represent degree centralisation for all posters.

As Figure 5.1 demonstrates, degree centralisation metrics are in alignment with forum facilitation: with FP being highly facilitated, and Water representing an unmoderated course. Two courses with both staff and teaching assistants' involvement (Excel and FP) had the highest degree centralisation for both the all posters' and regular posters' networks. In contrast, Solar's degree centralisation was lower despite the inclusion of community assistants. The Water MOOC had an unmoderated forum and the impact of this approach is reflected in this course having the lowest degree centralisation.

Degree and Edge Distributions. An analysis of the structures of co-occurrence networks suggested that the forum core and regular posters have a higher potential for co-occurrence. This finding was further validated in the degree and edge distributions across the networks (Table 5.4; Figures 5.2–5.4). The forum core's co-occurrence networks contained a small group of highly interconnected individuals that interacted with many people multiple times, while the majority of participants across all analysed networks interacted infrequently and with only a few.

For the all posters' networks (Table 5.4), 75% of them most commonly interacted with 1-21 people throughout the course, with the median being four and five people (with the exception of FP where 75% of the people most often co-occurred with 8 people). The frequency of co-occurrence was most often once or twice, with three times in FP. To generalise, on average, the majority of actors in the all forum posters' group co-participated in a conversation with about four people and that could have happened up to two times. Such co-occurrence numerically is quite low and suggests little potential for the development of familiarity, trust or any kind of interpersonal relationship.

In the regular posters' networks, the potential for co-occurrence was higher. 75% of those posting in any three weeks of the course or more, co-participated in discussion threads with 1-30 people – a number similar to the one reported for the all posters' networks. However, the norm per person represented by the median was twice as high: from 8-10 (with 18 for FP). In other words, participating more often resulted in co-occurring with more people. The fact that the frequency of such co-occurrence is still as low as once or twice for most of those interactions (median edge weight) indicates a low potential for interpersonal co-occurrence for most of the regular posters in the MOOCs. A low potential for co-occurrence was observed for 75% of the forum core's

networks. Here, a forum core's participant co-occurred with up to 14-19 people, with a frequency of such co-occurrences for 75% of cases still within the range of 1-2 times.

Two conclusions can be drawn from the results outlined in this chapter. First, regardless of the regularity of participation, the frequency of co-occurrence with the same people in a MOOC forum is low, according to the analysed courses. However, the exposure to more people is more likely to be characteristic of those who contribute more frequently. However, these indicators are higher for the remaining 25% of cases within each co-occurrence network. Furthermore, indicators of frequency for all posters, regular posters and forum core posters for the top quartile and maximum were in alignment. This was interpreted as forum core indicators being present within the other subsets. The maximum number of actors a participant co-occurred with is at its highest in the all posters' network. However, this co-occurrence is more meaningful among the regular and forum core actors. That is, some participants in the forum core's and regular posters' networks co-occurred with almost every actor within that network. It is likely that these same people have also co-occurred with many other occasional posters – thus, the highest number of co-occurrences rose up to 600-700 in all posters' networks. It is also worth noting that 25% of co-occurrences between forum core participants took place frequently, sometimes as frequent as more than 100 times for the same two people. While some of these co-occurrence frequencies are extreme, Figure 5.2 and Figure 5.3 demonstrate the spread of both degree and edge weight distributions across the courses with outliers removed. Outliers in these cases were fewer than ten people in each course.

The details regarding degree and edge distributions reported above show that an overall potential for the formation of interpersonal relationships through shared posting activity is low, but not absent. Figure 5.4 evaluates, per person, an average of co-occurrence frequency, slightly different from Figure 5.3 that offers a distribution of co-occurrence frequency in the entire network, not an average per actor. On average, observations reveal that the top 25% of the forum core's participants co-occurred with 18-33 people (Table 5.4, degree Q3). These individuals also demonstrated a high level of regularity of co-occurrences by co-participating an average of 2–6 times, depending on the course. The series of Figures 5.2, 5.3 and 5.4 also indicate that there are course-contingent differences, and again, that courses with a higher teacher and teaching assistant involvement generally had higher co-occurrence indicators.

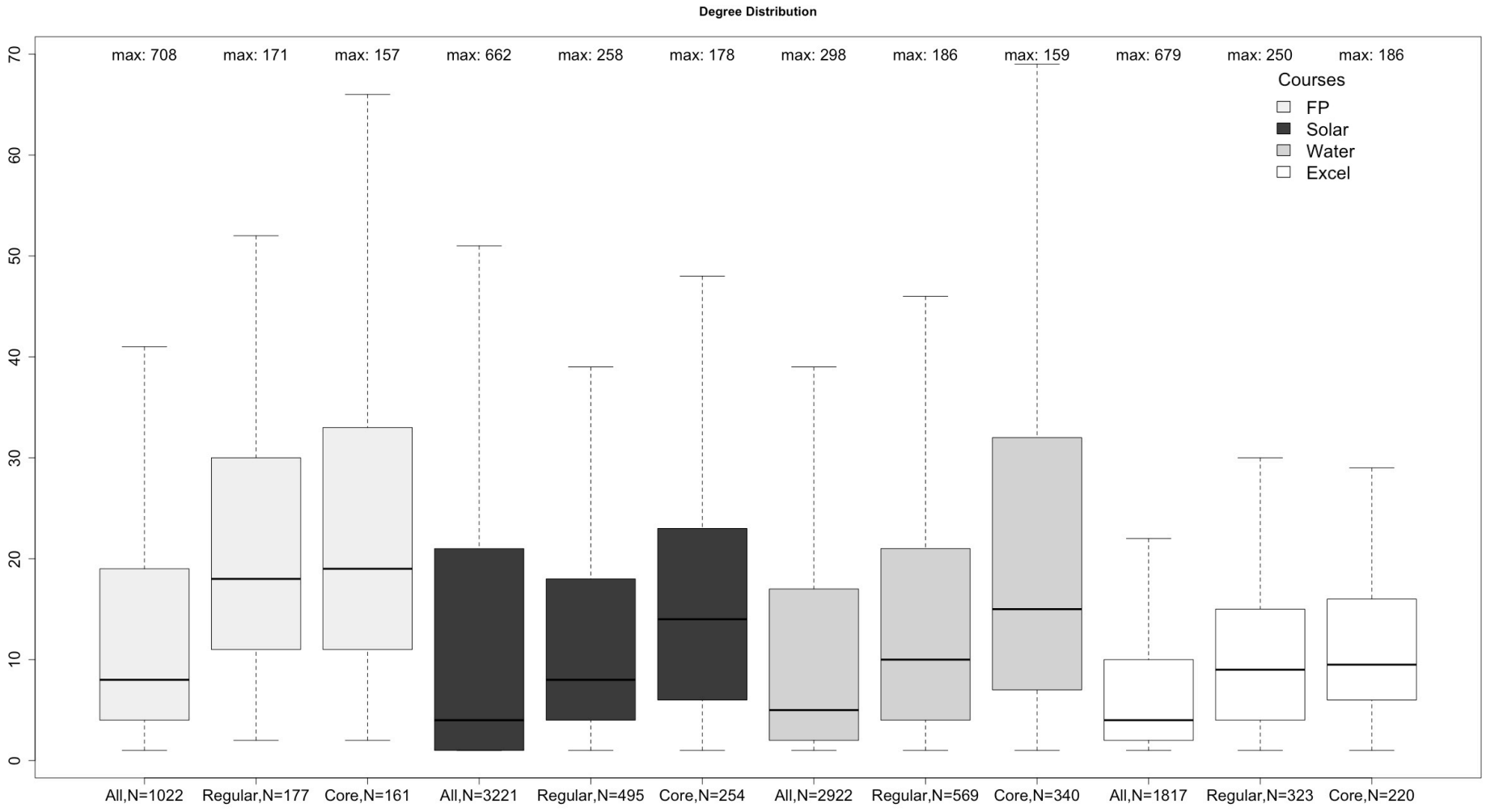


Figure 5.2 Degree distribution across the networks, with outliers removed. Maximum degree noted on the top of the figure.

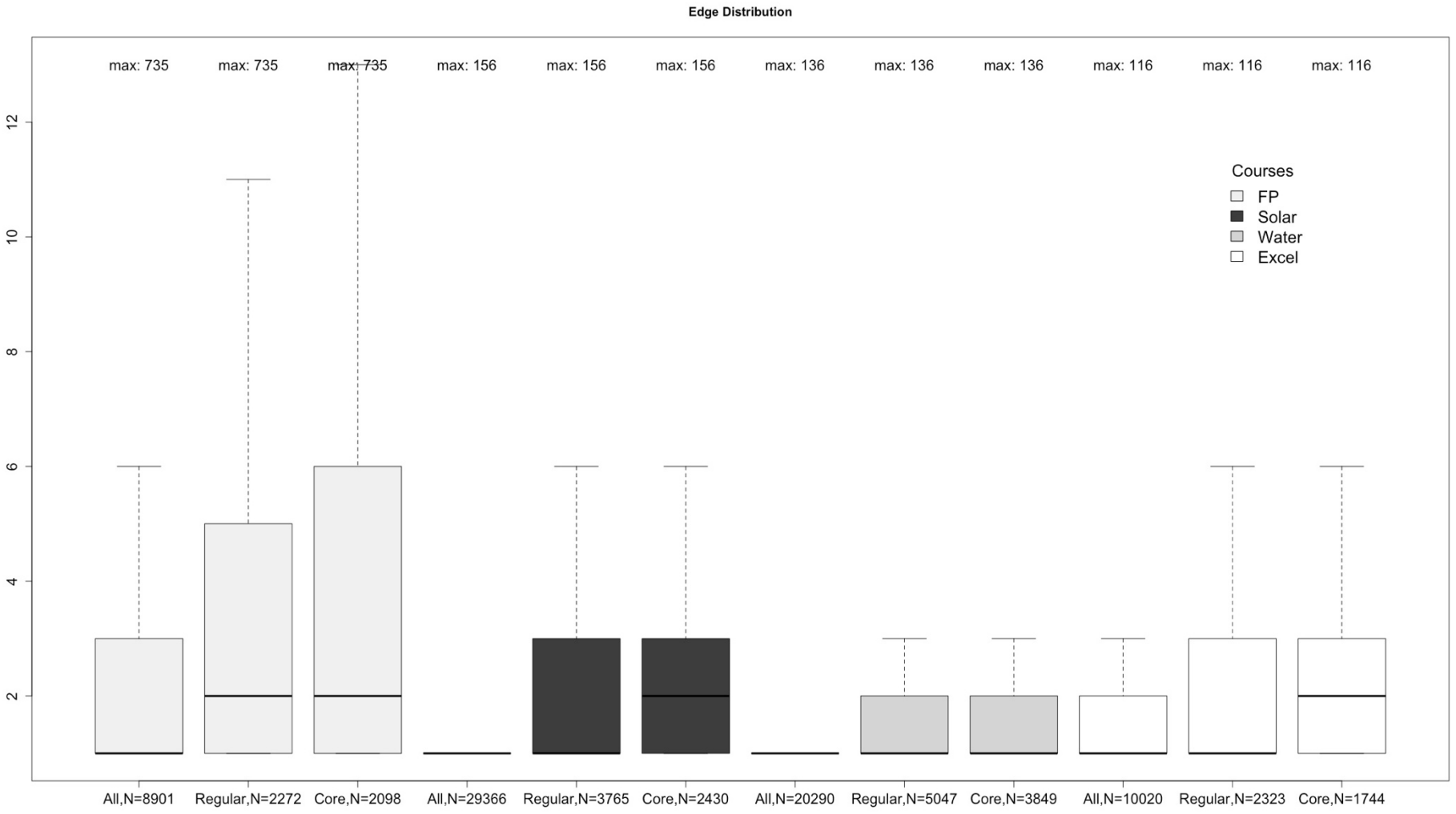


Figure 5.3 Edge weight distribution across networks, with outliers removed. Maximum edge weight noted on the top of the figure.

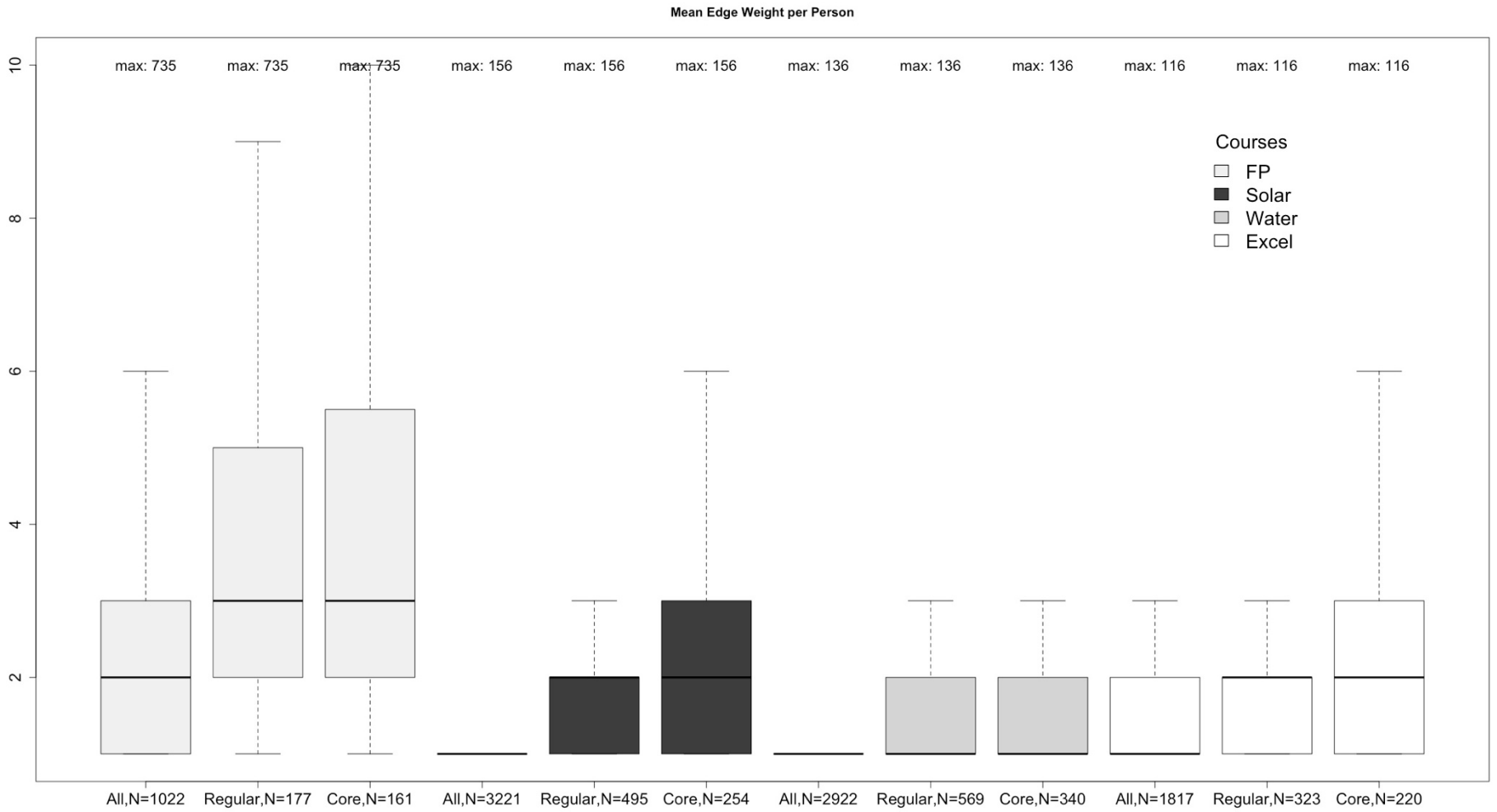


Figure 5.4 Distribution of average edge weight per person in each network, outliers removed. Maximum noted on the top of the figure

Given that 25% of a forum core in these cases comprised 30-50 people, the findings suggest that there is an establishment of interpersonal relationships within the MOOCs. The indicators above should also be understood as situated within a viewing activity that was not included in these analyses. That is, some people also posted in single conversation threads several times, and therefore, also had a high viewing rate of the various messages, given the forums' structure format (post/reply).

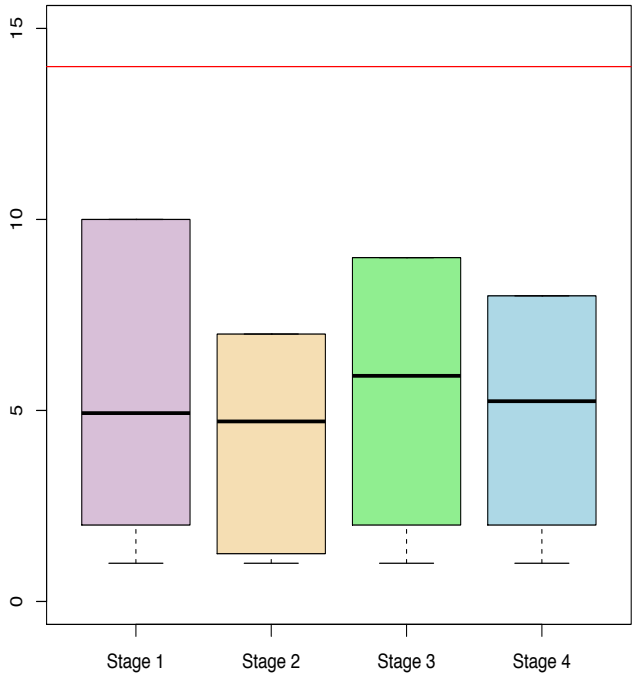
5.4.2 Co-Occurrence Over the Course of a MOOC's Progression

To better understand the dynamics of co-occurrence over time, degree distributions for each of the courses were analysed at a stage (time) marked by the implemented assessment (see Methods, Table 5.3). A boxplot representing degree distribution and edge distribution for poster co-occurrence for the entire cohort and forum core participants was undertaken exclusively for each stage (Figure 5.5). The observed patterns of development across the courses were similar, and are exemplified in Figure 5.5 for the Solar MOOC.

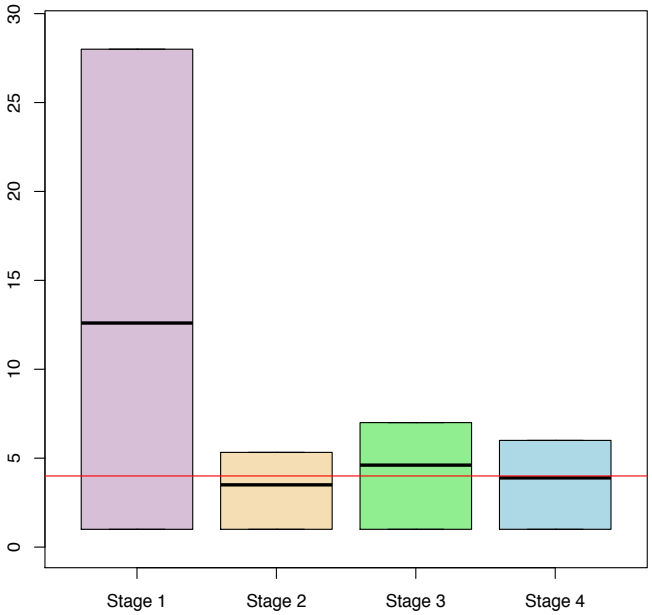
Analyses of degree distribution over time indicates that the overall level of activity across networks was stable, with people co-occurring with a similar number of people at each stage of the courses. Figure 5.4 represents degree and edge weight distributions of co-occurrence for the all posters' groups of the Solar MOOC, across four course stages. Depicted is that the overall number of people co-occurring weekly remained stable and similar from stage to stage. The first week yielded the highest median number of co-occurrence due to the introduction threads and higher number of people sampling the forum. The lower edge weight in the first week is also explained by the shallow nature of discussions in the introductory week. This then slightly grows over time (note that the scale for edge weight plot differs from the degree one).

The red line in both plots represents the median for degree and edge weight, respectively. This median was derived from the analysis of the entire course's co-occurrences. For the entire posting cohort, the median degree was the same from stage to stage, and remained the same for the entirety of the course. The median edge weight however was lower than weekly frequencies of co-occurrence. That is, though sometimes some posters may have been interacting more frequently in a given week, they were not interacting with one another again in the following weeks. In other

words, co-occurrences from week to week in the entire forum were between different people in each network segment, and not aggregate.

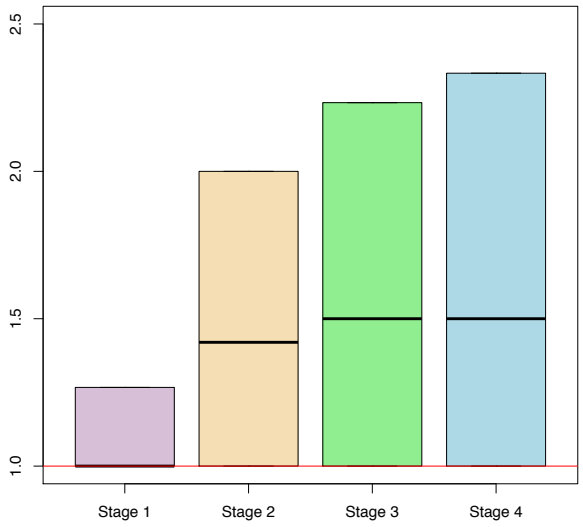


a) Degree Distribution at Different Course Stages, Forum Core Network

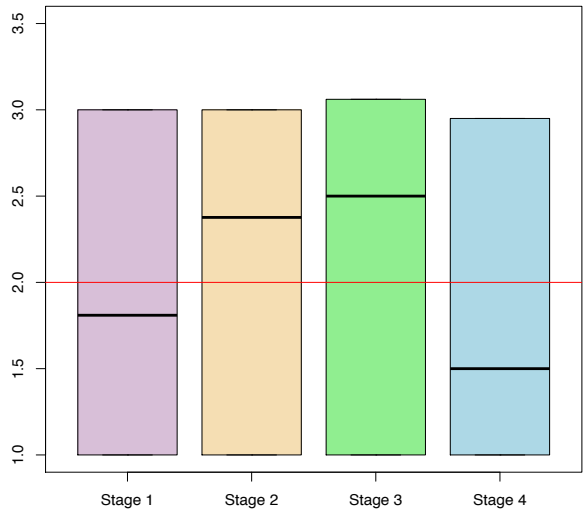


b) Degree Distribution at Different Course Stages, All Posters' Network

Figure 5.5 Accumulation of edges and ties in the co-occurrence networks in a Solar Course. Red lines in all graphs represent a median of either degree or edge weight in the forum core and all posters aggregated networks of the entire course. In other words, red median represents the outcome of the interactions that are otherwise represented as time snapshots.



c) Edge Weight Distribution at Different Course Stages, Forum Core Network



d) Edge Weight Distribution at Different Course Stages, All Posters' Network

Figure 5.5 (continued). Accumulation of edges and ties in the co-occurrence networks in a Solar Course.

In contrast, for regular posters, while the degree and edge weight remained similar from stage to stage, the median of the overall number of co-occurrences and frequency of co-participation was higher for the entirety of the course. In other words, the same people repeatedly interacting with one another from stage to stage was detected. The implication thus, is that despite the seemingly low co-occurrence activity on a stage-to-

stage basis, forum core participants accumulated their shared co-occurrences over time, with a shared history of the course and a shared history of forum interactions. These patterns repeated in all four MOOCs (see Appendix D).

5.4.3 Effect of Time on Degree and Edge Weight Accumulation across different Forum Participation Types

To understand the effect of time on the individual participants' degree and edge weight at the beginning of each MOOC, in the middle of each MOOC and at the end of each MOOC, permutated linear mixed models analysis has been conducted.

Degree Accumulation Model. For the degree accumulation models, each learner degree at the end of the first week of the course, their degree by the end of the middle week of the course (cumulative), and their degree by the end of the last week of the course (cumulative) was normalised within each course. Then the datasets of four courses were merged for the analysis. The final model reported below included random effects of the student and the course. The fixed effects within the model represented the interaction between the type of participation (regular, occasional or core), as well as a time in the course (beginning, middle, end). Normalised learner degree served as a dependent variable in the model. Such dependent variable represented overall number of people a learner interacted with in the course (breadth of interaction).

The likelihood ratio tests indicated that *Interaction of Forum Participation Type and Time Degree* model yielded a significantly better fit than the null model with χ^2 (13) = 2797.9, p =0, R^2_m = .03, R^2_c = .89. Bootstrapped R-squared (calculated from a 100 iterations of randomly sampled dataset of 2000 rows from N=19351 dataset) for the *Interaction of Forum Participation Type and Time Degre*e model predicting the degree was R^2_c(95CI): [.92, .94].

Edge Weight Accumulation Model. For the edge weight accumulation models, a mean of each learner's edge weight was calculated at the end of the first week of the course, by the end of the middle week of the course (cumulative), and by the end of the last week of the course (cumulative). All the individual measures were normalised within each course. Then the datasets of four courses were merged for the analysis. The final model reported below included random effects of the student and the course. The fixed effects within the model represented the interaction between the type of participation (regular, occasional or core), as well as a time in the course (beginning, middle, end).

Normalised mean edge weight of each student served as a dependent variable in the model. Such dependent variable represented the overall intensity of learner interactions with the same participants (depth of interaction).

The likelihood ratio tests indicated that *Interaction of Forum Participation Type and Time Edge Weight* model yielded a significantly better fit than the null model with χ^2 (13) = 4491, p = 0, R^2_m = .13, R^2_c = .82. Bootstrapped R-squared (calculated from a 100 iterations of randomly sampled dataset of 2000 rows from N=19351 dataset) for the *Interaction of Forum Participation Type and Time Edge Weigh*t model predicting the degree was R^2_c (95CI): [.83, .93]

Table 5.5. Bootstrapped coefficients (SD as a unit of analysis) for mixed-effects models predicting degree and edge weight from forum participation type and time of the MOOC.

Measures:	**Normalized Degree (95%CI)**		**Normalized Edge Weight (95%CI)**	
Effects:	**β**	**SE**	**β**	**SE**
Beginning of the course	-0.11***	0.06	0.18***	0.04
Middle of the course	0.9***	0.07	0.29***	0.05
End of the Course	1.4***	0.08	0.51***	0.05
Course Beginning for Occasional posters (Core posters as baseline)	-0.07.	0.06	-0.21***	0.04
Course Middle for Occasional posters (Core posters as baseline)	-0.92***	0.03	-0.49***	0.02
Course End for Occasional posters (Core posters as baseline)	-1.43***	0.04	-0.7***	0.02
Course Beginning for Regular posters (Core posters as baseline)	-0.1.	0.06	-0.15**	0.04
Course Middle for Regular posters (Core posters as baseline)	-0.87***	0.03	-0.38***	0.03
Course End for Regular posters (Core posters as baseline)	-1.31***	0.05	-0.53***	0.03

Note: * p < .05; ** p < .001; *** p < 0. Fixed effect coefficient (β). Standard error (SE). N= 2000 randomly sampled from N=19351 observations for 100 iterations.

Results demonstrate that overall time has effect on the accumulation of the degree and edge weight accumulation, with beginning of the course negatively associated with higher degrees, whereas the end of the course having an impact of 1.4 standard deviations on the degree and edge weight. Interaction effects of the type of poster and time of the course illuminate, that both regular and occasional participants overtime are

lagging behind the accumulation of the degree and edge weight, as compared to the core participants. Although this may be an expected outcome for the occasional participants: later stages of the course show higher negative correlation with their degree and edge weight suggesting that fewer of them appear as the course progresses. More surprising is that regular posters demonstrate similar dynamics as occasional posters. That is, they accumulate higher degree and edge weight in earlier course stages, and fewer of them appear at later course progression. Figure 5.6 below demonstrates both the temporal effects on edge weight and degree accumulation across the three groups, highlighting that forum core posters appear to both reach to significantly more people over time (degree) and do it more often (edge weight), whereas the change in degree and edge weight for other groups across time is less prominent.

Mixed effect modelling of learner accumulated breadth (degree) and depth (mean edge weight) demonstrates that core forum participants were much more likely to experience communication with more people and in a more frequent manner over the course duration. The modelling shows that the temporal dimension had very little effect on the accumulation of the degree and edge weight for the occasional posters. Also, although regular posters (posting in three weeks or more, without necessarily receiving a reciprocation in the same week when the post was made) had some effect of the time on their accumulation of degree and edge weight, the levels of degree and edge weight are lower than those observed for the forum core. In fact, by the end of the course, the degree and edge weight accumulation of the regular posters only achieves comparability with the levels of degree and edge weight accumulation that described the activity of the forum core by the end of the first week of the MOOCs.

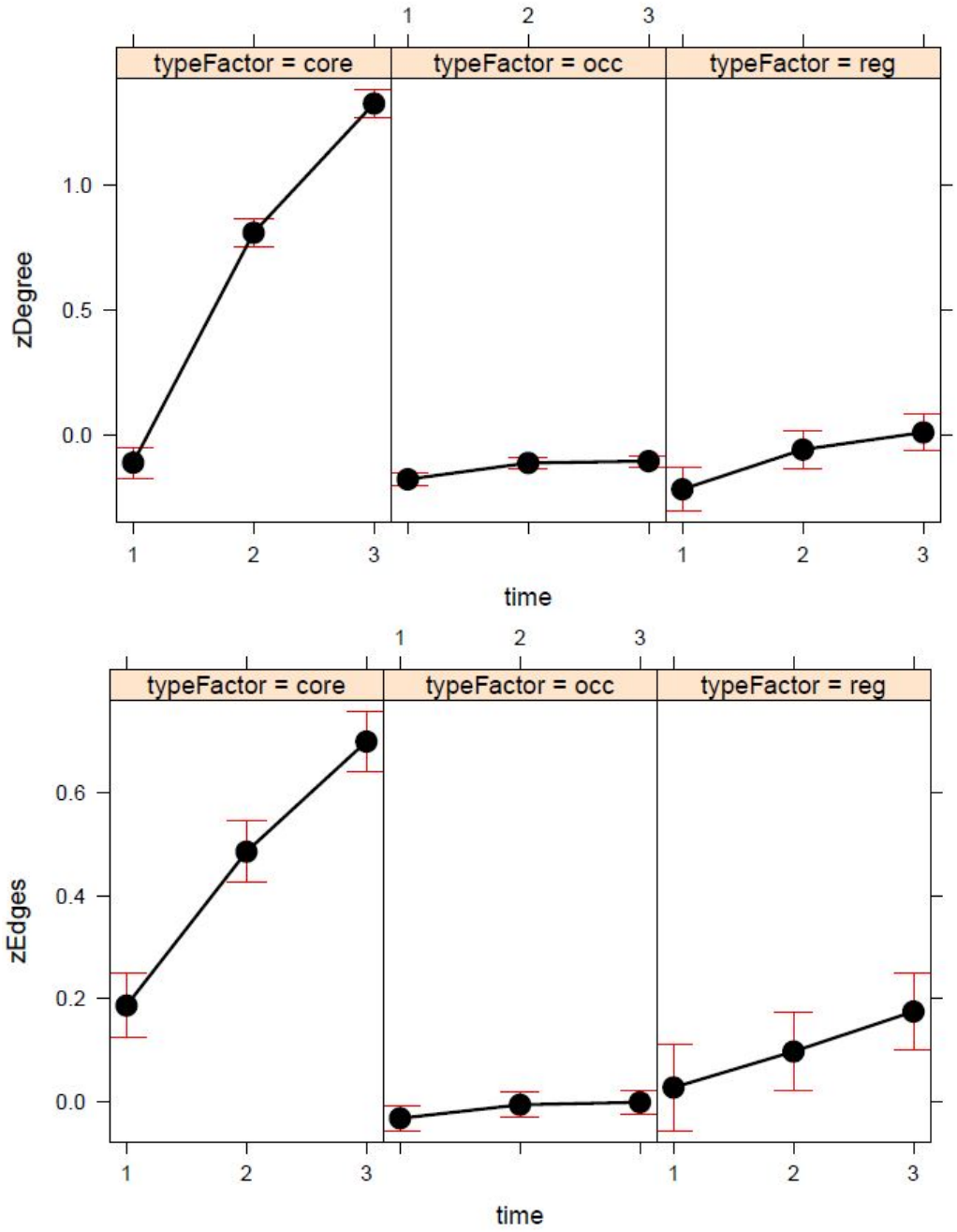


Figure 5.6. Visualised effects of time (x-axis) over normalised degree and normalised edges (y-axis, scaled in standard deviations) for different participation types (Core, Occasional Posters, and Regular Posters) respectively calculated from mixed models over the entire dataset. Bootstrapped coefficients from iterative analysis were presented in the Table 5.5 (above).

5.4.4 Differences in Frequency of Co-Occurrence

Descriptive social network analyses suggested that forum core participants have the highest frequency of co-occurrence with one another and other participants, as

compared to other types of posters. To test this hypothesis, a series of permutated analyses of variance were conducted for each of the four MOOCs examined. Edge weight representing the frequency of co-occurrence were treated as a dependent variable. Edges were grouped as independent variables, representing co-occurrences between different groups:

- co-occurrence of forum core participant with forum core participant;
- co-occurrence of forum core participant with occasional poster;
- co-occurrence of occasional poster with occasional poster;
- co-occurrence of forum core poster with regular poster;
- co-occurrence of regular poster with regular poster; and
- co-occurrence of regular poster with occasional poster.

In short, for each of the analysed networks, the frequency of co-occurrence was viewed as an outcome of interactions between particular groups.

As the data were non-parametric (i.e. non-normally distributed, and violates the assumption of independence necessary for ANOVA) an analysis of variance (ANOVA) was not feasible. To overcome these violations, a permutated ANOVA test was applied. A permutated test assumes that the weight of edges between groups belongs to one sample, and keeps reshuffling the labels of the group categories where the edge weight belongs to, while comparing them to one another. The results are used to indicate whether there are observed differences between the sampled groups, or if an edge weight belonging to different categories of posters can be a part of the same distribution. Significant differences between the groups give rise to disproving the null hypothesis that the analysed sub-groups are equal to.

The null hypothesis that there are no significant differences between the edge weight in the analysed groups was not supported in any of the courses. Permutated ANOVA demonstrated significant differences between the groups after 500 iterations in the Solar course (df=5, p<.000), Water course (df=5, p<.000), FP course (df=5, p<.000), and Excel course (df=5, p<.000). To understand which groups are different, bootstrapping estimation of mean in each of the analysed relational groups was applied as with a strategy to mitigate the effects of outliers in non-parametric and dependent samples.

Table 5.6 presents an overview of the results. Forum core participants had the highest level of co-occurrence in the network. Furthermore, it appears that the forum core had the highest co-occurrence with all other participants – emphasising the strong role these learners played in network formation. Surprisingly, those who posted regularly experienced only a marginally different frequency of co-occurrences with other posters, even in the courses where regular posters' groups were almost in alignment with the forum core's.

Table 5.6: Bootstrapped means for co-occurrence frequencies in various posting groups

	Co-Occurrence Frequency					
	Core to Core	**Core to Occasional**	**Occasional to Occasional**	**Core to Regular**	**Regular to Regular**	**Regular to Occasional**
SOLAR						
N*	2430	6734	16138	1176	158	2730
CI**	[3.32, 3.89]	[1.69, 1.8]	[1.17, 1.19]	[1.66, 1.84]	[1.13, 1.39]	[1.14, 1.19]
WATER						
N	3849	5466	7955	1071	127	1822
CI	[2.52, 2.87]	[1.6, 1.75]	[1.2, 1.24]	[1.4, 1.7]	[1.1, 1.3]	[1.21, 1.32]
FP						
N	2098	4570	1950	172	2	109
CI	[6.68, 9.09]	[2.5, 2.7]	[1.5, 1.6]	[2.7, 4.3]	[1, 2]	[2,3]
EXCEL						
N	1744	3691	2171	545	34	414
CI	[2.93, 3.52]	[1.55, 1.64]	[1.19, 1.24]	[1.76, 2.05]	[1.14, 1.44]	[1.26, 1.43]

Note: *N – the number of observations within the groups (number of edges of a particular type). **CI – 95% confidence intervals for the bootstrapped mean for co-occurrence frequencies, derived from 1000 iterations. Permutated ANOVA was conducted for each course and resulted in significant differences between the groups.

5.4.5 Summary of Results

The research question addressed served to investigate the type of posting activity that can yield opportunities of co-occurrence for MOOC learners. To address this question, network analytical techniques were used to analyse the co-occurrence networks of all posters, regular posters and forum core posters. The degree distribution and accumulation of edge weight throughout the progression of the MOOCs were examined. The frequency of co-occurrence between individuals was statistically compared

between the different groups of posters. Throughout these series of analyses, forum core posters were observed as a group of posters with a higher potential for interpersonal relationships. In addition to having a higher frequency of co-occurrence, forum core posters' ties were accumulated over time. The accumulation of the edge weight over the progression courses resulted in a temporal quality underpinning the ties. This temporal quality could indicate an exposure to a shared history underpinning the forum activity. In summary, network structures demonstrated that despite a low potential for interpersonal relationships, a network's forum core participants may experience interpersonal interactions that yield emotional benefits, such as familiarity and trust. The results and findings lead to further questions that require a more qualitative analysis of the interactions taking place between forum core participants.

5.5 Discussion

This chapter analysed the development of co-occurrence networks (entire cohort, regular posters, forum core posters) evolving from participation in MOOC discussion forums. The analyses demonstrate that the network structures observed between these groups are similar. Regardless of whether forum participation has been sustained or occasional, the networks representing interpersonal interactions are loosely connected, clustered, with hubs of activity, linked by individuals who have a higher degree of participation. Similar to the findings by Gilliani and Eynon (2014), analyses showed that structurally limited conversations occurred in fragmented groups, and a small group of people participated across them. In these disparate conversations, around 75% of an entire cohort, including those who posted regularly, were likely to have one-time encounters with the same person. While it is evident that the vast majority of connections made in MOOC forums could be classified as weak and infrequent, a quarter of the interactions in the MOOCs' forum core's networks were of a recurrent nature. Within this group of posters there were pairs of individuals who interacted with each other in over a hundred instances. The high level of co-occurrence suggests that among a diverse and disparate network, strong relationships are being established.

Although co-occurrence networks resemble each other structurally, they may be characterised by different modes of peer production processes. Haythornthwaite (2009) explained that in open online environments, individual commitment to collaborative knowledge production ranges from lightweight to heavyweight. Within this continuum,

a heavyweight-mode represents strong-tie affiliation with members, its purpose and peer-negotiated norms. From such a standpoint, infrequent ties formed through forum activity signal lightweight participation made up of interest-based contributions with a low-level commitment to maintaining or creating relationships. Given that most ties between participants seem to be one-time occurrences, learning in a MOOC forum can be described as 'learning in the crowd'.

While the commitment to social interactions for most participants in centralised MOOC forums can be regarded as impersonal and lightweight, there are also frequent interactions among the same individuals. A quarter of interactions taking place in a forum core's network may be indicative of strong ties typical for a heavyweight commitment to forum participation. Heavyweight peer production refers to sustained contributions to a perceived community, as well as monitoring its viability (Budhathoki & Haythornthwaite, 2012). This study demonstrates that forum core networks contain more active and engaged participants. They are central to forums, and are thus more likely to have frequent encounters with their fellow learners. Thus, in addition to being active contributors, forum core participants are capable of brokering information between discussion threads. Further, the frequency of their co-occurrences in forum core networks is accumulated over time. Such co-occurrences may result in a shared history, and may shift from impersonal contributions to ones where participants identify each other. It is also plausible that norms of behaviour, interaction and participation were negotiated in the examined MOOCs, through this shared history. Processes of repeated interaction, norm negotiation and commitments to the quality of collective products are atypical to crowds, but characteristic of communities.

Thus far, the claim that networks represent overlapping social entities defined by different social processes has been assumed. Characterising the content of the more frequent ties was beyond the scope of this chapter and is undertaken in Chapter 6. The strength of the ties is not necessarily reliant on the frequency of relationships, but more so on the quality of the interactions (exchanges). The study reported in this chapter did not offer insights into the nature or quality of more frequent dyadic relationships developed in MOOC forums. More frequent ties could be sites for a higher percentage of knowledge construction incidents than the low 7% observed by Kellogg et al. (2014) for an entire cohort. It is also reasonable to expect that individual active students are the hyperactive individuals keeping the spirit of the forum (Huang, Dasgupta, Ghosh,

Manning, & Sanders, 2014; Papadopoulos, Sritanyaratana, & Klemmer, 2014). Milligan's (2015) measures of forum use effectiveness may be interpreted as indicating that learners with more sustained interactions are more proficient in learning from many others. However, these considerations drawn from the literature are mere extrapolations of the findings from research spanning an entire MOOC cohort without distinction of the forum participants. Further inquiries are required to identify how the strength of a relationship between individual actors in a MOOC influences the quality of discussions and depth of knowledge construction.

Some practical implications, such as recruiting active forum core posters who broker between a forum core and other forum posters, could be inferred from the analyses, and may assist with network formation. However, it is not clear what these more co-occurring individuals experience throughout the duration of a MOOC, and if this extends beyond to-the-point informational exchanges known to describe MOOC discussion threads (Anderson, Huttenlocher, Kleinberg, & Leskovec, 2014). Examinations of the nature of network exchanges between forum core members could illuminate if they contained social negotiations, indicative of group formation.

Networked learning provides a sound theoretical framework for describing the various overlapping relationships that co-exist in the complex social organisations that manifest in educational settings (Jones, 2004a, 2004b; Jones, Ferreday, & Hodgson, 2008). Analysing a social entity from a networked learning perspective does not bias or privilege the strong relationships that imply closeness and unity of purpose within a group of actors (Jones, 2004a, 2004b; Jones & Esnault, 2004). Prior research has well demonstrated the benefits of using networked learning as an interdisciplinary framework for the analysis of learning due to its capacity to address "multiple scales of groups at multiple granularities of analysis, with multiple methods and theoretical foundations" (Suthers, Hoppe, de Laat, & Shum, 2012, sec.1). Integral to this work has been the development of methodologies to investigate networked learning. For instance, de Laat, Lally, Lipponen, & Simons (2007a) have outlined the potential of social network analysis to inquire about the nature of networked learning, while Jones (2004b) suggested that network analyses of the links and relationships in networked learning environments need to be supplemented with qualitative analyses of their nature. Consequently, there exist a number of methodological frameworks that utilise social network analysis and complementary methods and diverse techniques to

contextualise it (de Laat, Lally, Lipponen, & Simons, 2007a; Suthers, 2015). The examination of interpersonal interactions in MOOCs from a networked learning perspective and through networked learning methodologies could further enrich our understanding of learning at scale. Such an approach would allow for the capture of social learning ties of differing strengths, as well as the defining of their role and meaning through a qualitative analysis of such ties, strengths, and socio-cultural dimensions, underpinning network structures and formation.

5.6 Conclusion

The findings in this chapter show that despite the generally low opportunities for interpersonal relationships in MOOCs, forum core posters demonstrated a high potential to establish stronger dyadic connections. The posting activity of these individuals is characterised by their sustained, reciprocated and timely interactions. About a quarter of forum core participants co-occurred more frequently over the full duration of their course. The level of reciprocated interactions suggests that their dyadic co-occurrence may well translate into both interpersonal trust, underpinned by a social heavyweight commitment to participation. However, the level of relational quality that is established requires further investigation.

The next chapter continues to investigate the nature of the interactions between the MOOCs' forum core participants. Rooted in the assumptions about the socio-cultural nature of interactions, the next chapter moves beyond an analysis of student activity observed through the volume logs of students' postings. To complement indicators of co-occurrence frequency, the nature of the forum core's network exchanges is examined through an analysis of the meanings and issues that drive group formation.

6 Forum core and groups in formal online courses

This chapter analyses different discussion types driving the formation of the forum core. The findings highlight its emergent nature. The chapter also discusses the similarities and differences between forum dynamics in formal education contexts and MOOCs. Discussions signifying the presence of relational processes are further examined in the following Chapter 7.

6.1 Introduction

Chapter 5 established that forum core posters (i.e. regular posters receiving timely replies) have increased opportunities for establishing interpersonal connections. These increased opportunities are embedded within the frequent co-occurrences between the various forum core posters. In addition to a higher frequency of co-occurrence, forum core posters sustain their interactions with a group of "familiar" learners over the progression of a MOOC. It is hypothesised that the development of a temporal quality of forum core co-occurrence assists forum core posters to develop a stronger interpersonal impression than other learners and develop a more personal connection. If so, the interactions between learners could obtain a relational quality of trust, familiarity, or comradery associated with communities of learners. To gain further insights as to whether or not communities develop among cohorts of thousands of learners, this chapter enquires about the presence of relational processes within the forum core's discussions.

To ascertain the presence of relational processes, network analytic approaches employed in Chapter 5 are complemented by the qualitative dimensions of network ties. Chapter 6 analyses the content of the discussion threads underpinning the forum core co-occurrences. The integration of network analytic with qualitative content analyses can offer insights into the learning context. The nature of the learning situation can illuminate why an interaction resulting in learner co-occurrence took place. Contextualising interactions according to the purpose of communication is also aligned with a networked learning approach (Jones, Ferreday, & Hodgson, 2008). Understanding or gaining insights into the kind of information exchanged through co-occurrence ties can reveal that both the context and content of the learners' discussions

(de Laat, Lally, Lipponen, & Simons, 2007; Haythornthwaite & de Laat, 2012). According to Haythornthwaite (2005), such interpersonal exchanges may contain for example, social and professional norms, explanations of skill, or an offer of social support. This information, support processes or other goods can be seen as exchanges through discussions of a different nature, and can signify the rationale for communication.

In the context of this thesis, the presence of discussions containing off-task communication indicates the occurrence of interpersonal exchanges of a social and interpersonal nature. In formal online education, off-task communication has been taken to indicate that more personal communication is taking place within a group (Gunawardena, 1995; Walther, 1996). Although off-task communication does not bring the learner immediate informational benefits, it has a certain relational quality of sociability that can aid in community development (Kreijns, Kirschner, & Jochems, 2003). Hence, shifts in discussions within a forum core, particularly between on-task and off-task communications, indicate the presence of relational processes (Section 2.3.5), and therefore the possible evolution of a community of learners.

This chapter, describing the content underpinning the forum core's co-occurrences, is structured as follows. First, the role of off-task communication as indicators of relational processes is presented in Section 6.2 through a review of research on computer-mediated communication (CMC) and content analyses in MOOCs. Section 6.3 presents the chapter's research questions. Section 6.4 describes methodological considerations, such as the unit of analysis, the process of qualitative coding and the coding framework used for content annotation. Section 6.5 details the results of the chapter. The types of discussions underpinning the forum core's co-occurrences at different times in the courses' offering are outlined. Section 6.6 discusses the results in relation to the prior literature, and compares the nature and development of the forum core to a community of learners in formal online classrooms.

6.2 Analysing Student-Produced Text in CMC and MOOCs

Before presenting the choices used to annotate the content underpinning the forum core's ties, this section reviews: (1) studies that have employed content analyses in CMC environments to understand patterns of the emergent social context in formal courses; and (2) content analyses of MOOC forums. The first part of the review

demonstrates the significance of off-task discussions, and unfolds the meaning of social presence theoretical construct. The second part of the review demonstrates that the current MOOC-forum literature fails to offer insights about the social context of such a setting owning to the focus on content-related, on-task posts, or other decontextualised content analysis. Understanding the state of the art in CMC research in formal courses and MOOC forums positions this chapter's contribution within the broader existing body of research.

6.2.1 CMC and the Development of Social Context

Online education researchers have employed content analysis methodologies to investigate learning and communication processes within text produced by students (e.g. De Wever, Schellens, Valcke, & Van Keer, 2006; Schrire, 2006; Strijbos, Martens, Prins, & Jochems, 2006). The first attempts to research electronic student-produced texts highlighted the availability of a rich source of data that could provide new insights into learning processes. The optimism surrounding this early research into electronic student textual artefacts is well reflected in Henri's (1992) claim that such sources of data are "a gold mine of information concerning the psycho-social dynamics at work among students, the learning strategies adopted, and the acquisition of knowledge and skills" (1992, p. 118). Since the initial analyses of electronic student texts, a vast number of frameworks have been suggested. These frameworks aim to make sense of the rich insights student-produced text can offer. For example, Weltzer-Ward (2011) noted that between 2002- and 2010, some 56 independent coding schemes were in use for the analysis of content evolving from online forums. Other examples of coding schemes included cognitive and meta-cognitive knowledge and skills (Henri, 1992); critical thinking (D. R. Newman, Webb, & Cochrane, 1995); phases of knowledge construction (Gunawardena & Zittle, 1997; Pena-Shaff & Nicholls, 2004); presence of a community of inquiry (Archer, Garrison, Anderson, & Rourke, 2001; Garrison, Anderson, & Archer, 1999; Rourke, Anderson, Garrison, & Archer, 2007); perspective-taking (Järvelä & Häkkinen, 2002); and patterns of exchange (Fahy, Crawford, & Ally, 2001), among others. The frameworks addressed the varying elements of learning in a computer-mediated environment with a focus on the cognitive dimension of learning and knowledge construction.

The cognitive aspects of knowledge construction have often been emphasised, whereas affective and social aspects of learning have been often taken for granted. Kreijns and colleagues (Kreijns, Kirschner, Jochems, & Buuren, 2011; Kreijns et al., 2003; Kreijns, Kirschner, & Vermeulen, 2013) argued that computer-supported collaborative learning (CSCL) researchers tended to "limit interaction possibilities to the cognitive dimension (i.e. social interaction only serves learning) and the task context (i.e. the social interaction only takes place in the task setting)" (2013, p. 2). However, analyses of on-task settings are insufficient to explain how a social context emerges because relational processes between learners have been associated with off-task communication.

In online classroom environments, off-task communication and its gradual increase have been linked to the development of social climate and the evolution of a community of learners. Henri (1992) defined off-task communication as "part of a statement not related to formal content of subject matter" (p. 126). Early research on social climate in online education centred on the presence of social "interpersonal" messages, often around issues not directly related to tasks, in contrast to "impersonal" task-oriented communication. Interpersonal communication at the initial stages of a course was found to be low, but increased as the course progressed (Walther, 1996; Walther, Anderson, & Park, 1994). The slow development of interpersonal communication was attributed to the observation that online groups in formal courses take longer to establish than in face-to-face settings. Despite the slow pace, online learners increase their interpersonal communication over time, adding an interpersonal quality of relatedness to communications. Later studies also supported these findings by describing socially-oriented communication as increasing over time along with an increase in informal language and the development of social climate (Akyol & Garrison, 2008; Hara, Bonk, & Angeli, 2000; McDonald, 1998; Oren, Mioduser, & Nachmia, 2002; Swan, 2003a, 2003b). Through these studies, researchers suggested that interpersonal relationships in formal online courses develop gradually, and that the informality of communication and the affective quality of discussions can reflect the emergence of a social climate.

In addition to the shifts between on-task and off-task communication, the construct of social presence has been instrumental to investigations into social contexts in online education. In formal online education, social presence has been viewed as the precursor to interpersonal relationships and foundational for the evolution of group climate. Short, Williams, and Christie (1976) first described social presence as the level of "salience"

emerging from interactions between two or more participants when their communication is mediated by technology. The construct has been adopted by various scholars (Gunawardena & Zittle, 1997; Rourke et al., 2007; Tu, 2002), and generally refers to the establishment of a perception of another person as "being present and being real" (Lowenthal, 2012, p. 33). As reflected in its definition, social presence evolves from inter-personal, dyadic exchanges. For instance, Tu (2002) explained social presence as the degree "of person-to-person awareness" (p. 34). Similarly, Rourke et al. (2007) approached social presence as a perception based on interpersonal salience. It is captured within their social presence instrument, which includes statements referring to interpersonal familiarity such as "Getting to know other course participants gave me a sense of belonging in the course" and "I was able to form distinct impressions of some course participants" (Arbaugh et al., 2008). Moreover, Kreijns et al. (2013) posited that social presence, based on interpersonal relationships between learners, mediates the creation of a social space composed of learner relationships. Numerous studies have shown, the construct of social presence to be linked closely to our understanding of how a social context evolves in formal online education.

More detailed accounts about the evolution of the social context in online education have been derived from analysis of the different elements that constitute the social presence construct; for example, affectivity, group cohesion and open communication. As a course progresses, the quality of relationships underpinning learner interactions has been found to intensify, leading towards the development of a community of learners. More specifically, in an early study, Swan (2003a) used affective, cohesive and interactive indicators of social presence to quantify the change of online behaviours. In this study, Swan demonstrated that the number of affective indicators within student discourse increased as a course progressed. Affective indicators in Swan's study included features of text conveying emotions, for example, using words indicating feelings, and self-disclosure. Self-disclosure, in particular, comprised more than half of affective messages. In essence, the increase in the number of affective messages was accompanied increase by a large number of self-disclosure posts. A similar pattern was described by Akyol and Garrison (2008). In addition to the increase in affective expression, they also described a subsequent increase in the number of social presence elements related to group cohesion. These patterns of growing affectivity and growing

group-related processes reflect a dominant view of how a social context evolves through interpersonal interactions in formal courses.

To summarise, insights about the development of social context in formal online courses can be drawn from the CMC research. The findings from these works suggest that off-task communication can illuminate the development of the social context in online settings. Also, the patterns associated with social presence reveal that interactions containing self-disclosure (affective expression) increase as a course progresses. Growing affectivity involves a degree of personal emotional connection that takes longer to establish than in face-to-face settings. Once these interpersonal connections are established, social presence indicators shift towards perceptions of group cohesion. Comparing such social dynamics higher to those in a MOOC setting can help us understand the comparability between the two modes of online learning. However, as explained in the next section, current research of the content in MOOC settings has not yet offered similar insights enabling comparison between formal and open online settings.

6.2.2 Automated Content Analysis in MOOCs

Much like in formal online research, content analysis has been quickly adopted for the analysis of student-generated text in MOOC settings (Arguello & Shaffer, 2015; Brinton et al., 2013; Ezen-Can, Boyer, Kellogg, & Booth, 2015; Liu, Kidziński, & Dillenbourg, 2016; Wise, Cui, Jin, & Vytasek, 2017 among others). MOOC researchers have applied automated content analysis to help address the instructor's challenge in navigating the volume of forum exchanges. Such an approach has been shaped by instructors' need to make sense out of the vast amounts of information posted by thousands of learners. Due to the constraints imposed by the large amounts of qualitative data, manual approaches to content analysis that are characteristic of formal online education research are, at best, infrequent. Instead, automated discourse analysis and topic modelling have been favoured approaches for their ability to rapidly 'inform' or 'sort' student learning questions and support requests. For example, using automated content analysis, Rossi and Gnawalli (2014) analysed the forums from 60 Coursera MOOCs. The authors' automated approach aimed to differentiate the types of discussions that developed in MOOC forums. As a result, they noted the occurrence of discussion types, such as general discussions, study groups, assignment discussions,

logistics, and feedback. Another approach was undertaken by Atapattu and Falkner (2016) who used course curriculum content as a way to identify specific discussion threads that aligned with curriculum topics. Although these different analyses of forum content allowed instructors to capture an overview of learner discussions, they focused on different elements and aspects of the forum. As well as automating the content of the subject topics, researchers automated posts associated with particular forum behaviour, namely help-seeking and help-giving (Arguello & Shaffer, 2015; Hecking, Chounta, & Hoppe, 2016). They have also attempted to automate the functions of posts (Brinton et al., 2013). The classifications of forum posts into smaller and meaningful units for instructors were quite diverse, and the value of what instructors could actually glean from these automated analyses was not clear.

By observing the diversity of the unit of analysis in the automated labelling of forum content, Wise and colleagues (Cui & Wise, 2015; Wise, Cui, & Vytasek, 2016) questioned the generalisability of topic modelling of MOOC forum postings for gaining insights across different courses and platforms. They argued that the scope of topics in automated topic modelling has not been clearly and consistently defined (Wise et al., 2017). To overcome the generalisability problems of early MOOC research, Wise et al. (2017) suggested that content analysis in MOOCs focus on the types of posts that could be generalisable across different courses. The authors identified content-related posts as a distinctive feature of MOOC forums. They suggested that posts related to the curriculum or on-task behaviour differed from posts of a technical or social in nature. They defined content-related discussions as giving or seeking help in the learning of course material or sharing relevant resources. Overall, however, despite the benefits of focused content analysis of posts related to subject matter, the way this has been adopted throughout different studies limits our understanding of the social contexts underpinning them.

The current dominance of analysis of content-related posts (i.e. on-task communication) has diminished inquiries into other types of communication that arise in MOOC forums. The focus on content-related posts in preference to off-task communications is well illustrated in a MOOC study by Brinton et al. (2013). In this early work, the researchers established an algorithm to automate the detection of course-related discussions, including logistical queries, in ways that "filtered out" off-task communication. As Brinton et al. noted: "In the first few days, the forum is often flooded with 'small-talk',

such as self-introductions. The primary goal in this stage is to classify these threads and filter them out" (p. 347). The researchers emphasised a perceived irrelevance of such social posts by noting that reduced attention on "small talk" allows a greater focus on the remaining "valuable" threads. Similarly, Liu and colleagues (2016) suggested that any content analysis of MOOC forums needs to automate the detection of "important posts which were not yet properly addressed" (sec.48.3). In their study, the criteria for a low-relevance post clearly includes off-task communication; for instance: "1. off-topic - student has no intention to be even close to the content of the course, 2. slightly irrelevant - the relation is very superficial and main point of the post is off-topic" (p. 3). In other words, the research work to date suggests that the presence of off-task communication is irrelevant in the analysis of MOOC forums. Researchers have favoured on-task communication and subject-matter posts in their analyses.

Focusing on on-task communication and ignoring off-task communication can result in de-contextualisation of student's cognitive engagement. When analysis of cognitive engagement does not address the relationship between learning and its social context, the ability to draw broader inferences is problematic. The relationship between the social context and cognitive engagement is integral to socio-constructivist and socio-cultural theories (Section 2.3.1). The link between individual positioning within a social context and individual cognitive engagement is implicit when these theories are used in the interpretation of posts. However, more sophisticated research studies analysing levels of knowledge construction in forum posts through theoretical constructs appear to have taken this relationship for granted. Examples of research into MOOC forum content in relation to cognitive engagement, cognitive presence, knowledge construction, social presence, and similar theoretical constructs, include work by Wang and colleagues (Wang, Wen, & Rosé, 2016; Wang, Yang, Wen, Koedinger, & Rosé, 2015), Kellogg et al. (2014), Kilgore and Lowenthal (2015), Cheung (2014), as well as Watson and colleagues (2016). For example, Wang and colleagues conducted a series of studies that involved the content analysis of MOOC forum posts, their quality in relation to learning, and their association with learning gains. They first correlated the percentage of on-task and off-task posts, and identified that a higher ratio of on-task posts was associated with higher learning gains (quantified change between pre-test and post-test). However, it is unclear why the researchers separated off-task and on-task discourse for its correlation with learning gains. For the purpose of this study, it could

have been more informative to understand if learners who engaged in both on-task and off-task communication had higher learning gains than those who engaged in either one but not both. In a subsequent study, Wang and colleagues applied Chi's ICAP framework (2009) to automate content analysis of only the on-task discourses at interactive, active, constructive and passive levels (Wang, Yang, Wen, Koedinger, & Rosé, 2015). The study found that the discourses representing active and constructive discussion behaviours were strongly associated with learning gains. It is, however, unclear how the authors interpreted the levels of knowledge construction within on-task posts without acknowledging pre-existing interactions between the learners whose dialogue was being evaluated by the ICAP framework. For example, it is unclear if constructive dialogue is typical for two learners with a pre-existing history of interaction, and if constructive dialogue between learners who had only a single participation event is qualitatively different.

Similarly, a study by Kellogg et al. (2014) evaluated the content of posts through socio-constructivist theories, despite potential differences between the social contexts of formal and MOOC settings. The authors manually analysed MOOC posts for the level of knowledge construction (Gunawardena, Lowe, & Anderson, 1997). The study concluded that only half of forum discussions moved beyond the sharing of information towards entry into processes of negotiation and the co-construction of knowledge. Levels of knowledge construction were determined in relation to an entire cohort, in line with applications of knowledge construction frameworks used in educational research of formal online environments. By transferring these frameworks to a MOOC context, the authors assumed that the entire cohort they were examining was evolving socially in the same way as a group would in a formal online course. The researchers did not address the relationship between social context and cognitive engagement. In other words, it might well be the case that those learners who were interacting continuously as a sub-group demonstrated a different pattern of levels of knowledge construction.

In sum, the application of content analysis in MOOCs has advanced the automation of discussion posts that can provide instructors with practical aid in their teaching process. However, to date this work has focused on content-related on-task communication. Because prior research has failed to contextualise content-related on-task

communication, the findings do not from those studies do not offer opportunities to compare the evolution of a MOOC setting to that of a formal online course's.

Research on sentiment analysis in MOOC forums has, similarly, made few tangible insights into the development of social context in an open scaled setting, or patterns of learner discussions. Studies on sentiment analysis mine for learner emotions within the threads, without acknowledging the nature of the threads where a particular noted emotion is situated. For instance, Wen and colleagues (Wen, Yang, & Rosé, 2014) demonstrated a positive relationship between the number of negative sentiments expressed by a student and drop-out rates. However, as the authors pointed out, the results are highly dependent on the contextual conditions. For example, the simple expression of a negative emotion does not necessarily indicate a negative learner experience. The use of negative words may in fact be associated with the core vocabulary of the course (e.g. in humanities). Bakharia (2016) further demonstrated the limitations of the interpretation of sentiment analysis and generalisability of findings. To understand cross-course generalisability of text mining for affect, Bakharia replicated common text mining of confusion, urgency and sentiment. Her work demonstrated that confusion and sentiment were associated with specific curriculum issues, while urgency was associated with a specific technological issue. In other words, Bakharia's work clearly illustrates the critical need to contextualise sentiment analyses situationally. These limitations of sentiment analysis appear to be aligned with the limitations of the content-related posts – the lack of context as to why posts were made, or who the learners were who made them, and what kind of relationships connected them.

6.2.3 Summary

In conclusion, this section has presented two complementary literature reviews. A review of the CMC literature using content analysis explained that the temporal progression of formal courses is associated with a growing number of off-task communication, and growing interpersonal affectivity, both shifting a social context towards a web of relationships comprising it. MOOC literature demonstrated that current studies do not offer generalisable insights into affectivity or off-task communication, particularly its shifts within the same group of posters as a course progresses. MOOC research has failed to contextualise the insights gleaned from

advanced automated content analyses. This chapter poses a series of descriptive questions to contextualise learner discussions between forum core participants and provide insights into why such interactions take place.

6.3 Research Questions

To complement the analysis of the structure underpinning the co-occurrence in the MOOC forum discussion threads, the study posed the following overarching research question (RQ):

What is the role of on-task and off-task communication in the formation of a forum core?

To address the overarching research question, the study was conducted alongside these sub-questions:

RQ1: What discussion types underpin the forum core's co-occurrences (in volume)?

RQ): How did the nature of co-occurrence change throughout the course?

RQ3: Which were top 10 types of discussions with the highest centrality to the forum core network?

6.4 Methods

To address the research questions, a manual content analysis of discussion threads was conducted. First, the nature of the discussion threads was analysed by interpreting the content exchanged by forum core posters. The interpretation of discussion threads and the coding of the content was conducted at a group level. The unit of analysis was defined as a discussion thread with at least two forum core participants. To classify these discussion threads, the content was differentiated along three dimensions: curriculum-related, self-related and socially-related. These dimensions distinguished between on-task and off-task communication, as well as the context in relation to which they were conducted. Second, the analyses examined shifts in discussion types throughout the progression of the course since shifts in the dominance of certain types of discussion thread can provide insight into group-level dynamics. Finally, the types of discussion threads with the most forum core posters were identified. This section

explains each of these steps, and reviews the methodological considerations related to the unit of analysis, coding categories, and the process of coding.

6.4.1 Unit of Analysis

Content analysis requires certain initial methodological choices that are directed towards the so-called data reduction phase. According to Miles and Huberman (1994) this phase involves "the process of selecting, focusing, simplifying, abstracting, and transforming the data" (p. 10). Qualitative data must be transformed to make it more "manageable", and to enable sense-making in the context of the selected constructs. Broadly speaking, the analyses of transcripts was conducted at the levels of themes, discussion threads, individual messages and turn–taking (Schrire, 2006). In this study, the choices for the unit and level of analysis, and the segmentation procedure were determined by the operationalisation of constructs and the theory guiding the interpretation of written text.

Analysis of the content of forum postings was conducted at the level of discussion thread. This unit of analysis was determined from the findings reported in Chapter 5. Each tie in the network of the forum core's co-occurrences represented co-participation in a thread. In other words, the content of that discussion thread was of relevance to both participants. Thus, analysis of the discussion thread was used to help explain the nature of learner co-occurrence and its evolution over the course of a MOOC's duration.

6.4.2 Manual Content Analysis in MOOCs

Section 6.2.2 argued that research into the role of off-task communication in MOOC forums has been largely neglected. This lack of investigation into off-task communication may be due partly to the challenge of analysing MOOC forum data manually, and the inability of automated approaches to capture the situational aspects of communication beyond content-related discussions. To capture the types of discussions underpinning the forum core's co-occurrences, this study undertook a manual content analysis of forum posts.

The study's approach is built largely on the previous work by Stump and colleagues (Stump, DeBoer, Whittinghill, & Breslow, 2013). In an exceptional effort, this research group conducted a manual analysis of 96696 posts from the MIT MOOC Circuits and Electronics. Although the authors did not present the results of their analysis, they

discussed the methodological considerations adopted in their open manual content analysis of forums. Stump and colleagues (2013) carefully documented the process of open coding and interpretation of the MOOC's forum content. In their framework they distinguished between the different topics for in-forum communication: content-related posts, discussions of other courses, technical issues, posts that addressed online interface, posts that addressed course organisation, requirements, and so on. A separate category for social, emotional or community-building aspects of the class was also included.

In their reflections on the coding process, Stump's team (2013) addressed the problem of overlap between the categories during coding at such granularity. To avoid category overlap in the interpretation process, they suggested interpreting content in relation to the speaker's intention, rather than what appeared to a post's topic. The present study adopted this suggestion to interpret discussion posts in relation to learner intention. In doing so, the study related Stump and colleagues (2013) coding to broader factors that influence a learner's intention to engage in a social activity. In essence, analyses in this chapter replace the interpretations of "what the post was about" with "what intentions learners had when posting this message". Simply put, the analysis focused on why posters were using the forum, rather than what or how well they were learning.

Table 6.1 summarises the codes used for content analysis in this study. Coded discussion threads represented a learner's process of making meaning of the self, the curriculum and the environment. Discussion threads were annotated manually into three categories: (1) curriculum-related, (2) setting-related, and (3) self-related. Each of the broader categories included a smaller subset of codes that also represented (a) on-task and (b) off-task communication within a particular category.

The introduction of non-task and task–related discussions enabled the interpretation of the learner's intent. That is, the coding framework developed by Stump and colleagues (2013) was based on topics, i.e. content-related, other coursework, social/affective, course website and technology, course structure and politics, other, missing data, and posts not in English. The coding framework applied in this study embraced these topics by positioning them within the types of communication. For instance, the social affective category was split into social-task (ST) and social non-task (SN) discussions.

The characterisation of a post was dependent on whether the learner was making sense out of their own ability to perform a task, or commenting on the learning setting.

Table 6.1 Coding framework overview

	1. Curriculum-Related	2. Setting-Related	3. Self-Related
a) On-Task	**1a. Content-Task (CT) Discussions** *"How do you solve this task?"*	**2a. Administrative and Technical (INF) Queries** *"Can't submit the assignment, has the deadline passed?"*	**3a. Social-Task (ST) Discussions** *"This was the hardest tutorial I have ever done in my life."*
b) Off-Task	**1b. Content Non-Task (CN) Discussions** *"I was wondering why wind energy is not widely used in my country."*	**2b. Social Non-Task (SN) Discussions** (negotiation, group-related) *"I think we are all here to learn not to ace the content."*	n/a (These could have been posts, but not entire discussions)

Two additional categories emerged during the coding process. One, referred to as mixed discussion threads, was created to annotate the threads that combined several discussion types. For example, learners could typically have an exchange that first focused on their ability to undertake a task, then shifted to a more content-task-related exchange (CT) as to how they actually tackled it. Another discussion label, metacognitive discussions, was created to address learner discussions about strategies of learning. The established "Coding Book" is included as an Appendix E, presents the code descriptions, and specific examples. A description of each of the established categories is further elaborated on in the next section.

6.4.3 Coding Framework

Curriculum-Related Discussion Threads (Categories 1a and 1b, Table 6.1). The curriculum-related categories captured posts where the learner's intention was to inquire about the course's curriculum. Interpretation of the posts' content was conducted in relation to the learner's interaction intention. In particular, when learners sought clarification of the course concepts, lectures, assignments, and implementation of the subject matter in practice, these were interpreted as interactions related to curriculum. In other words, curriculum-related discussions revealed a learner's motivation to engage socially in aspects directly related to the subject matter, most likely also mentioned

within the course's materials. Curriculum-related discussions included two sub-types of discussion threads: content-task (CT); and content non-task (CN). The CT category was made up of discussion threads triggered by the content of a specific task. An example of a CT discussion is the learner's confusion about a graded assignment, or graded tutorial. Typically, learners would indicate the task or assignment they were inquiring about within the content of the discussion. Discussions triggered by an inquiry into a course's materials or concepts were categorised as CN. Such conversations ranged from clarification of lectures, to correction of course material mistakes, to interest-based discussion of out-of-the-course discoveries. For example, in a computer science course, CN discussion threads could include different additional resources helpful to understanding one of the key concepts in the course. In an engineering course, students could engage in discussions about the kinds of metals or plastic details for building solar panels. In a data science course, CN discussion threads could include conversing around the different versions of Excel and writing Excel commands in different languages. In other words, CN discussions are still related to and triggered by the curriculum, but are not directly linked to a perceived difficulty with assignment.

Setting-Related Discussion Threads (Categories 2a and 2b, Table 6.1). In setting-related discussion threads the learners inquired about the educational setting of a MOOC, trying to understand how the setting "works". Setting-related discussions were divided into two types: administrative and technical (INF) for close-ended queries; and social non-task (SN) for open-ended queries.

The INF category embraced informational queries posted about the educational setting. In particular, this category included discussions that prompted a closed-ended response or an answer. Used in the university environment, the range of discussions observed within this category would fall under the responsibility of an administrative student office or an IT-support desk. A subset of these discussions (technical) were triggered by misunderstandings related to the edX platform, or software that was required for an assignment. In these instances, technical queries were triggered by an individual seeking clarity on the use of the edX system in order to complete or undertake assignments or learning activities. As such, INF queries' category included conversations that typically involved issues such as how to download course videos, how to type mathematical formulas within html web-pages, how to read the progress bar, how to submit assignments, as well as common technical troubleshooting (i.e. IT-help desk queries).

Other queries in this category (administrative) were triggered by administrative issues that related to course deadlines, eligibility to start the course, course requirements, making sense of how many graded assignments were needed to pass the course, and so on. Such questions can be seen to relate to administrative student services at the university, or questions a student may ask a professor about the syllabus.

As stated, INF queries usually required a simple reply to a student concern. Hence, these queries could be described as a series of questions and answers (Q&A) around course technology and course organisation. However, learner sense–making of the educational setting of a MOOC was not limited to simple Q&A. As many MOOC forum users were adult learners, they at times wondered about larger issues within the course that may have necessitated more than one simple answer. A learner may have wondered why the content was being taught in a particular manner, or in a particular sequence, why an assessment was designed in a certain way, for what purposes and how best to use the forums, inquiring into the value of MOOC learning in relation to the learner's career aspirations, why other learners were taking the course, and so on. These discussion threads were labelled as SN as they were oriented towards the group, and were not required or necessary to complete any given task. In other words, these discussions were representative of a learner's need to reflect on the collective experience of learning in a MOOC around a perceived shared aspect. These SN discussion threads were generally open-ended. They included negotiations of how to organise communication in the forums, the ways of being together in the course, whether the course should be run the way it was, the intention behind the course and the expectations brought by the participants.

The coding framework included setting-related threads, because the learning needs may include the arrangement of a MOOC setting (INF discussions) as well as reflection upon the choices and practices a learner needs to make within the setting (SN). Learning about the social setting as well as about the tools, norms, values and practices could be a part of the in-course learning experience (Section 2.3.1). In a MOOC setting, when norms, rules and practices were unclear or not in place, learners discussed them, creating a culture around forum practices.

Self-Related Discussion Threads (Category 3a, Table 6.1). A self-related discussion thread included an emotional statement that a learner shared with peers that was related

to the curriculum task they were undertaking. Although statements and discussions were triggered by a learning task, the intention of these messages was to share or reflect upon a personal skill in relation to the learning experience. This reflective process was often triggered by either an assignment or learning activity. Such discussions were coded as ST. They typically involved emotional statements as participants self-disclosed their emotional reactions to a course's assignments. The ST discussions were similar in nature to the affectively expressed CT conversations. A CT post may have included emotional content but would have been framed as a search for information, possibly for a difficult task. In contrast, an ST discussion would have been framed as a complaint about the difficulty of a task. Although both might have noted that a learner found a particular discussion difficult, the difference in phrasing provided information for determining the intention of the post. A CT discussion sought for help, while an ST discussion appealed to emotional support, venting, or reassurance. In cases where a clear distinction was impossible, the discussion was labelled as mixed (MIX).

The coding framework identified self-related threads to emphasise that learning reaches beyond content-related concepts of the course. While cognitively engaging with the concepts, individuals are also learning about themselves. Self-related learning outcomes are explained through the theories of self-efficacy (Bandura, 1982) or self-theories (Dweck, 2000). When undertaking educational tasks, individuals commonly establish a belief or perception of the relevance of the task and about how well they will perform. These beliefs about one's ability are closely tied to an individual's motivation to continue to learn and perform. For instance, learners with lower self-efficacy or fixed mindset about their own ability to learn, may not persevere in undertaking the tasks they see as difficult (Dweck, 2000). Alternatively, observing somebody else struggling with the task, but completing it, may serve as motivation for other learners (Bandura, 1982). Self-efficacy theory notes that a learner's self-efficacy and motivation to learn may increase when observing their peers achieving a level of mastery. Even though the achievement is not a direct personal success, it is in a context that the individual can identify with and relate to. By reflecting on the task and their own ability, individuals also achieve learning outcomes that become a part of their own perceptions of themselves and their own abilities, and influence their future learning experiences. In short, self-related discussion threads captured individuals' processes of reflection upon themselves, as well examples of disclosure that were offered to the group.

Other Discussion Threads. In some cases, two coders annotating a forum's content experienced some disagreement about the intention of the discussion threads (see Section 6.4.5 for details on inter-rater reliability). If none of the rules specified in the code book applied clearly, and the coders could not reach agreement, the code MIX was applied. The intention of the discussion post was not sufficiently clear to capture within the code book, without undertaking a further and more granular level of coding. The researcher opted not to move to a more granular level of coding in order to maintain a high level broad coding. Ultimately, the three high-level categories (curriculum-related, self-related, and setting-related) represented the three domains of learning outcomes. Mixed discussion threads included instances of overlap, demonstrating that separation between outcomes could be synthetic, rather than natural. During the coding process it was also noted that some courses had a much higher amount of MIX discussions. Untangling the meanings and reasons for the existence of mixed discussions were beyond the study's scope; however, to ensure that the content of the MIX discussions was addressed, such threads were added to the qualitative analysis presented in Chapter 7.

Besides the three categories of curriculum-related, self-related and setting-related discussions, both coders also came across a particular kind of discussion that was not defined before coding began. In these instances, the intention of the MOOC posters was related to a need to know how best to learn in such environments (META). The content included constructive suggestions on exam preparation strategies, how to navigate information, and how to best organise one's learning activities. These discussions reflected a poster's interest about how to learn to prepare for certain tasks or master certain content. They differed from the CN discussions about learning resources on the topic of interest. These particular discussion threads captured a learner's specific strategies for study, exams, and note-taking from video-lectures. In some ways, the posts can be considered to cut across the curriculum, the self, and also the setting-related categories. That is, they were related to studying or mastering the curriculum to facilitate academic success. The threads revealed an individual's personal learning strategies while also disclosing his or her perceived weaknesses or strategies for learning. For instance, one learner suggested downloading all the pdfs of lectures and searching across them when preparing for exams.

In conclusion, the coding framework used to annotate the forums' posts interpreted whether or not the group of forum core posters were engaging socially in order to: understand the curriculum, reflect on the self, make sense of the educational environment, or inform their process of study. These discussions took place through both on-task (CT, ST, INF) and off-task (SN, CN) communications. Although the high–level codes may not fully reflect the granularity of discussions, they do offer a generalisable framework that includes the types of discussions privileged within the current MOOC research, while addressing the off-task communication and disclosures (ST) that can inform about the social aspects of forum interactions. It should be noted that an evaluation of the quality of the posts was deemed beyond the scope of the study. The focus for the present study lay in determining whether or not the forum core's activity included any discussions beyond curriculum-related and on-task aspects, and if so, the interest was to capture the volume and dynamics of these other types of discussion threads in relation to well-studied subject matter–related postings.

6.4.4 Process of Coding and Validity

Content analysis has roots in the tradition of qualitative methods. Schrire (2006) compared it with the case study approach owing to its implicit "interpretation in the context". As a result of the interpretive elements, content analysis is heavily reliant on human coding (Rourke, Anderson, Garrison, & Archer, 2001). To ensure the generalisability and validity of content analysis, Naidu and Jarvela (2006) provided a set of guidelines that serve to ensure a high level of rigour. According to the guidelines, the unit of analysis, segmentation, and coding categories are evaluated in relation to their reliability. To ensure internal validity, particular attention is applied to match the constructs with their operationalisations. Inter-rater reliability is frequently used as a means to ensure internal validity. For the present study, this was achieved with two coders working through the samples of text (discussion threads). Moreover, as external validity is determined through the broader applications of a coding framework, in this instance, the external validity of the framework was well demonstrated through the application of a coding framework to four different MOOCs.

The process for coding the discussion threads was as follows. Each conversation shared between at least two forum core learners, (i.e. a situation that a co-occurrence tie was based on (Section 5.3.1, para.2)), was interpreted in relation to what the participants

were making sense of: the course curriculum; themselves; the MOOC setting, or the process of studying. If several of these could be applied, then the MIX category was assigned. The initial annotation of the content was approached abductively (Blaikie, 2007, p. 88). That is, the coders were interpreting the text in relation to pre-specified categories open to the emergence of new categories, based on the interaction of their prior experiences and backgrounds with the interpreted text.

6.4.5 Inter-rater Reliability

An additional coder was trained to annotate the forum discussion threads alongside the primary researcher. The individuals involved in the coding process, first annotated 100 discussion threads in each MOOC. Upon completion of the first 100, the coders compared their results through the inter-rater reliability metric of Cohen's kappa (Cohen, 1960). Any disagreements in coding categories were resolved through discussions and re-clarification of the coding protocol. Required changes or adjustments to the coding framework were made following such discussions and the resolution of perceived coding inconsistencies (Appendix E). Coding by the two individuals continued until a third of the discussion threads had been coded for each course. Cohen's kappa statistic was calculated for each of the course. The score represents the level of agreement between the coders for each sample beyond the agreement that could have occurred for the sample of this size by chance. For the present study, the kappa statistic showed substantial levels of agreement (Kvålseth, 1989): the agreement levels ranged from 0.65 to 0.71 across the four courses. As a range of scores from 0.6 and 0.8 is considered "substantial", with agreement scores over 0.8 being referred to as 'almost perfect' (Landis & Koch, 1977, p. 165), the remaining content was annotated by the researcher alone. A total of some 4260 discussion threads comprising 16195 posts were manually coded across the four MOOCs (Table 6.2).

Table 6.2. Overview of coded content

Course	All coded discussion threads, N	Posts comprising the discussion threads, N	Discussion threads coded by two researchers, N	Inter-rater reliability, kappa
FP	889	4281	300	0.67
Water	1167	4250	400	0.68
Solar	1072	3880	350	0.71
Excel	1132	3784	300	0.65

6.5 Data Wrangling

The content analysis presented in this chapter was closely related to the network analyses discussed in Chapter 5. The datasets used for the analysis of co-occurrence networks were transformed to enable the descriptive analysis presented in this chapter. This sub-section details a number of choices around data wrangling.

Dataset Structure. First, as explained in Chapter 5 (Section 5.2), for the analysis of co-occurrence networks, raw forum data were transformed into the datasets using the following structure: Learner ID, Discussion Thread, Time Stamp. These datasets were manipulated into a network structure for the analysis of co-occurrence networks (Chapter 5). That is, Learner A had a tie to Learner B, based on the discussion thread they shared. The weight of this tie corresponded to the number of times they co-occurred. Timestamps were used to separate learner interactions into the different course stages, delineated by course assessment deadlines (Chapter 5, Table 5.4).

The first research question of the study presented in this chapter inquired about the volume of the different discussions within the MOOCs. To analyse the content underpinning the co-occurrence networks constructed in Chapter 5, the same datasets were used, with the structure: Learner ID, Discussion Thread, Time Stamp. After discussion threads were manually interpreted and coded, the dataset structure looked as follows: Learner ID, Discussion Thread, Type of Discussion, Time Stamp. To address the first and the second research questions, discussion types were counted. The counts were done at a post level, whereas the coding was done at the discussion level. As a result, the counts presented in the analyses represent the volume of discussions (the number of posts comprising discussion types) of a particular type across a given MOOC. In this way, impressions of what was discussed (the volume, rather than the count of discussion threads) was quantified in ways that a learner might experience.

Temporal Boundaries. The second research question inquired about the change of discussion types throughout a course's progression. To understand how the volume of these different discussions changed, the counts of discussion posts into discussion threads were aggregated in two different ways, so as to enable in-course interpretation, as well as inter-course generalisability.

For in-course interpretations, temporal boundaries defining the different stages of a course were aligned with that course's assessments (Chapter 4, Table 6.4). The implemented temporal boundaries provided an opportunity to observe how discussion patterns ebb and flow over the duration of a course and in the context of the implemented learning design. In the present study all MOOCs presented differing assessment designs. For instance, one course was separated into three stages (three deadlines), while another had six stages (six deadlines).

Because of the differences in assessment design, the variation in temporal boundaries did not allow for inter-course comparisons. For example, one stage in the Water course lasted three weeks, whereas in Excel a stage would last one week. Therefore, to enable inter-course comparability of the patterns of discussion change, the data were re-analysed within simple yet equivalent time slices for all four courses: week 1, mid-course, three quarters into the course, and the week of the last lecture. Within the results (see Section 6.6.2), the counts of the coded discussion thread types are presented for both assessment-defined stages, and inter-course comparable stages.

Centrality of Discussion Threads. The third research question inquired about the centrality of particular discussion types to formation of the forum core network. For this purpose, the dataset of the structure "Learner ID, Discussion Thread, Discussion Type" was transformed into a bipartite network; that is, a tie between a learner and the discussion thread they posted into was generated, (Learner A made a post into Discussion B). Degree centrality was calculated for the network nodes representing discussion threads. The weighted degree (Newman, 2004) of a discussion thread was used as an indicator of its centrality. A measure of weighted degree combined the highest number of individual contributions from a forum core poster being made to the thread with the highest number of forum core posters contributing to the thread. Social network analysis analyses were conducted using the igraph package (Csardi & Nepusz, 2006), and data wrangling was conducted in R (R Core Team, 2014).

6.6 Results

The overarching research question posed in Chapter 6 inquired about the role of on-task and off-task communication, social and content-related in nature, in the formation of a forum core. The types of discussion threads were analysed as related to the curriculum (content-task and content non-task), related to the self (social-task), and related to the

MOOC setting (informational and social non-task). The volume and nature of discussions were examined throughout the duration of the four courses. The influence of the discussion types in relation to their centrality to the forum core's posters was also investigated. This section describes the specific results pertaining to each of the sub-questions.

6.6.1 Dominant Discussion Types

The first research question unpacked the types of discussion threads where forum core posters co-occurred. The study found that forum core members co-occurred mostly in curriculum-related discussion threads (in 74% of the co-occurrences), and on-task setting-related discussions about informational and administrative issues (10% of the co-occurrences). Self-related discussion threads were at a low level (4%). Likewise, the off-task setting-related discussions (7%), i.e. social non-task, and meta-cognitive discussions about learning strategies (0.5%) were less frequent. The volume of discussion threads was operationalised through the number of posts in the threads. Table 6.3 presents the ratio of posts belonging to a specific type of discussion thread in relation to the entirety of the analysed content.

Table 6.3. Overview of the nature of discussion threads exchanged within the forum core. The ratio of discussion types is calculated in relation to the total number of posts within the course, and adds to approximately 100%

	Category	Discussion Type	Ratio		Course			
			Mean, %	*SD*, %	FP	Solar	Excel	Water
1	Curriculum-Related	CT, %	45	11	39	51	56	33
		CN, %	29	14	39	17	18	43
2	Self-Related	ST, %	4	2	3	2	6	5
3	Setting-Related	SN, %	7	2.8	6	9	4	9
		INF, %	10	4	7	16	9	7
4	Other	META, %	0.5	0.5	0.6	1	0.2	0.1
		MIX, %	5	2	4	5	7	2
5	Total Posts		4112	406	4703	3865	3630	4250

Curriculum-related discussions. The results (see Table 6.3) demonstrate that more than two-thirds of the forum core's discussions were related to the curriculum; that is, the discussions threads were coded as CT and CN and served as the primary opportunities for co-occurrence. The mean ratio across the courses demonstrates that

curriculum-related discussion threads (mean CT and mean CN) accounted for an average of 74% of the 4112 posts. That is, around two-thirds of the total content generated by the forum core was triggered by issues directly related to topics of the course. First, some 45% (*SD*=11%) of the discussions across the four MOOCs were characterised as CT. Half of an average of 4112 posts in each of the forums took place in discussion threads triggered by issues, questions and concerns raised in relation to specifically graded course tasks. Second, 29% (*SD*=14) of conversations were curriculum-related issues but not necessarily aligned to graded assessments (CN).

The balance between content non-task and content-task discussions varies across the courses. There are several possible explanations. The higher number of CN can be attributed to the difficulty of concepts in the lectures (FP), or the presence of general topics related to the curriculum encouraging hobbyist chats (Solar). The higher ratio of CN discussions in Water may in part be a result of the higher number of students present in the forum in the first week, and the lack of assignments to be submitted during that week.

Self-related discussions. Self-related discussion threads represented instances where a forum core's posters reflected upon their abilities to perform the tasks, and disclosed their emotional reflection to other posters. A typical self-related discussion would include a variation of the following: "The task was so difficult. I tried really hard, but my preparation/ knowledge/background did not allow me to do it properly". Alternatively, the learner may have said: "I love the task. This was so exciting to conduct this experiment at home. I thought it was difficult/easy but very enjoyable". In essence, these discussions invited similar reflections from other learners.

The average volume of ST discussions across the courses was four per cent (*SD*=2%) of all co-occurrences. Due to their nature, these discussions contained a high degree of self-disclosure statements. Disclosure was both positive (e.g. indicating learner excitement about the task just completed) and negative (indicating frustration with the task or an assessed concept). Self-disclosure was also found in many mixed discussion threads that contained similar self-task posts that were also included in another category type. Overall the level of self-related discussion threads appeared to be low. Considering the vulnerability associated with self-related discussion threads, this low volume suggests that learners had a low propensity for self-disclosure.

Setting-related discussions. The third major category of discussion types was triggered by the need to understand the educational setting of a MOOC.

SN discussions were triggered by some puzzlement, misunderstanding or reflection upon the social and educational setting of a MOOC. On average there were seven per cent (*SD* = 2.8) of SN threads. The topics raised in these discussions were diverse. They included discussions about the purpose of learning, reflections about the teacher or course design, and disagreements between learners on how to use the forum. These discussion threads reflected the social negotiation processes between the forum core members, that is, learner reflections on why things were done in a certain way in a MOOC, and whether the learners agreed or disagreed with that. Participation in such discussions could indicate the learner's commitment to "reside" in a MOOC forum, rather than merely visit to obtain necessary information. Such conversations did not bring any direct benefit to the learner, such as an answer to an informational need or conceptual confusion. They were also open-ended and subject to discussion and opinion exchange. Although low in volume, SN threads were diverse in topics and rich in content. Chapter 7 discusses the content of the SN discussion threads in more detail.

Administrative and informational queries represented closed-ended searches for information about the course. The level of information and technical discussion threads was relatively consistent across the courses, at ten per cent. The Solar course did emerge as a slight exception, recording the highest level (16%) of this conversation type. There were numerous Q&A exchanges in these types of discussions. They manifested learner confusion around course navigation, forum use and locating assessments, and included questions about interface issues or course deadlines.

Other types of discussions. Meta-cognitive discussion threads were infrequent and varied between the courses. The total volume of conversations in this category was negligible for the Water course, representing 0.1% of all discussions. However, this increased to 1% for the Solar course. As noted in Table 6.3, the level of metacognitive discussions was low as compared to the other category types. The content of these discussions included tips for assessment preparation, the best ways of studying lectures, among others.

MIX discussion threads were found in each courses, at an average of five per cent of all the forum core co-occurrences. Essentially, MIX discussion threads represented

combinations of discussion threads already presented. Typical examples would be a combination of self-disclosure intertwined with the request for help with a task (social-task and content-task combination). Another example could be a combination of an experienced difficulty triggered by a task combined with a reflection upon the course. Instead of having shorter, focused exchanges, in some courses learners tended to carry on the conversation in one discussion thread, shifting from one need to another. Analysis of why mixed discussion threads exist and the types that are more frequent is beyond the scope of the study. Introducing more granular units to represent these discussion threads would modify the higher level granularity of the categories. To acknowledge the potential relevance of the themes discussed in the mixed discussions, they are included in the qualitative analysis presented in Chapter 7.

Summary. Content analysis of information exchanges between forum core networks was driven largely by the curriculum. Where self-disclosure occurred in ST discussions, it was at low levels. Social negotiation was used sparingly in SN discussions. In comparison, INF queries maintained a relatively higher presence across all the courses. There were very low levels of META conversations about study techniques. Some courses appeared to foster more fluid MIX discussion threads.

6.6.2 The Change of Discussions Throughout the Course

The second research question examined the dynamics of the discussions where a forum core's posters co-occurred throughout the duration of a course. Temporal patterns of changes in the volume of discussion threads indicated that:

- curriculum-related discussion threads grew over time;
- self-related discussion threads tended to increase as the courses progressed and fluctuated in alignment with frustrating or exciting course events;
- social non-task discussion threads decreased over time;
- informational and technical queries remained stable throughout the duration of courses;
- metacognitive discussion threads emerged at later course stages in most courses.

This sub-section describes these findings in more detail. To understand the shifts in volume in the discussion threads, the posts comprising them were aggregated in different stages over the duration of the courses. Temporal divisions (stages) mirrored

those imposed in Chapter 5 and were distinct across the courses. The stages within the courses were separated based on the first course week followed by the temporal stages delineated by the courses' assessment deadlines. Additionally, for inter-course comparability, the discussion thread types were aggregated at similar time points in the course. Table 6.4 presents the counts and ratios for each individual MOOC. Figures 6.1, 6.2 and 6.3 present the shifts in discussions visually to capture the patterns between courses using comparable time slices. That is, temporal boundaries (i.e. time windows for aggregation) are aligned in the figures in order to ensure a comparability of the temporal patterns between the courses.

Curriculum-related discussions. The volume of curriculum-related exchanges CT and CN) was higher that other types of conversations between the forum core posters (see Table 6.4). Aggregating the post counts within assessment-related temporal boundaries demonstrates that the number of CT and CN discussions remains high throughout the course. This result supports the previous finding for the dominance of content-related threads and their role in driving the use of forums for solving course-related problems.

Table 6.4 Overview of the counts of discussion types throughout the course. Stages are defined in alignment with the learning design presented in Chapter 5 (Table 5.3)

a) Functional Programming (FP)

	Stage 1		Stage 2		Stage 3		Stage 4		Stage 5	
	N	%	N	%	N	%	N	%	N	%
CT	957	38	329	41	277	43	168	38	114	32
CN	1012	40	294	37	216	34	188	42	147	41
ST	73	3	11	1	21	3	22	5	22	5
SN	156	6	49	6	66	10	11	3	23	6
INF	216	8	29	4	40	6	41	9	11	3
MIX	71	3	70	9	15	2	10	3	14	4
META	25	1	3	0.3	2	0.3	0	0	0	0

b) Water Treatment (Water)

	Stage 1		Stage 2		Stage 3	
	N	%	N	%	N	%
CT	9	1	574	27	809	51
CN	280	48	1116	53	435	27
ST	4	0.6	114	5	121	7
SN	178	30	163	7	72	4
INF	106	18	64	3	107	6
MIX	0	0	4	0.1	2	0.1
META	6	1	72	3	14	0.8

Table 6.4 (continued) Overview of the Counts of Discussion Types Throughout the Course – Stages are defined in alignment with the learning design presented in Chapter 5 (Table 5.3)

c) Solar Energy (Solar)

	Stage 1		**Stage 2**		**Stage 3**		**Stage 4**	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**
CT	148	24	524	48	881	62	399	51
CN	136	22	176	16	171	12	167	21
ST	18	2	20	1	20	2	15	1.9
SN	169	27	88	8	59	4	38	5
INF	104	17	190	17	181	13	133	17
MIX	9	1	18	1	16	1	2	0.2
META	30	4	67	6	60	4	26	3

d) Taking Data Analysis to the Max (Excel)

	Stage 1		**Stage 2**		**Stage 3**		**Stage 4**		**Stage 5**		**Stage 6**		**Stage 7**	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
CT	279	41	588	55	252	58	217	50	204	52	278	65	212	64
CN	154	23	180	17	70	16	88	20	59	15	40	9	48	15
ST	66	9	56	5	22	5	13	3	38	9	21	5	7	2
SN	50	7	34	3	19	4	10	2	1	0.2	2	0.4	12	4
INF	91	13	42	9	51	11	28	7	28	7	54	13	43	13
MIX	0	0	4	0.3	0	0	0	0	0	0	2	0.4	1	0.3
META	28	4	69	6	24	5	54	12	55	14	28	7	8	2

Figure 6.1 illuminates the relationship between curriculum-related posts (i.e. content-related communication) and posts containing self-disclosure (i.e. ST) and social negotiation (i.e. SN) (i.e. social-related communication). The graph plots the changing ratio of the discussions of socio-emotional nature (ST and SN) to the curriculum-related discussions (CN and CT) and demonstrates the ratio of self-related and SN discussion threads to the curriculum-related discussion threads at four comparable time points across the analysed MOOC forums. During the first week of the MOOC, the ratio of socio-emotional discussion threads (ST and SN) to the curriculum-related threads was between 1:5 and 1:3 in FP and Excel respectively, and a ratio over 2:3 in Solar and Water (Figure 6.1).

As the courses progressed, ratios dropped to 1:10 and around 1:6 respectively. In other words, the number of curriculum-related discussion threads grew over time in the courses, with every nine posts out of ten being made about curriculum-related issues.

This suggests that throughout the progression of the course, the forum core learners increasingly engaged in and focused discussions on more curriculum-related issues.

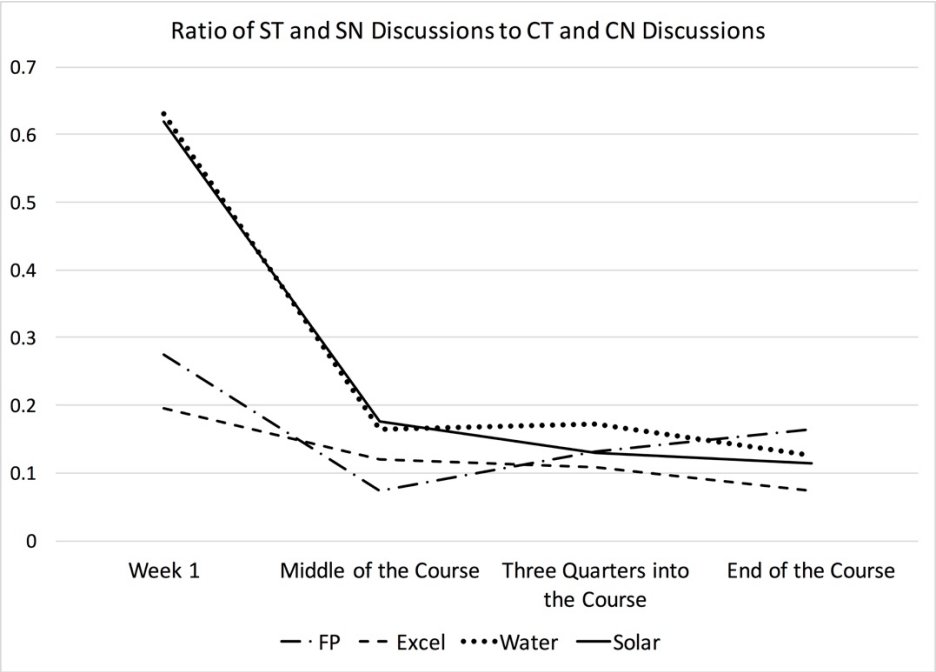


Figure 6.1 Ratio of the sum of ST discussion and SN discussions to the sum of CT and CN discussions threads at comparable time stages, in four MOOC forums

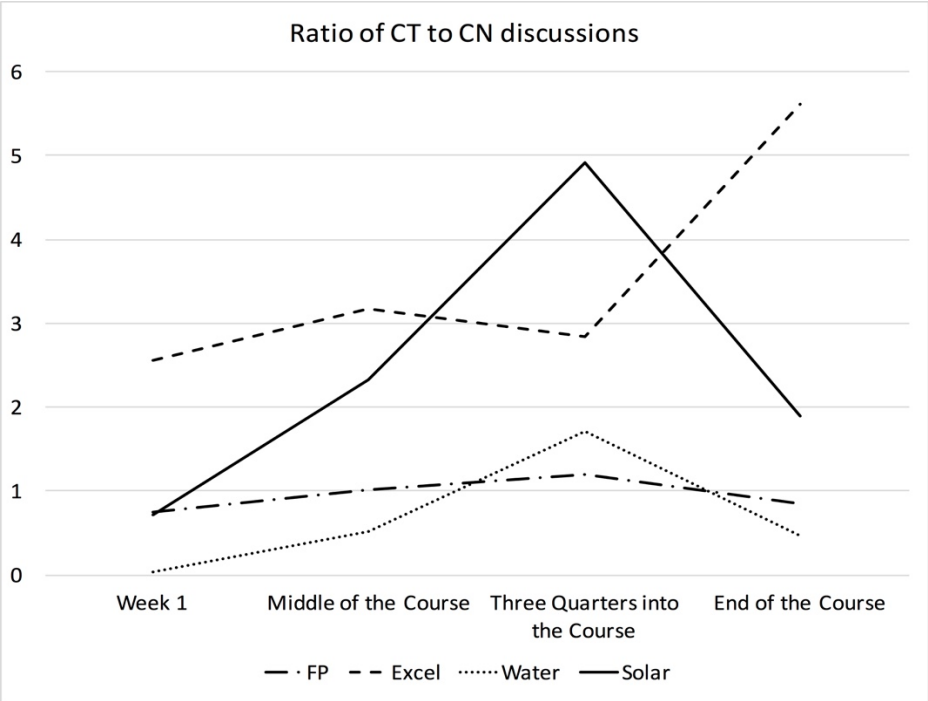


Figure 6.2 Ratio of CT to CN conversations at comparable time stages, in four MOOC forums.

No clear pattern emerges between on-task and off-task curriculum-related communication. Figure 6.2 reveals large variations between CT and CN discussion threads across the various MOOCs (note that the scale for Figure 6.1 differs from Figure 6.2). This could be caused by a combination of different in-course factors, such as difficulty of the lecture concepts, the presence of general discussion topics within the course design, and a high number of participants interested in general broader discussions around the topics. Figure 6.2 maps the ratio of CT discussion threads to CN discussion threads at four comparable time periods in the four courses involved in the study. It can be observed that at the onset of the course, in both Solar and FP, the CN and CT discussion threads are in balance, with the ratio close to 1:1. In the first week of the Excel course, there are fewer than one-third of CN discussions compared to CT discussions. And in the unmoderated Water MOOC, there were very few CT discussions compared to CN discussions. Figure 6.2 demonstrates the temporal differences in how the individual cohorts conducted these different discussions. By stage 3, three-quarters of the way into the course, the proportion of CT discussion threads in relation to CN threads has increased significantly. By the final stage of the course (stage 4), the proportion of interest-based discussion threads, or those seeking conceptual clarification, has again increased and the proportion of CT discussions has decreased slightly. The exception was in the Excel course, where CT discussions had spiked by the end of the course. The reasons for these differences may be related to the combination of general interest, personalities and the moderation policy, and a course design that encouraged or motivated additional discussions.

Social-task and social non-task discussions. The coding included two types of situations of a social nature (ST and SN) belonging to distinct categories of self-related discussions and setting-related discussions. First, self-related conversations, ST discussions, were triggered by a learner's reflection upon her undertaking of a particular task in the course. Second, SN discussions, were triggered by learners' reflecting upon the social system of a MOOC, such as why things are the way they are, what is the purpose of the MOOC or the forum. In both cases, these individual reflections prompted learners to share and seek out opinions from fellow learners. As Section 6.6.1 showed, ST discussions occurred at an average of four per cent, and SN discussions at an average of seven per cent. These discussion threads are of interest as they are linked to the emergence of social contexts in MOOCs.

ST discussions appear to grow in volume and in ratio throughout a course's duration. Further analysis of the posts indicated that the surge in ST discussions reflects certain course events. For instance, in the first weeks of Excel, learners were encouraged to post their work settings and the spreadsheets they were working on. Many offered creative and emotional exchanges of these images. Another surge in the Excel course can be observed in in Week 5. Analysis of the posts indicated that learners were frustrated with using the new software, which required a steep learning curve, and expressed that they were having difficulty coping. Alternatively, in the FP MOOC, the increase in the number of ST posts occurred in stage 2. That is when the learners were offered a complex functional programing concept of monads, which caused frustration, and generated ardent discussions about its meaning for several weeks. It appears that the slight increases in the number of self-related posts were not limited to either positive or negative emotions. To better understand the overall patterns in ST discussion threads, Figure 6.3 plots the counts of ST discussion threads at comparable time periods across the courses. It shows that across all courses, except Excel, ST discussions seemed to be at higher levels in the middle of the course.

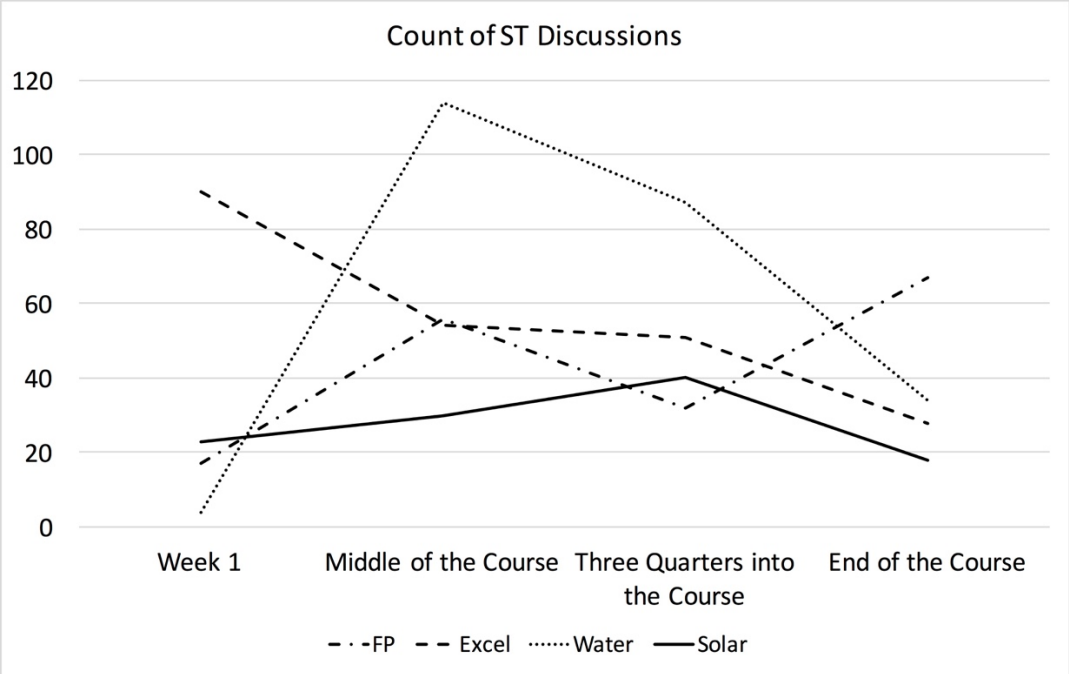


Figure 6.3 Counts of ST discussion threads at comparable time stages, in four MOOC forums.

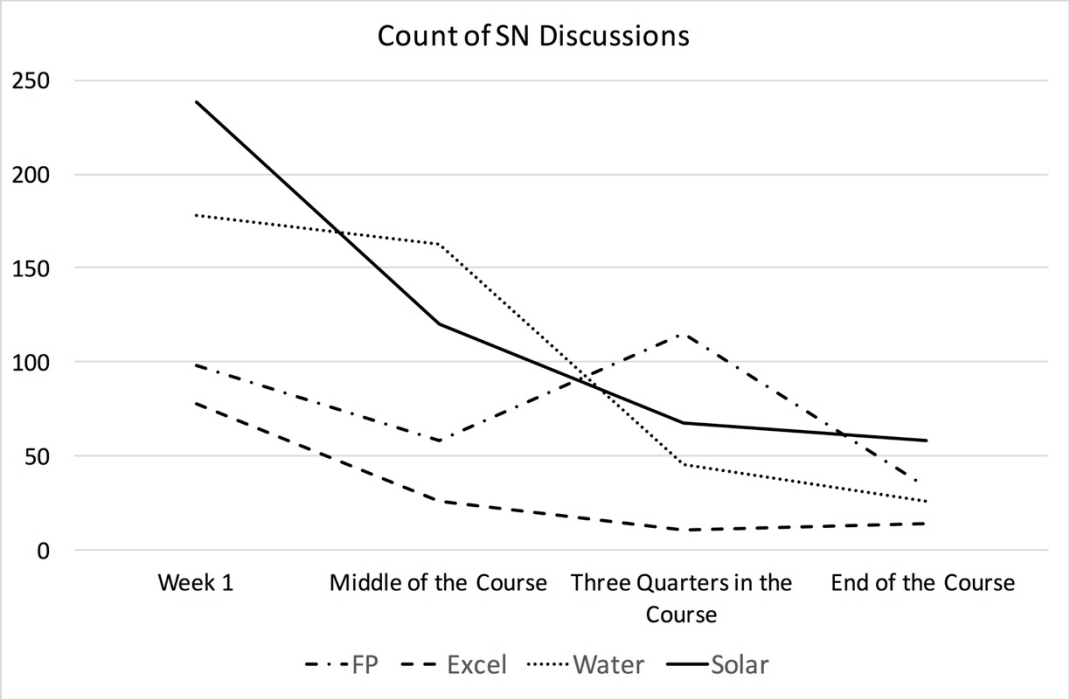


Figure 6.4 Counts of SN discussion threads at comparable time stages, in four MOOC forums.

While ST discussions, (except in the Excel course), appear to have grown in volume as the courses progressed, SN discussions appear to have decreased (Figure 6.4). ST discussions seem to be aligned with emotionally exciting or frustrating course events. In contrast, the number of SN discussion threads drops as the courses progressed, with the highest volume appearing in the first weeks of the courses. Figure 6.4 shows that the decrease in the number of SN posts was typical to all the courses, except in FP, which shows a sudden increase of SN discussion threads in stage 3. There were no obvious reasons for this spike.

Administrative and Informational Queries. The percentage of INF queries in the total number of the discussion threads fluctuated as the course progressed. The range observed in the Excel course was 7%–13% (Table 6.4d), in Solar 13%–17% (Table 6.4c), 3%–9% in FP (Table 6.4a), and in Water 3%–18%. In terms of the volume of discussion threads, they were at their highest in the first week in every MOOC (Table 6.4). The ratio of INF discussions appears to grow in the final course stages from that in the middle of the course. This might be explained by learner anxiety in understanding how the progress bars work, and whether their last assignment or exams met the bar for passing the course. Some of the INF queries asked throughout the courses were related

to issues previously discussed, and had been raised again by newcomers. Forum core members would address the issues for them, or point them to the relevant previously held discussions. Other purposes for these queries throughout the courses related to understanding assessments, issues with software and technical bugs (i.e. non-conceptual issues) within the interface, lectures or labs, and similar.

Metacognitive discussions. The number of META discussions related to study strategies among the forum core was low, at an average of four per cent (Table 6.4). There were six META learning-related discussion threads in Excel and seven in Water. Both FP and Solar had higher counts. In all courses, but FP, META discussion threads with learning-related strategies appeared or grow as the course progressed. In FP, a course where the topics were strictly sequenced by the instructor, most of the META discussion threads occurred in the first week. Sharing of one's learning strategies later in the course, as observed in other three MOOCs, could be linked to the openness of learners regarding their struggles in learning, hence, they were seeking solutions. The analysis provided no additional information to explain the low number of META discussions, except flagging the commitment to help others learn and be open about one's own learning practices.

Summary. This sub-section reviewed analyses of the changes in volume for the different types of discussion threads over the course of the four MOOCs examined. First, it was demonstrated that curriculum-related discussion threads were prominent throughout the duration of the entire course. Further, they grew in number as the course progressed, demonstrating learner engagement with one another around specific course tasks. Second, the analyses demonstrated that ST discussion threads also increased in number with the course progression. This suggests that there is a willingness for learners to self-disclose positive and negative emotions among the forum core. SN discussion threads appeared to be at their highest level in the early stages of the course. They may occur then to define the MOOC forum space and set the ground rules for who the forum is for and how it is to be used. INF queries were raised throughout the courses' progression, indicating a persistent need for technical support for learners throughout the course. META discussion threads were low in number and tended to occur at the later stages of the course.

6.6.3 Discussions Central to Network Formation

The third research question investigated the discussion threads that yielded contributions from many forum core posters. A particular kind of discussion thread might have been low in number in the content overall, but could have served as an opportunity for learners to engage in denser exchanges, resulting in higher familiarity and/or trust. To identify major discussion threads in each of the analysed forums, networks of learners and discussion threads were constructed. The central discussion threads yielding the highest co-occurrences from forum core posters, along with a high numbers of contributions, are presented in Table 6.5.

Table 6.5. Central conversations from forum core posters

a) Functional Programming (FP)

Rank	Posts, N	Type	Stage	People, N
1	54	ST	Stage 5	9
2	45	MIX	Stage 2	19
3	32	CT	Stage 4	17
4	31	SN	Stage 3	14
5	30	CT	Stage 3	11
6	30	CT	Stage 2	6
7	25	CN	Stage 1	7
8	25	CT	Stage 1	6
9	24	CT	Stage 3	9
10	24	CN	Stage 3	8

b) Solar Energy (Solar)

Rank	Posts, N	Type	Stage	People, N
1	42	CT	Stage 3	10
2	39	CT	Stage 3	15
3	28	SN	Stage 1	15
4	23	INF	Stage 3	4
5	23	CT	Stage 4	8
6	22	INF	Stage 1	13
7	21	SN	Stage 1	11
8	21	CT	Stage 3	15
9	18	CT	Stage 3	11
10	17	CT	Stage 3	5

c) Taking Data Analysis to the Max (Excel)

Rank	Posts (N)	Type	Stage	People (N)
1	23	CN	Stage 2	20
2	21	CT	Stage 7	6
3	17	CT	Stage 6	3
4	15	CT	Stage 2	8
5	15	CT	Stage 2	11
6	13	CT	Stage 1	6
7	13	CN	Stage 1	8
8	13	CN	Stage 3	4
9	13	MIX	Stage 4	5
10	13	CT	Stage 4	4

d) Water Treatment (Water)

Rank	Posts (N)	Type	Stage	People (N)
1	44	CT	Stage 2	16
2	31	CT	Stage 2	23
3	29	CT	Stage 3	17
4	29	CN	Stage 1	4
5	28	CT	Stage 3	13
6	25	CT	Stage 2	16
7	22	CT	Stage 3	9
8	22	CT	Stage 3	11
9	21	CT	Stage 2/3	11
10	20	CT	Stage 3	8

The results confirmed the overall nature of the forum core as identified Sections 6.6.1–6.6.2. Curriculum-related discussion threads were central across all four MOOC forums. In all four courses, central CT-related discussions took place half-way through the course. They were associated with difficult exam questions and demonstrated learner commitment to content mastery. These central discussion threads were also different from discussions with two or three conversational turns, previously found to be typical of MOOC forums (A. Anderson, Huttenlocher, Kleinberg, & Leskovec, 2014). Instead, CT discussion threads attracting contributions from many forum core posters were longer. They appeared to contain learner engagement with difficult course concepts at later course stages.

The Solar MOOC highlighted the importance of early INF queries or its forum core (Table 6.5b). Across all courses, these kinds of discussions involved closed-ended

questions or direct answers to narrow questions about the use of technology or deadlines. In Solar, the informational discussions central to the forum core's interactions were related to strategies on making the best use of the course's technology. Both informational queries were about how to better use the videos. In one discussion, a learner asked about ways to download videos. Another central INF discussion thread of this nature initiated upon an inquiry about video subtitles. Both discussions were replied to with suggestions that revealed learner sensitivities to the varying limitations of Youtube across the globe. In that sense, both conversations were representative of exchanges about the tools for learning.

Forum core posters from both the Solar and FP MOOCs participated in SN discussion threads (Table 6.5a, 6.5b). The centrality of SN threads in the two courses is of particular interest, given their overall lower volume within the discussion threads – an average of seven per cent. Although these discussion threads were generally few in volume, their centrality indicated that their content was of interest and of relevance to many forum core members. Hence, one possible interpretation could be that these SN discussion threads served as opportunities for forum core formation. Posters eagerness to contribute to such discussion threads that were not required by the course or by the instructor also highlights the commitment of some forum core posters towards engaging in issues beyond the curriculum and common interests. For instance, in FP, a conversation that yielded the highest number of contributions from forum core posters was titled "Passing course, but somewhat lost". The opening post for this discussion thread with 54 contributions was as follows:

> I have just about finished the homework/labs for week 8, and to my astonishment, I am now on 56%, so it looks like even if I put on a blindfold, get completely drunk, have a frontal lobotomy (which almost seemed like a good idea, after the first brush with monads), I will pass the course. However, while I am comfortable now with recursive functions, guards, folds, and list comprehensions, I'm not at all happy with monads, and the various right arrow combos, and don't feel I could write any meaningful programs other than getting my code golf handicap to be slightly less outrageously over par on something like Project Euler. Having said that, it has been a very interesting experience for my old brain, and I will pursue functional programming more, though probably not Haskell, as I would probably be fired if I started trying to use it for work.

This post brings to light the intertwined nature of a student's learning progress (what was learnt) and reflection on that progress (how the learner is feeling). The thread contains similar reflections from other learners, their personal reactions and discussions

about why the course was the way it was. The content of the post is of a social nature and is not related to a given task, and nor is it a self-reflection upon a task. It is, however, a discussion about the course (not its content) and is also used for sharing emotions. The social nature of this post demonstrates that there was a certain number of people had established a certain affectivity and solidarity through in-course communication.

To conclude, examination of the types of discussion threads that were central to the forum core's activities confirms previous observations. The forum core participants were driven by the need to engage with and solve curriculum-related issues. However, central discussion threads demonstrated that issues related to the learning of difficult curriculum occurred later in courses. Also, at times INF queries were of interest to forum core members. Finally, SN discussions signified learner commitment to using the forum beyond its function as a place where learners drop-in to exchange Q&A and leave. These additional insights suggest that the forum core is a learning-oriented group that experiences some community-related negotiation and discussion processes. To further understand these, Chapter 7 delves into the thematic analysis of discussion threads that were of a socio-emotional nature; that is, the SN and ST discussion threads.

6.6.4 Summary of Findings from Research Questions

The first research question inquired about the types of communication dominant in forum core discussions, underpinning forum core co-occurrences. The study found that the opportunities for forum core members to co-occur were dominated by curriculum-related discussion threads (some 74% of the content), and closed-ended questions about INF issues (some 10% of the content). ST discussion threads were at a low level as were SN discussions reflecting on the nature of the courses, as well as META discussions about learning strategies.

The second research question addressed the change in the volume in relation of different types of discussions throughout the course. Analysis demonstrated, over time, the growth of curriculum-related discussions, an increasing pattern of (an overall low number of) self-related discussions, and a decreasing pattern for the setting-related off-task discussions. META discussions were emergent at later course stages, whereas INF remained at a similar level throughout the course.

The third research question inquired about the centrality of particular discussion types to the forum core. The results offered more evidence about the significance of the curriculum-related discussion threads to the forum core. The difficult and lengthy exam task-related threads, as well as those tackling complex concepts, were usually central for forum core member co-occurrence. The centrality of discussion threads also revealed that in some courses, forum core members felt compelled to engage in a number of setting-related off-task discussions that negotiated some aspect of the MOOC, its set up, or its impact on the learners.

6.7 Discussion

The overarching research question inquired about the role of on-task and off-task communication, social and content-related in nature, in the formation of a forum core. Results indicate that both on-task and off-task curriculum-related communication comprised the bulk of exchanges within the forum core, whereas off-task setting-related discussions appeared to have been central to the forum core's formation in some of the MOOCs. Analysis demonstrated that the nature of the discussions among the forum core was predominantly curriculum-driven. The number of on-task curriculum-related communication (CT) increased over time, whereas off-task curriculum-related discussions (CN) fluctuated depending on the course's design. ST threads associated with self-disclosure were low in number, despite gradual growth. This indicates a low increase in the number of more personal messages, which were directly related to the social climate of a forum. Off-task setting-related communication (SN) was also low in number. However, their centrality in the network points towards the relevance of off-task setting-related discussions (SN) for the forum cores' members. The centrality of SN discussions indicates that the forum core undergoes social processes of community development, despite low levels of affectivity.

The findings suggest that to some extent the development of the forum core resembles the development of formal online groups towards learning communities, underpinned by relational processes. However, some notable differences were also observed, particularly in the low levels of interpersonal relatedness. The following section discusses these findings in relation to the existing literature in both formal online settings and MOOCs.

6.7.1 Forum Core as a Curriculum-Driven Group

The findings support the premise that forum core posters predominantly engaged in social activity (discussion forum) to address shared curriculum-related issues. The majority of discussion interactions were curriculum-related for both on-task and interest-based. This finding was consistent for all courses. The analysis also demonstrated that curriculum-driven discussion threads, particularly those related to the tasks in the course, increased in volume as the course progressed. The centrality of curriculum-driven, task-related discussion threads to the forum core activity further supported the role of these discussion types for the forum core formation. Analyses of the discussion thread, volume, over-time change and centrality all point to forum core posters discussions around the tasks and common interest around the course content as primary reasons why learners join the forum.

The study's finding about the growing number of curriculum-related threads poses an important question: What defines forum activity and how forum activity should be evaluated? Prior studies have found that the forum activity decreases exponentially as the course progresses (Brinton et al., 2013; Rossi & Gnawali, 2014). Such decrease of the number of posts in the forum with every subsequent week is obvious given the exponential decrease of the posters. The findings of this study raise the question of whether the decrease in posts is equal to the decrease in forum activity. As shown by this study, within the overall decreasing number of posts, the number of posts by regular posters and which are related to the curriculum and assignments increases steadily to a large volume. Given the nature of increasing posts-types and that they are contributed by learners who use the forum consistently, it could also be said that forum activity by those learners who appear motivated to engage in social learning increases. The instructor evaluating forum activity may consider the growth of curriculum-related discussions, among a clearly smaller group of learners, a positive reflection on a forum activity.

In addition, this finding casts doubt unto the existing findings around the cognitive presence in MOOC forums. Previous research has shown that the overall level of cognitive presence in the forums is low (Kellogg et al., 2014; Kovanović et al., 2017). The current study was not focused on evaluating the quality of learning within the curriculum-related threads; however, its results suggest the increase in content-related

activity, and the presence of more complex discussions with the course progression. The centrality of curriculum-related threads reveals the long turn-taking exchanges around the curriculum that take place about half-way through the course. The timing of these more involved turn-taking curriculum-related discussions coincides with the time of the surge in curriculum-related posts. It can be speculated that as the content becomes more advanced, learners both need to engage more with it more, as well as to engage in deeper knowledge construction around it. Consequently, the levels of cognitive presence pertinent to the posters who use the forum consistently may as well differ from the indicators for the entire cohort.

6.7.2 Relational and Affective Aspects of Forum Core Development

This study's findings also revealed the low-level development of the social context. Some discussion threads were of a socio-emotional nature: on average, there were four per cent ST threads and seven per cent SN threads. The number of ST discussions reflecting learner disclosure increased over time, whereas the number of SN discussions decreased over time. Despite their low number, in some courses SN discussions were also among the threads with highest centrality. Many forum core posters chose to contribute several posts to these discussions.

To interpret the results, parallels needed to be drawn from research regarding the social context of formal online courses. It appears that in this study, the patterns of socio-emotional discussion threads exchanged by the forum core in MOOC settings are similar to those reported for more formal online courses. As discussed in Section 6.2.1, Swan (2003a) used affective, interactive and cohesive indicators to demonstrate the evolution of the elements of social presence in a formal course. Analysis from this study suggests that the pattern of ST discussion threads aligns with the development of affective indicators of social presence, found by Swan. In her study, the number of affective messages containing disclosure grew over time. In a MOOC, self-disclosing statements appeared in ST categories. Because only reciprocated interactions were a part of the forum core analysis, the growth of ST discussions also reflects an increases in self-disclosure and acknowledgement from others, possibly with the offer of social support. The gradual growth in the number of ST threads with course progression can be interpreted as evolution of the forum core towards building social climate and social support.

Besides the increase in self-disclosure, Swan's study (2003a) also described a decrease in the so-called cohesive indicators of social presence. Swan found that the number of cohesive indicators of social presence decreased as the course progressed. This category is similar to the SN discussion threads. Swan's cohesive indicators included greets and salutations, group references, social sharing, and most importantly, self-reflection. Self-reflection was actually defined as "reflection on the course itself, a kind of self-awareness of the group" (p. 142). Swan's definition of cohesive indicators ultimately overlaps in meaning with the SN discussion types coded in this study. The MOOCs' SN discussions contained negotiations around why the forum and the course were conducted in a certain way, as well as how forum users should act within the shared online space. Like Swan's cohesive indicators of social presence, the forum core's SN discussion threads decreased in volume over time.

The mere presence of SN discussions, as well as their centrality to the forum core, may indicate group formation processes. To gain insights into the quality of the socio-emotional discussion threads, further analysis is necessary. Such processes could potentially include discussion of ground rules and common ground establishment. Many MOOC forum core members chose to introduce themselves in the first week's discussion threads, which were oriented towards meeting one another. In studies that have automated the analysis of posts, the discussion threads in which MOOC learners introduce themselves are often referred to as undesirable. Brinton et al. (2013) described them as noise hindering the algorithms for meaningful thread detection. However, the high number of forum core members in those threads could be interpreted as indicating personal motivation to be known, using the forum space as a common space to be. Moreover, considering their low volume, the high centrality of some SN discussion threads demonstrates their importance as compared to the low volume of these discussion threads. Similarly, their centrality may signify a commitment of forum core members to make the forum core their own, to negotiate its meaning, and contribute to and shape its development.

Despite a similarity in the pattern of social context evolution across forum core in a MOOC setting, this study's results suggest that forum core development unfolds through different mechanisms. The MOOC forum core did experience social processes leading-up to the formation of groups and communities, as indicated by the presence and patterns of ST discussions and SN discussions. The pattern of development of these

discussions as the course progressed reflects the relational processes observed in the formal courses. However, despite this similarity in pattern, the development of a MOOC's forum core appears to take place through a somewhat different mechanism. Given the low quantity and generally low co-occurrence of ST discussions, it is unlikely that the formation of a forum core is based on interpersonal affectivity.

The contrast between strong group cohesion and low affectivity found in MOOC forums in this study is supported by existing MOOC research which examined MOOC learner perceptions of social presence as a part of the Community of Inquiry questionnaire (Arbaugh et al., 2008; Garrison & Akyol, 2013). When combined with the results of this study, empirical work by Kovanović et al. (2017) and Poquet et al (2017) suggests that the evolution of a forum core into a community is not based on interpersonal affection and close bonds, as assumed in the learning communities of formal online classrooms. According to the study by Poquet et al. (2017), MOOC posters with a regular (i.e. a sustained) level of participation, reported a low level of established interpersonal trust and familiarity, a result that resonates well with the low level of self-disclosing discussion threads found in this study. In addition, in the study by Poquet et al., regular MOOC posters reported higher perceptions of group cohesion. That also resonates with the presence of cohesive indicators (i.e. SN discussion threads) in early MOOC stages. Similarly, the findings of Kovanović et al. (2017) support the supposition that interpersonal dyadic affectivity may not be a part of social presence in a MOOC setting. The authors conducted an exploratory factor analysis of learner perceptions related to the community of inquiry (Garrison et al., 1999). The factor loadings associated with social presence perceptions were different within the data collected from a MOOC setting than they were previously within the data collected from formal settings. In a MOOC setting, factor loadings associated with the affectivity based on interpersonal familiarity were independent from group cohesion and open communication. In other words, it appears that the social context in MOOCs is not based on affective interpersonal relations; nevertheless, group-related processes may still be taking place.

6.7.3 Comparing Forum Core with Groups in Formal Online Courses

There are some similarities and differences between MOOC forum core and groups in formal online courses. First and foremost, both experience a continuity of interpersonal interactions and establish a shared history of curriculum-related and SN communication.

Although the learning communities in both formal courses and MOOC forum core groups appear to be driven by a need to learn the curriculum, they are founded on differing social motivations. This study's content analysis of four forum cores demonstrated that the content exchanged via network ties was predominantly driven by course curriculum (RQ1), specifically by graded tasks. This finding is much in line with prior CMC research which suggests that CMC environments are naturally more task-oriented (Walther et al., 1994). The volume of META activity found in forum core discourse also compares to that in formal online courses. McDonald (1998) referred to meta-cognitive discussions in formal online courses as "almost non-existent". Another difference between the discourse reported in formal online courses and this chapter's analysis of forum core discourse is also observed in the volume of INF queries and its change throughout the course. In formal courses, the volume of such conversations decreases over time, whereas in this study's four MOOCs' forum core, it remained at consistently high levels. First, the centrality of the INF queries in some courses suggests the importance to the forum core's learners of tools of all types (e.g. deadlines, videos, progress bars). Secondly, sustained high levels of INF queries could be interpreted as discussions that take place between forum core members and later-joining posters. It is likely that some of the same questions discussed by a forum core in the early weeks of their MOOC are asked later by intermittent newcomers. In that sense, forum core posters may use these kinds of discussions as opportunities for introduce the newcomers to the ways of communicating on the forum.

In relation to the dynamics of social context, as observed through ST and SN discussions, the forum core differs from the evolving groups in formal online courses, which are underpinned by interpersonal attachment between its members. The patterns of slow growth of interpersonal messages and engagement in group-related MOOC discussions mirror formal online courses. However, low levels of affectivity suggest that group attachment in the forum core is based on principles other than interpersonal familiarity and trust.

This finding has a significant implication for the transfer of formal educational theories and models. In formal online education theory, constructs related to the social context are largely dependent largely on an interpersonal dyadic level of interaction and salience of the self. Lowenthal (2012) explained that conceptualisations of social presence all refer to it as resting on an interpersonal relationship that requires a distinct perception of another person (see Section 6.2.1). The overwhelming bulk of existing approaches to social context development in formal courses favours interpersonal relations as foundational. These conceptualisations seem to be unfit to capture the mechanisms of social relations within a MOOC setting.

To conclude, this chapter enables for a comparison between the forum core in MOOCs and participants in formal courses. Analysis of the content of network exchanges between the forum core posters reveals the growth over time of curriculum-related discussion threads, with the slow intensification of social support and establishment of social rules in the background. Similar dynamics were reported in formal online courses (Akyol & Garrison, 2008; Hara et al., 2000; McDonald, 1998); however, the low levels of affectivity cast doubt on the mechanisms of group formation in a forum core. Groups in formal courses evolve based on interpersonal relationships, and it appears that the forum core evolve into groups and further into communities without established interpersonal relationships.

Despite the low opportunities for co-occurrence, a forum core could be analysed using socio-constructivist tools derived from formal classrooms. Additionally, given the similarity of the content exchanged via forum core ties and the patterns of development of a learning community, one can presume that cognitive presence and socially situated learning within this group could be enhanced and built as the course progresses. However, the study also suggested that the mechanisms for group evolution in MOOCs are different; hence, teaching models need to be re-applied with caution. The next chapter investigates the potential mechanism for group formation in a forum core.

6.8 Conclusion

Chapter 5 analysed the association between the type of forum participation and higher chances for co-occurrence between pairs of learners in a MOOC forum. The findings revealed that forum core has the highest potential for co-occurrence and therefore holds greater opportunities to form relationships. Interpersonal relationships could develop a

social context in a MOOC that is conducive to learning in such informal and scaled contexts. To better understand the nature of the interactions between forum core posters, Chapter 6 the types of discussion threads yielding forum core contributions.

The findings in Chapter 6 suggest that that a forum core's development is driven by informational and learning needs, but is unlikely to be based on interpersonal affectivity. However, the centrality of SN threads signifies that the some forum cores engage in social processes of community development. This differentiates the MOOC forum core from a community in formal online courses, which are underpinned by the interpersonal attachment of its members. Chapter 7 sets out to understand the role of SN and ST discussion threads in the formation of group attachment, which is based on shared identity, rather than one-to-one relationships.

7 Identity formation processes in forum core

This chapter analyses the issues discussed by forum core posters within social-task, social non-task and mixed discussion threads. These discussions threads were analysed qualitatively using the theoretical constructs of identity-based communities and communities of practice. The chapter interpreted these discussions of a social nature in relation to the development of shared identity.

7.1 Introduction

Analyses in Chapters 5-6 suggested that a forum core is a group of MOOC-takers engaged in sustained social learning activities. However, Chapter 5 demonstrated that person-to-person interactions within such sustained social activities was rather minimal for most forum core posters. In addition, findings from Chapter 6 showed that a forum core was largely driven by the informational exchange and interest in solving tasks and discussing curriculum. These insights signify that on average, interpersonal interactions between two learners although continuous in time, were scarce and mostly on-task. As a result, interactions between forum core posters were unlikely to have been underpinned by one-to-one emotional connection and interpersonal trust, perceived as an achievable learning outcome in small groups in formal online courses.

Despite low affectivity and infrequent co-occurrence, forum core discourse included discussion threads of a social nature associated with relational processes of group formation. Chapter 6 established that forum core exchanges contained self-disclosure (social-task discussions) and group-related negotiations (social non-task discussions). The presence of such discussions of a social nature suggested that forum cores were not stripped of the social processes of group formation. The centrality of group-related negotiations to the formation of forum cores within two analysed courses (FP and Solar) further highlighted that the social processes of group formation were central to the groups' development. In these courses, the forum cores could be hypothesised as evolving towards emergent groups and, possibly communities.

To examine the social processes within the forum cores, this chapter analyses discussion threads of a social nature. Such discussion threads are not related to course curriculum, and include social-task, social non-task and mixed threads. To examine these threads,

first, the literature review in Section 7.2 argues that forum core posters can be approached as identity-based communities, where the mechanism of attachment is based on a self-identification with shared group identity (Prentice, Miller, & Lightdale, 1994). Further, the literature review suggests that the communities of practice theory can be employed to interpret the development of a shared group identity within a forum core. Section 7.3 formulates the research question, and Section 7.4 documents the process of analysis. The qualitative analysis of the issues discussed by the forum cores within social-task, social non-task and mixed forum threads is presented in Section 7.5, whereby the issues are interpreted as negotiations around situational norms of the MOOCs' and forum use. The discussion in Section 7.6 relates these negotiated issues and ways they are conducted to the processes of identity-formation.

7.2 Literature Review

This section argues that forum core posters can be approached as identity-based communities, and that the communities of practice lens is suitable to examine group-level learning processes that forum core posters undergo in the formation of a shared identity. The findings from Chapter 5¬ and Chapter 6 concluded that forum cores are unlikely to form communities based on dyadic relationships of familiarity. However, community formation is not exclusively based on the interpersonal bonds underpinned by one-to-one trust. An alternative mechanism of group formation is that of identity-based communities where emotional attachment to a group is based on shared identity. Grounded in the characteristics of identity-based communities, Section 7.2.1 argues that a forum core can evolve into one. Sections 7.2.2 and 7.2.3 then discuss how the communities of practice theory can be used as an analytical lens to understand the development of a shared identity amongst forum core posters.

7.2.1 Understanding Attachment in Group Formation

Social psychology has presented two distinct mechanisms describing individual attachment to a group in group formation. Attachment, i.e. "members' affective connection to and caring for an online community in which they become involved" (Ren et al., 2012, p. 842, footnote) can be an outcome of interpersonal bonds (one-to-one). Interpersonal bonds imply that two familiar individuals develop an affective, "friendly" connection (Allport, 1924; Festinger, Back, & Schachter, 1950). In such instances, a group's strength is characterised by who knows whom, and the attraction

between the similar members. An alternate model explaining attachment to a group is based on shared identity (one-to-many). In this instance, attachment to a group results from an individual's self-identification with a group's purpose, such as in topic-based groups, such as sports teams or a school newspaper (Prentice et al., 1994). In identity-based communities, the strength of a group is within the attachment to shared purpose. The two mechanisms represent two distinct approaches to understanding the nature of relational processes within groups.

Through prior analysis it appears that forum core posters do not experience attachment based on interpersonal bonds. The co-occurrences between the posters is generally low, and interpersonal relationships of trust and familiarity seem unlikely. Given the presence and centrality of setting-related threads, an identity-based approach to group attachment is relevant to the analyses of the social context within a forum core. The heightened anonymity, focus on the topical curriculum-related exchanges, and high member turnaround are features associated with the emergence of an identity-based community. These features are also characteristic of a MOOC setting.

First, MOOC forums are notorious for their heightened anonymity (Section 2.4.1). Postmes, Spears and Lea (1999) asserted that environments described by a degree of anonymity where individual identities are less distinguishable, enhance the salience of group identity. According to their empirical work, anonymity fosters a forging of social identity in groups where individuals do not have cognitive representations of other individuals. Hence, group formation based on identity is more plausible in the forum core where familiarity and interpersonal co-occurrences are limited.

Secondly, in addition to the lack of familiarity between members, prior investigation of online communities from the identity-based perspective have suggested that the lack of topical breadth is a distinct feature within such communities. For instance, Sassenberg (2002) applied an identity-based and bond-based distinction to understanding the perceptions of IRC-chatters. His analyses confirmed that the individuals who engaged in on-task chats were more attached to the group than to individual members, while those engaging in off-task chats demonstrated interpersonal attachment. Similarly, Grabowicz and colleagues (Grabowicz, Aiello, Eguíluz, & Jaimes, 2013) demonstrated that identity-based communities on Flickr engaged in more focused discussions through narrower topics. In relation to a forum core's context, curriculum-driven interactions,

and particularly task-related discussions have been shown to be prevalent, suggesting that the breadth of topics is limited to curriculum concepts.

Finally, the turnover of membership does not impact individual behaviour as much as in common-bond communities where members will leave as a result of following their 'friends'. According to Turner (1985), when attachment is motivated by identity, other members are perceived as interchangeable. Such would explain the sustainability of a forum core despite varying patterns of forum participation among its members, i.e. participating anywhere from any three weeks to every week in a course.

Summing up, investigating community development through identity-based attachment of one-to-many appears relevant in educational settings such as MOOCs where: (1) the scale and non-privacy of the environment aggravate the development of bond-based relationships; (2) members come together, driven by information exchange; and (3) the open nature of participation may create high membership turnover.

7.2.2 Participation and Communities of Practice

Understanding learning as occurring at a group-level can offer analytical tools to investigate how social identity is developed in a forum core. Collective learning considers shared identity as an outcome, developed from individual participation in collective practices. Becoming a member of a particular group is viewed as a learning trajectory, captured by learner participation in social practices (Koschmann, 2002; Lave & Wenger, 1991). Learning as participation in social practices reaches beyond mere assistance with knowledge construction (Sfard, 1998), and emphasises the relational aspects of human interaction in educational contexts. Sfard wrote that when approached at a group level rather than at the individual level, learning therefore could be conceived as "a process of becoming a member of a certain community" (p. 6). Becoming a member of that community includes, "the ability to communicate in the language of this community and act according to its particular norms" (ibid.). When individuals participate in social practices, an emergent community develops the ability to know things collectively, by negotiating and collectively constructing them, as well as situating them in a shared discourse and practice. Consequently, investigating if a forum core has engaged in discussions around shared practices and has negotiated forum participation collectively would inform if collective learning processes leading up to identity-formation were in place.

Among well-established theories approaching learning as participation and identity development is the theory of the community of practice (Lave & Wenger, 1991; Wenger, 1998). The term 'community of practice' (CoP) refers to groups sharing a concern, a set of problem, or an interest. The main activity of such groups is to deepen knowledge and expertise by continuous interaction (Wenger, McDermott, & Snyder, 2002, p. 4).The foundations of the community of practice theory are suitable for examining the development of a forum core into identity-based communities. According to Wenger (1998), the development of a shared identity is critical for the formation of a CoP, while friendship or affection are not required. Instead of friendships, collective intention becomes the central point around which the social identity of a community can develop. Through the trajectory of participation in practices around a collective interest, an individual learns about the group's shared history and values, and becomes a full member in community practices.

Theoretical constructs salient in CoPs can be used as an overarching analytical framework to shed light unto the mechanisms of identity formation in forum cores. The essential components of an emergent CoP comprise: i) mutual engagement; ii) a joint domain of practice; and iii) a shared repertoire. Mutual engagement is the key element related to actual participation in shared practice. Within a CoP, mutual engagement implies membership, as "people are engaged in actions whose meanings they negotiate with one another" (Wenger, 1998, p. 73). Membership in a CoP is more than a category or a label. In Wenger's view, being included in the matters of importance to the group is the requirement for being engaged in the community's practice, as it also defines belonging. Mutual engagement therefore eventually determines social identity, defined by Tajfel (1978) as "that part of an individual's self-concept which derives from …knowledge of …membership of a social group together with the value and emotional significance attached to that membership" (p. 63). In other words, membership in a community comes to be negotiated through the issues of relevance to the group. The importance of negotiation of common issues for membership offers support to viewing the centrality of social non-task discussion threads as relevant for group-formation processes in forum cores.

Besides the element of mutual engagement, a CoP also presents other components critical to the evolution of social identity. An emergent CoP reifies participation and mutual engagement by establishing material or abstract artefacts that capture a shared

repertoire. Shared repertoire, i.e. reified agreements around shared practice, can also be constituted through routines, words, tools, ways of doing things, stories, concepts and metaphors (Wenger, 1998). By creating a shared history reflecting the creation of a shared repertoire, a CoP's members also create a regime of mutual accountability. According to Wenger (2000), mutual accountability is more than a set of rules; it is the system of relationships that enforces behaviour to serve a particular group. In the process of creating a shared repertoire, history and accountability, a group learns as a whole about their shared domain of practice. Therefore, the learning outcome of an emergent community are also the agreements and reified artefacts around the practice that support social identity.

CoPs can illuminate how group attachment is enacted. In Wenger's view (2000), an underlying mechanism of one's participation in the social system involves making sense out of the system, and an individual's position in it. During the engagement stage, an individual interacts with the practice and a community around it. By ways of personal reflections upon the practice, an individual undergoes an active re-construction of her understanding of the world in order to be able to explain the practice. Within this process, an individual constructs a certain vision of the social system where the practice takes place. If the imagined vision of the practice aligns with an individual's sense of what is and should be, a learner can potentially identify with the practice. Within this process, an individual coordinates previous 'perspectives, interpretations, actions, and contexts' with that of the emergent community. Wenger (2000) described this process of alignment of what is observed with what is imagined as key to understanding how attachment to a group is assigned.

In sum, the theory of CoPs offers an analytical framework that highlights elements of participation in a social practice that lead to the emergence of shared identity. Understanding how these elements are manifested in a forum core are of relevance due to a forum core's potential to evolve into an identity-based community, i.e. a social entity motivated and governed by individuals' attachment to the group's purpose.

7.3 Research Question

To gain insight into what motivated forum core posters to mutually engage in forum practices beyond the actual learning of the subject matter, a qualitative analysis of a smaller sub-set of the discussion threads was guided by the following research question:

What were the forum core posters talking about within social-task, social non-task and mixed discussion threads?

The issues that forum core members engaged in within these threads are further interpreted through the CoP theoretical lens. CoP constructs were employed in order to explore how identity formation and group attachment were enacted through forum core communication.

7.4 Methods

To address the research question presented in the previous section, social non-task, social-task and mixed discussion threads within each of the four course examined were extracted from the forum data. The number of threads in each course ranged from 91 to 357 (Table 7.1).

Table 7.1 Number of non-curriculum discussions analysed across the Four MOOCs

	Course			
	FP	Excel	Solar	Water
Discussions, N	91	188	357	238

The discussion threads were coded in an open manner using line-by-line coding within each course. Then, codes across the courses were merged and structured into larger categories. The nvivo codes were subsequently organised to represent dominant themes that were found to cut across the analysed discussions of the four courses. Section 7.5 presents three major themes that emerged from the coding.

7.5 Forum Core Discussions Beyond the Curriculum

A thematic analysis demonstrated that in the discussions of a social nature (social-task, social non-task and mixed), forum core posters conversed about course assessment, teacher presence and teaching, as well as course organisation and MOOC forums. These topics allowed learners to exchange with their different opinions about MOOC settings. Through exchanges of the divergent expectations of being a MOOC learner, individuals revealed their assumptions about the values of education, role of the teacher, and purpose of social activity in a MOOC. The ways of communicating in the forums and resolving these tensions prodded negotiations of norms in the forums.

Discussions along the themes of (1) assessment, (2) the role of the teacher, (3) the course and teaching and (4) communication, collectively created shared and (5) intersubjective rules and norms as aspects of shared identity. Discussing the situational norms around communication and the MOOC itself complemented curriculum-related exchanges, and comprised the emergent shared repertoire of group-related beliefs.

This section describes these prominent themes and exemplifies them with participants' quotes. The role of these discussions in the identity formation of a forum core is addressed in the next section.

7.5.1 Discussing Assessment

Assessment was a subject taken-up by the forum core in each of the four examined courses. Forum core posters felt compelled to address the types of questions posed in assessments, alignment correspondence between assessments and lectures, and assessment policy. For instance, across all four courses, forum core learners had reacted to the course's assessment design. In debating the relevance of their respective teacher-chosen assessment design, forum core posters expressed their distinct views on the validity of both open-ended and closed-ended questions for learning. For instance, one learner pointed out a disconnect she observed between the assessment and what she perceived was being taught: "Some questions remind me of my kids 'spot the difference' puzzles - nothing much to do with programming". In a similar manner, common to discussions were the number of attempts allowed on graded quizzes within assessment policies. In courses where quizzes and assignments allowed for only one submission attempt, many learners publicly deemed them as being too narrow and performance-driven:

> I strongly agree that for HW [homework] there should be at least **2 attempts**. Cause If I got result very bad I see no point to learn here anymore cause I can't write test again like in University to improve my final grade. And for task like that it should be more attempts.
>
> I think you've made an error in only offering one submission per question. Would the Karate Kid have learned anything with only one attempt to get each move perfected?
>
> In my opinion, we are not trying to get another degree here. It is a training and getting acquainted with basics of water and waste water treatment.

In addressing their agreement or disagreement with the status quo, posters often espoused their beliefs about learning and performance. They extrapolated what they

observed as valuable in university education, and whether MOOCs were similar or different from formal educational experiences. As demonstrated in the above quotes, the value of learning was often juxtaposed to the value of credentialing or achieving. Learners, however, differed in their understanding of how that applied to their experiences in a MOOC. Along with proponents of the multiple-attempt assessment policy, there were learners who believed the opposite:

> The point of the test is to let the lecturer know how you are progressing, and hopefully yourself. If you can re-answer, after seeing the answers, what's the point? You have seen the answer and hopefully realized what you did wrong. Learn from your mistake! If you are making silly mistakes, like I have too, learn from them and learn to read and understand the question before answering!
>
> [The Teacher] is setting his standard here which I think is totally fair. Just think harder and take responsibility for your answer.
>
> For my part, I want to learn, I want to get the certificate, and I trust the number of answers was calibrated for the vast majority to pass the certificate. There are no grades attached to the certificate. Yes, learning is the objective, and everyone has the chance to watch the videos as many times as they want, read, take notes, think about the questions and answers.

Discussions about assessment could lead to self-disclosing one's attitudes around learning. For instance, one particular learner questioned what counts as learning: "I think it is better when questions have the option of writing WHY are you choosing what you are choosing. This way, you actually see what students did learn, and evaluate critical thinking". As shown, the learner expected to be tested on understanding rather than memorisation. Similarly, some posters questioned if assessments should at all address curriculum areas that were not explicitly covered by the teacher.

In addition to what counts as learning, another forum core value espoused through discussions on assessment was that of failure. The learners associated the number of attempts for graded assessments with the likelihood of making a mistake. As noted by this learner: "I don't feel that homework is the place to implement a policy of having no margin for error …Learning means being able to fail". Such conversations uncovered attitudes developed in prior formal experiences: "I'm here to learn stuff, not being scared by a grading process". The statements about attitudes to learning were often strong and disclosing. They may be relevant to the creation of a learning culture where individuals are openly acknowledging the difficulty and effort implicit within their learning processes. Such vulnerability would also encourage other less confident learners to seek help through the forums more actively.

To summarise, the MOOCs' assessment design triggered negotiations of community views on failure in learning, reasons for learning, and ways of assessing knowledge. As learners reflected on something in assessment they felt strongly about, they at times continued to espouse why they held such beliefs. Through discussing these issues, the learners appeared to express their views on what they imagined as the "norm" for learning in a MOOC, and what underlying values towards learning they held.

7.5.2 Discussing Teacher, Staff and Course Organisation

Another popular topic of non-curriculum discussions was the role of a teacher. Similar to learners' views on assessment, a teacher's involvement, presence and decisions in the MOOC appeared to concern learners. Analyses demonstrated that a diverse student body carried distinct expectations of whether and how much a teacher should be involved in the teaching-learning process. In this area, learners also appeared to have strong and differing opinions. Often, posters were certain that a good and effective teacher needs to be spending more time in the forums. Such an expectation is exemplified through the following quote:

> My personal experience is that staff involvement during the course is a key success factor for a MOOC. The MOOCs that I enjoyed most and that I stuck with until the end, were mostly the MOOCs were the professor made a connection to the student and showed their passion for (the quality of) their course via emails, announcement and mostly via the forum. This course seems to be prepared well, and [this teacher] seems an energetic and enthusiastic person. I'm therefore even more surprised to not see him, nor anybody else (except a tiny little bit of [staff member's name]) here.

These expectations were not limited to instructors teaching the courses, and would often be extended to all the teaching staff in a MOOC. The plea for staff to participate was captured within this quote: "One of the most important issues STAFF must work is to be more present in the POST area, since otherwise we, students, can feel alone, disconnected". Although learners expected their teachers and staff to engage, expectations for the extent of such engagement were varied. It is plausible to assume that learners' cultural backgrounds played an important role in their expectations upon the role of a teacher in forum interactions. These divergent approaches to a teacher's role were well reflected through one's way of talking about and to the teacher. For instance, for this learner, the teacher's role in the course was that of a role model, as he wrote: "After all, we do look up at our teachers and expect them to be perfect, don't we?" For other students, however, the teacher became the object of critique, at times

accompanied with quite strong language, or, the teacher became an object of attempts for friendship. Posts made as remarks about teachers beyond their ways of teaching, reactions, or involvement in the courses exemplify underlying differences between what was deemed as acceptable talk towards a teacher, and about a teacher. Remarkable, some posts extended beyond how much a teacher should engage in forums, and included comments on teachers' personality or quirky characteristics, gender, clothing or manners. Such remarks are known to be impolite, especially in a face-to-face setting or in smaller online classes where members can be identified. The quotes below exemplify the types of remarks posters permitted themselves in the MOOC settings:

> @[teacher_username]: The energy is great! Don't lose it! I do agree, though, that when your movement brings you in front of the slides, blocking them, it is distracting and makes the slides hard to follow.
>
> Maybe i enjoy the teachings of the male professors than the female ones, or maybe i was on a different level of understanding.
>
> The third week T-shirt of the instructor is not in par with the two preceding amazing shirts. Is there a subliminal message.
>
> … next time, may I suggest you have the presenter practice it a few times before taping? There were a lot of pauses and 'um's and a few 'oh, wait, I have to go back here' moments that could have been easily avoided by a few practice runs before taping.

As shown, these messages depict direct criticisms of an instructor's movements, gender, clothes, and/or unpreparedness. Some of the criticisms contain tinges of bullying, likely influenced by the anonymity aspects of the forum environment. The remarks, as previously noted, were not always insulting. Some were made in the spirit of comradery, as the one below:

> I must compliment the lecturer on clothing style – it's surprisingly brilliant idea! Once I open the folder with all the lectures downloaded I immediately see which files belong together because they have same shirt in them. I do hope you won't run out of new shirts before the end of this course :)

Course design was yet another area that forum core posters commented on. Learners' remarks about the courses' design are not surprising because a course's organisation, choices of materials, and teaching as such, are all decisions associated with the teacher, and manifest teacher presence in the course (Anderson, Rourke, Garrison, & Archer, 2001). Much like attacking critiques that are addressed to a teacher, the comments targeting course organisation were also strongly voiced through a range of comments across the four courses. It appears that critiques of teaching competence was even

stronger than that of the personal remarks, as highlighted through the selected quotes below:

> When 'inverting the classroom' for 22,000 students – most of whom will have no support other than forums – it is essential to work on both presentation and content. You clearly have the TCK, but please read http://en.wikipedia.org/wiki/Technological Pedagogical Content Knowledge and accept constructive criticism.
>
> If you're going to 'teach' a formula, actually teach the formula. A small screen flitting it behind you without providing additional description/instruction on how to structure the elements of the formula is pretty useless.
>
> [To the teacher about assessment] That was a very badly constructed question. If you are going to do this, please do it properly. Be clear on what you are asking. Define what sheets you are asking students to work on. Please try and appreciate that students are doing this remotely.
>
> The course just give[s] a first impression of a poorly thought out learning experience, including the exercises which didn't really make me feel like I had learned or achieved anything.

These quotes demonstrate that forum core posters freely critiqued their teachers' competence. Another area of the course design that posters at times chose to discuss was that of the stated course pre-requisites. Learners venting around this particular issue was captured through the following quote: "This course said 'no prerequisites' but I really think they should have mentioned that you'd do better with some chemistry under your belt". Alternative critiques included misalignments between lectures and assessments, a lack of practical application of taught content, and pedagogy that reaches beyond the course. The quotes below demonstrate learner frustrations around some of these issues:

> …there is no rational relationship between conceptual content of lectures and mathematical equations contained in the assignments.
>
> I'm not against learning new stuff. I'm against learning useless, meaningless or just broken stuff.
>
> This course sucks – very rudimentary teaching, you are better off just using a book or website and read about the subject.

Tensions between the academic and the practical aspects of knowledge were brought up by the students as well: "Sorry, but I'm interested in solving real problems rather than learning another weird syntax of academic language that did not move one step further since 1990". These discussions were particularly present in the MOOCs that attracted a higher number of professionals who were confident to express both their possessed

background knowledge and opinions about how courses should be taught. Disagreements related to a course's needing to be more practical or more academic can be explained by the wide range of learner diversity. MOOC participants who were university students could have been more used to conventional academic formats of teaching, also hugely dependent on context, while professionals whose daily application of taught knowledge is within the practice domain might have been the ones arguing for the value of a more practical perspective to be taken on by teachers.

The sub-section highlighted that in addition to assessment, the role of the teacher, teaching behaviours and choices around teaching itself were subject of learner discussions. Strikingly, the language chosen to express some of the reflections, as well as the points raised, may appear as quite inappropriate for face-to-face settings. Similar negative sentiments on MOOC forums were previously addressed by some MOOC instructors. For instance, Comer and colleagues (2015) reported that they experienced a particularly high number of negative comments from learners when teaching a MOOC. Examples presented by Comer and colleagues observe a similar tone and level of critique to the quotes from this study's four examined MOOCs. Critical and negative discussion threads did not simply voice strong opinions, but put forth inappropriate prodding disagreements between posters. Such negativity could be explained by the so-called deindividuation theory positing that in online environments, the effect of what is being said upon another person is not immediately observed. Besides, speakers lack interpersonal affiliation to moderate the strong expression of opinions. Hence, individuals may not exercise the conventional amount of consideration and respect towards another person, as they may in a face-to-face setting. Prior research established that such effects of deindividuation that disrupt the conventional norms of communication are exercised in CMC environments prior to the establishment of mutual accountability (Postmes & Spears, 1998). Thus, it comes as no surprise that negative expressions of views around the courses' issues eventually prodded discussions about the forums' rules and norms.

7.5.3 Discussing Communication, Rules and Norms

Beyond discussions around what learning is for, and how the teacher should facilitate it, a MOOC's forum, its organisation and ways to communicate became a dominant theme in the threads of a social nature. Ultimately, discussions around forum use were

triggered by learners' inability to find what they expected to within forums, or by their experiences of other posters' communicating in ways they found unacceptable. These dissonances prompted learners to share opinions about forum use, and negotiate as to how the forum should be used collectively. In phrasing their disagreements around forum communication and use, in similar ways to their expressions around course assessment, design and teachers, some learners chose strong and rather impolite language, and others felt that such ways of communication were not acceptable. This sub-section briefly reviews the discussions that arose in relation to the purpose of forums and forum behaviour.

Purpose of Social Learning and Ways of Communicating. As the courses progressed, learners engaged in negotiating acceptable ways of communicating to one another in MOOC forums. Discussions around forum behaviour would often start as a reaction to critique of a teacher, course organisation, and similar issues mentioned earlier. The distinct feature of the discussions about behaviour was that they were not focused on the teacher's behaviour or decisions. Here, learners discussed how they themselves should act in the shared forum spaces. The quotes below exemplify a range of behaviours negotiated by learners, such as being aggressive, being critical and judgemental, and courteous:

> That was quite aggressive... I am feeling ill at ease reading that. I think everyone can post their opinion without throwing stones at each other, let alone at the entity providing us with this great opportunity of learning.

In terms of the ways in which the forum core's participants' posting were being challenged, the MOOCs' discourses reflected a variety of advices on how people should behave, ranging from: from "we need to be more analytics and less critical" to "let's be more polite" to "I don't think we should be attributing thoughts and feelings (judgmental). …We all need to be courteous and civil to both the course organizers and each other." Discussion around how to communicate reached beyond the choices of one's tone. Learners also discussed what was appropriate to post and when, e.g. if an answer or hint could be posted before a submission deadline expired, as well as how to structure the forums, where to place them in the course, and what forums were there for in the first instance. In this sense, discussions around forum use espoused learner attitudes towards the purpose of social learning in a MOOC and the values forum posters held about it. Some learners explicitly stated their interest in learning and intent

on how to go about it: getting the information they needed was the ultimate outcome. Others were more interested in the climate or inclusivity of conversations that would embrace diversity. For instance, as demonstrated in the quote below, a particular poster was interested in the collective power of learning from others:

> If you did not understand, ask and many will help you (*this really happens*) and if you already know, teach others and that teaching will help you to consolidate your knowledge ***even further***... the biggest treasure in a MOOC are the **Discussion Forums**.

Although this may generally reflect what other people think about forums as well, translating these expectations into social learning practices were underpinned by different ideas. Among examples of the divergence of opinions were discussions as to what language was acceptable to use in the forums. In some courses, posters would insist on using English only, while in others, learners appeared content in exchanging posts in their own languages. Using English exclusively was not an issue for all MOOCs. For instance, in the Water Treatment MOOC, the students openly engaged in forum threads in other languages; mainly Spanish. In Solar, however, there were participants, and teaching assistants for whom this was an issue: "If everybody decided to write in his language then this POST area would lose a lot in participation" or "Just imagine if everybody decides to write in their home language: it will be a mess and no one will talk to no one." Clearly, for some learners, connecting to other individuals by using a common language was an easier and faster way of enabling the exchange of information. For others it meant not being inclusive, against what they imagined the ethos of an intercultural course should be, or even against their expected learning outcomes, if these included the practice of English use.

Emergent Ways of Forum Use. Discussions about what was acceptable in terms of communicating to one another were reified into accepted practices, as well as rules and norms of behaviour in forum groups. The reification of the purpose of the forums, i.e. shared practice, was best observed when the rules as to how to make a post were established. It appeared that the majority of posters seemed to perceive the forum, first and foremost, as a place for sharing, finding and exchanging information. Consequently, some posters demanded that questions be asked in a way that the others could find them, and, most importantly, understand them. In every forum there were individuals who made sure to remind others about the rules of posting. A common advice here was to post in the 'right' place, not to post same text twice, and to ask precise and specific

questions. More elaborate users suggested that forum threads be re-structured in a way they thought would work better, and some encouraged others to upvote good answers – although even upvoting was a subject of debate in one of the MOOCs, as some posters felt it promoted individuality and damaged group attitudes. Regardless, once a norm had been established for seeking answers, for example, unacceptable ways of going about this would be met by a demand to comply by the rules as a way of getting the information. For instance, two students gave the following advice to a newcomer seeking information in one of the forums:

> Person A: Hi [student's username], I have no idea what the title of your post is supposed to mean. Did you have a question? Was it in the title of the post? All you posted was a formula and although that formulation certainly works if you first sort your data by rep it's probably not the most flexible. Now what if I asked you for the bonus for Jones or Sorvino? Are you going to have to change the range of your formula every time I ask for a different bonus? By the way, since this does lead to an actual solution, I removed your formula. Please do not post solutions or formulas that lead to the solutions to exercises. Please review the [Collaboration Guidelines][1] and [Forum Guidelines][2] so that everyone has an opportunity to truly learn the material. [1]: https://courses.edx.org/c4x/DelftX/EX101x/asset/collaboration guidelines.pdf [2]:
>
> Person B: Also, please, use only English language for the forum discussion, thank you!

The excerpt above represents an all-encompassing set of rules in a forum interested in efficient question-and-answer exchanges. Sets had been previously negotiated, and were subsequently offered to a newcomer for obligatory compliance. The discussion excerpt also illustrates a commonly negotiated concern of whether solutions to graded tasks should be posted, and if, so, when and how. The issue of 'cheating' or 'helping' can have multiple interpretations in divergent cultures and contexts. Offering a hint to the graded task can be approached as cheating, yet, a hint may really be related to intermediary feedback helping a learner get unstuck in order to progress in solving a quiz. Aside from the simple decision of whether to post answers or not, learners also negotiated ways around the problem of cheating. For instance, an answer did not always have to be revealed, as others could help one another get 'unstuck' by asking questions, or sharing their mistakes when undertaking assignments. Decisions around how and whether to share and support each other before the final deadline of a task differed across the four courses. Sometimes the way the 'rules' came about depended on staff decisions. For instance, in the Excel course, staff adopted a very strict moderating policy and anything that went beyond the guidelines for forum use was heavily

punished by being cleaned out from the forum by the moderators. Thus, friendly fellow students would warn each other:

> **PLEASE edit your POST urgent!!!!** You are **showing the answers**. That is ***against*** the rules. PUT **XXX** and **XXX-1**, they will understand it …. That ***is not the reason you got it wrong***.

In contrast, in Solar and Water, learners negotiated on their own that posting solutions was not acceptable. Many students wanted to discover the answers on their own: "The intention is however for students to help other students with their questions, as long as they don't give the solutions". In Solar, many students agreed that they liked doing "all the problems … without help from fellow students, but I don't mind instructor hints". So they had adopted hints as a way of discussing contested areas. In FP, no hard policy around revealing answers or hints was formulated. However, a teacher introduced a so-called EULA for learning, i.e. a humorous end-user license agreement that requested learners to acknowledge their familiarity with the lectures prior to engaging with the homework. Therefore, if somebody sought advice on a graded assignment, the learners would first refer to the EULA, e.g. "I think you clearly violated the EULA for this week!!!!". Such a comment would suggest that the answer to the question was within the lecture. If unsure, students would ask the teacher, who had a very strong forum presence: "Can I post the details here? Or is it against forum rules?". Such statements signalled that in this particular course, revealing the hints and answers could be negotiated through individual cases, between a student and the teacher. In fact, the in this course, the teacher did offer hints, seemingly being more generous as the course came to a close.

Reification of Forum Use Practices. The previous section presented various instanced of negotiations of rules and norms. While in progress in a discussion, however, it may have been difficult to discern what the final decision on a particular issue was. Yet, through repeated instances and experimentation, norms were eventually set. These formed norms were best observed within encounters between new forum users, arriving half-way through a course, and a forum core's welcoming of them. For instance, posters who joined a MOOC late would commonly ask if it was still okay to join, and where to start. Quite often, newcomers were encouraged to remain:

> Language is not really a barrier for this course...if someone is an expert with the language used, same person could be weak with the math for example!! So dont

> worry about it, just put in your best efforts on this course and you WILL successfully complete it!! Good Luck.

Active students and teacher assistants would serve as gatekeepers, also due to their broad participation history within the course. They were often the ones introducing the rules and norms, with some excitement or politeness as the latecomers appeared. However, the gatekeepers did not welcome re-negotiations of norms. For instance, the quote below is a forum core poster's response to a newcomer who joined an older conversation related to norms and rules:

> You are in disagreement [to what we have said here before you posted], because you were not in here in the first 10 days of this course. Now, thanks to everybody feedbacks, things are going exponentially better. The time-frame of your comment is far away from the context at with this topic was written. Take in easy on how do you write comments, ok? You are commenting [on] a post written 2 weeks ago.

This last quote highlights how re-opening the negotiated elements of the forum's social practice was not encouraged by this particular poster. In some courses, learners could engage in re-negotiating the norms based on specific instances. A poster has asked, for example, "Can I post this?", in reference to breaking an existing rule, prior to the actual transgression of it. In the FP course, both the teacher and learners made exceptions to the general norms of forum behaviour in order to accommodate specific circumstances. These different agreements around a social practice of forum use can be interpreted as a representation of the existing mechanisms of mutual accountability against an established situational norm.

Forum Decisions and the Powerful Others. Different aspects of how to use the forums created tensions in the relationships between learners and teachers, or learners and staff, as well as between learners. The language used by posters demonstrated that they often self-categorised as a group of learners with a shared identity. Such self-categorisations would be characteristic of posters seeking interpersonal interaction or human support through forums. Forum core posters who appeared to have a stronger commitment to shaping their forums, self-categorised as distinct from teachers, or other figures helping a MOOC's decision-making. The circumstances that prodded posters to differentiate themselves from a 'powerful other' differed across the courses. Yet, all the courses had a 'powerful other' to stand across from in attempts at negotiating arrangement for their forum. For instance, in the programing course, a teacher with a strong forum presence and personality was the sole receiver of both positive and

negative attention from learners. The learners revered his behaviour and supported him, or strongly opposed his choices and challenged him. In the unmoderated Water Treatment MOOC, students directed their feedback to an external ephemeral staff. Attempts to make changes or seek advice from staff was helping learners acquire their group identity. In Solar, posters similarly grouped as distinct from the staff whose attention they were seeking – strongly voicing it. In addition, in Solar, the policies around teaching assistants were not clear, yet they were also grouped as the powerful others, along with the teaching staff. Interestingly, in the Excel course, the "powerful other" in was neither the teacher, nor the teaching assistants. The teaching team used carefully chosen language indicating that they identified with posters and their problems. Teaching staff, and particularly one teaching assistant acknowledged some of the concerns projected by learners by attributing problems to quite a different "powerful other" – the MOOC provider and the platform. The teaching assistant explained that there were limitations embedded within, and that staff could only do so much. Hence, learners often addressed their concerns to a responsible third-party, referred to as "the edX". That is, in an Excel course, it was often the edX that needed to realise what they, learners wanted and needed, by creating a better learning environment for them. Through these examples, it appears that the presence of a 'powerful other' facilitated the formation of group identity, and became another power-holder to which a forming group could increasingly direct shared criticisms in an effort to changing the status quo.

Although the current analysis did not focus on the linguistic features of discussion threads, tensions between the in-group and the out-group 'powerful others' were well-captured within the language used by forum posters. The us-them tension between the emergent group of posters and the 'powerful others' was well-reflected through the use of pronouns within the discourse. For instance, some posters had observed that "it is just us students" present in the forums, and that "they" [staff] need to consider "our feedback" so "we can all pass the course", and that the course will be improving "after so many problems we all faced (students and STAFF)". The last quote particularly demonstrates that for a student, "we" does not include STAFF, and so additional explanations are used to make the "we" more inclusive. These linguistic observations support the supposition that group identity was gradually emerging in relation to in-group and out-group decisions.

7.5.4 Summary

Qualitative analysis of the discussion threads of social nature in the forum cores' exchanges demonstrated that some learners found forums useful for discussing their MOOC's assessments, course organisation and facilitation, as well as the teacher. The way a forum was used within the course, and how learners acted within it, prompted negotiating rules and norms for forum use. The interpretation of these discussions through the CoP lens is presented in the next section.

7.6 Discussion

To gain insight into the issues of forum core discussions beyond the subject matter, this chapter presented a qualitative analysis of social-task, social non-task and mixed discussion threads. Analyses demonstrated that discussion threads of a social nature centred on issues related to the organisation of a MOOC and its forum. This section interprets these issues in relation to the formation of a forum core as an emergent community of practice (sub-section 7.6.1). Further, the negotiation of situational norms in relation to MOOCs and their forums is presented as a mechanism of identity-formation in a MOOC setting. The section also explores the link between the negotiation of situational norms, negotiated control obtained by posters, and the shortening of psychological distance between learners, i.e. growth of relatedness. The discussion concludes with implications in light of this analysis.

7.6.1 Forum Core as an Emergent Community of Practice (CoP)

The forum core posters mutually engaged in discussing the following aspects of their shared experience:

- role of the teacher, including teacher/staff presence, competence and personalities;
- assessment, discussing if it was unfair, had too few attempts, testing what had not been taught;
- course/teaching, including misalignment with one's expectations of what the course should have been, whether such content should be taught differently, whether the course should be more academic or more practical, etc.;
- communication, addressing the ways in which people communicate, or poor forum organisation resulting in an inability to find information or communication in a forum.

These thematic issues found within the discussion threads that were social in nature support the understanding of a forum core as a CoP. Three elements define a CoP. First, mutual engagement constituted collective engagement into the matters of collective importance. Analyses in this chapter demonstrate that although the MOOCs' forum cores were curriculum-driven learners, posters engaged in discussions that related to learning and forums, which are at the heart of their shared learning experiences. The second element of a CoP is that of a joint enterprise. Wenger (1998) noted that joint enterprise, or a shared domain of practice, is not necessarily the stated purpose of a community, but the negotiated one (p.80). Analyses in this chapter have demonstrated that the forum cores negotiated the meaning of their joint enterprise, i.e. social learning in an open course. And in line with Wenger, joint enterprise also involves the creation of mutual accountability. Examples in the analyses section highlighted that learners eventually settled on inter-subjective (non-consensual) ways of behaving and acting in the forums. It has been noted that these behaviours, i.e. how to talk to one another, how to ask for help and when to post, became monitored by the more active learners. When implementing the regime of accountability, forum core posters referred to previous conversations, posting links to previous discussions where negotiations has taken place. In that, forum negotiations themselves were turned into a shared repertoire, i.e. a reification of the social practices of forum use and learning, collectively. In other words, it can be said that by debating the course set-up and discussing setting-related issues, the forum core evolved by creating a shared domain of practice and a shared space for social learning.

7.6.1 Mechanisms of Identity Formation in a Forum Core

Learning shared identity as a group is another element defining the emergence of a CoP. This section proposes a framework that can help analyse if a forum core has developed a shared identity. A forum core's identity could develop as a result of learners obtaining negotiated control over some course elements (including forum-related issues), triggered by discussions around situational norms. Through the process of negotiating, learners may identify with emergent situational norms, and with the salient group identity. Such mechanisms can be explained by Dron's (2007) transactional control theory, Wenger's modes of identification with the CoP (Wenger, 2000), and the role of situational norms in an environment that holds heightened anonymity. Combined, these

elements, supported by this study's empirical insights, offer an interpretation of how a forum core establishes a common identity.

7.6.1.1 Negotiating Control as a Mechanism of Relatedness

Dron's (2007) theory of negotiated control asserted that learners who negotiate control over course elements can successfully acquire a higher sense of closeness between one another since their transactional distance is shortened. To note, Moore's (1993) seminal concept of transactional distance refers to the psychological space between individuals separated by the space and time embedded within a web of relationships in a distance course. According to Moore's theory, this distance is shortened through emergent dialogue between the structure imposed by the teacher and the autonomy of the learners. In other words, interpersonal relationships are established through the dialogue. Dron (2007) posited that negotiations of control between the two parties serves as the mechanism that invokes dialogue to decrease transactional distance. In other words, by having negotiated control over some areas in their course, learners and instructors can shorten the transactional distance between one another. Given that transactional distance is a psychological construct, its decrease indicates the emergence of a relational quality in interpersonal interactions.

Analyses showed that the four examined forum cores engaged in negotiating various aspects pertinent to forum use and the courses. Given the autonomy of the learners in the analysed courses to take ownership over the forum space, it could be argued that the forum cores, as emergent CoPs, were negotiating control over their shared domain of practice, learning in a MOOC through its forums. Analyses also showed that a forum core was juxtaposing itself against an out-group, or a 'powerful other'. This 'powerful other' represented a force each forum core may have been trying to negotiate with. In courses where a forum core took ownership (alone or with the instructor) of their shared space (Solar and FP), control over some areas of the joint enterprise was successfully negotiated. Through the premise of Dron's (2007) transactional control theory, it can be argued that the successful negotiation of control over certain issues in a forum resulted in higher levels of relatedness between the learners who negotiated, and a shortened transactional distance. If so, an effective negotiation of control represents a mechanism that increases the quality of relatedness, underpinning a forum core's interactions.

7.6.1.2 Negotiating Identity Through Situational Norms

Analyses in the chapter suggested that learners negotiated control over various course issues, addressing MOOCs' assessments, forum rules, and a course's organisation. The discussions around these issues largely revolved around situational norms, i.e. norms specific to the participant groups and possibly different to other groups and other situations (Postmes & Spears, 1998). Negotiations and the emergence of situational norms are critical in group identity for large groups that have a heightened deindividuation, i.e. where the salience of individual unique selves is low. Postmes and Spears conducted a meta-analysis to identify the effects of deindividuation, i.e. the lack of salience of the self, on the antinormative behaviour in computer-mediated settings. Their meta-analysis demonstrated that a situational norm impacted social behaviour through group identity in settings where deindividuation was heightened. According to Postmes and Spears (2010), when individuals were not familiar with one another, or awareness was reduced, participants' sensitivity to situational norms would increase, along with their responsiveness to cues indicating what is appropriate and desirable in a particular context (p. 10). By implementing a research agenda to support their Social Identity Deindividuation Effects (SIDE) theory, Postmes and Spears (1998) showed that anonymity, large groups and reduced self-awareness had actually caused participants to behave more in accord with established social norm (as opposed to antinormatively). They further noted that antinormative behaviour associated with negativity of discourse, for instance, would still appear in settings characterised by heightened anonymity, in cases when accountability for one's actions has not yet been established.

From the perspective of the SIDE theory, the negativity towards instructors and other peers uncovered through this study's analysis can be interpreted as resulting from deindividuation, and taking place prior to the establishment of mutual accountability within the social practice of forum use. In light of the work by Postmes and Spears (ibid.), the negotiation and presence of the negotiation of situational norms signal and in part constitute a process of identity formation at a group-level. Since peculiar features of the individual posters in this study's four examined MOOCs cannot be differentiated due to scale and anonymity of the forum, the learners' behaviour and espoused beliefs about learning (or other negotiated issues) created the discourse that other forum core posters could relate to. In that, negotiations of the group-specific situational norms constituted the shared identity in the four MOOCs' forum cores.

7.6.1.3 Wenger's Modes of Identification

The successful negotiation of situational norms in a MOOC has been suggested as a plausible way that forum cores develop a shared identity, as well as an emotional underpinning to their interpersonal interactions. Wenger's (2010) modes of identification, previously referred to as modes of belonging, help explain how individual attachment may take place within such a process.

According to Wenger (2010), the production of identity is the outcome of learning in a CoP. In describing membership, Wenger noted that it is defined by an individual's identification with shared practice (p.3). Wenger further distinguished three stages of how an individual identifies with shared practice, i.e. establishes a perceived membership in a CoP: engagement, imagination and alignment. By mutually engaging in a MOOC, its forum core posters had shared experiences that may have been incongruous with their expectations. As demonstrated through the literature review, issues of assessment, course organisation, and learner forum behaviour, are all underpinned by varied values and expectations. Forum core posters accumulated these expectations and underlying assumptions about learning through their past educational and cultural experiences. In times of misalignment, i.e. when what occurred in the course failed to align with what learners thought should happen, some learners shared their views with others in their forums. The sharing could be seen as the process of imagination and alignment described by Wenger as integral to one's identification within a CoP. Most importantly, some of this sharing took the form of collective negotiations around one's expectations and beliefs. When expressing their opinions and "imaginations" of what could be, the posters also had room for disagreement. Disagreements usually occurred around open-ended issues, i.e. why lectures were organised in a certain way, why assessments tested a certain concept, why somebody should act differently in a forum, etc. At times, these negotiations resulted in changes of policies or the emergence of behavioural norms. The platform of the forum kept these negotiations as a reified history. If the change of a norm or a decision took place, a forum poster could then align her expectations with the newly emerged status quo. Such alignment would result in self-identification with the emergent group. Alternatively, a poster could self-categorise as distinct from the common practice.

These mechanisms of mutual engagement, imagination and alignment, described by Wenger (2000) as modes of identification, overlap with the processes of self-

identification and self-categorisation present in the theory of identity-based communities. Self-identification, a process of cognitively associating oneself with the collective, takes place along with the process of self-categorisation, when a learner first defines herself in relation to others. This process in a forum core is limited due to the lack of salience of other personalities. Alternatively, it can take place in relation to what values and opinions are espoused, rather than who the learners are. These contextual situational norms offer forum core posters opportunities to identify with the common practices and their underpinning values, therefore, aligning their personal identities with those of the group.

7.6.2 Implications

There are several relevant implications stemming from the analyses. First, analyses bring to the fore the need to design spaces that allow for group attachment and the development of group culture. The analyses showed that forum cores are more than a collective, they are groups with evolving identities. Situational norms that forge a group's identity rely on the purpose of the group, and depend on the motivation of the critical mass of forum users. As noted by Ren et al. (2011), people attached to a group's identity, want to be a part of that community and advance its purpose. Also revealed through the analyses, disagreements regarding the purpose of a social activity is revealed through various discussions about course set-up, where the meaning of collective experience and collective practice of learning socially is negotiated. The purpose of social learning in a MOOC needs to firstly be established and negotiated, along with the terms as to how it will be upheld in a specific circumstance. Negotiations around the purpose, and expectations, that drive forward the formation of a forum core, can to some degree be explained by the varying motivations of learners to come to a forum in a MOOC. As previously noted drawing on the residents and visitors framework by White and Le Cornu (2011), the motivations for using a forum's online space may differ: as some learners are more interested in the efficiency of information exchange, while others are more oriented towards the relational quality of a forum as a common space for learning. The culture of learning residing in these two perspectives may differ. In that, it appears that the role of the teacher and active posters in creating a forum's participation culture, underpinned by the forum's purpose and norms, is critical. What that culture becomes is manifested through the shared identity that others will either identify with or not. Hence, the instructors interested in fostering relational

qualities that underpin social interactions should focus on participating in these negotiations that take place in the early stages of a MOOC.

Second, developing a community of learners through identity-based attachment of one-to-many settings is useful where the scale and non-privacy of the environment aggravate the development of bond-based relationships. Empirical work has previously demonstrated that one's engagement in a social practice in identity-based communities has the same level of liveliness as engagement in bond-based communities (Grabowicz et al., 2013). Furthermore, Ren et al. (2012) demonstrated that identity-based attachment has stronger impact on the frequency of individual engagement than other conditions targeting interpersonal attachment. In other words, the benefits from one's membership in common-identity communities are comparable to benefits derived from having bond-based relationships, and the anonymity of the online setting may just as well strengthen such a tendency. Research programs by Ren and colleagues (ibid.) that have identified a set of design features and principles for the development of online informal communities are therefore also applicable to MOOCs' educational settings.

7.7 Conclusions

To seek insight into the relational quality of the patterns of interactions in a MOOC forum, this chapter presented qualitative analyses of the non-curriculum discussion threads between MOOC forum core posters. The forum core posters were compared to an emergent CoP where individuals self-identified with the group's ways of social learning. Analysed discussion threads indicated that forum core members engaged in negotiating the meaning of a shared practice, and situational norms of learning together. Non-curriculum discussion threads in relation to the negotiated meanings suggested that relational processes of identity-formation unfold within a forum core.

Combined with the insights from Chapter 6, it can be concluded that forum cores are becoming identity-based communities, driven by information exchange. Insights from Chapter 6 highlighted that only some of the courses examined in this study had a higher number of non-curriculum threads and that those threads were central to the formation of the respective course's forum core. In light of this chapter's analyses, it can be concluded that the FP and Solar courses have undergone identity-formation processes to a larger extent than the other two MOOCs.

The next chapter investigates the quantitative indicators capturing the structure of forum cores as becoming identity-based communities. Through the features of reciprocity, the structure of different forum cores is quantified to gain insight into the social processes underlying the formation of network structures in forum cores.

8 Modelling a forum core as an identity-based community

This chapter forms the last of the empirical studies of the thesis. In this chapter, the network structure of the forum core is modelled through exponential random graph modelling (ERGM). The prominent network forms of the forum core structure serve as indicators of the social context. The analysis reveals the differences in the network forms across four MOOC forums.

8.1 Introduction

Chapters 6–7 identified that a forum core evolves as an identity-based community, based on information exchange. In such a community, curriculum and task-related questions drive the majority of forum core exchanges (Chapter 6). Non-curriculum discussions within setting-related and self-related forum threads facilitate the processes of attachment to the group's identity. Through these discussion threads, a forum core negotiates situational norms and a shared domain of collective learning in a MOOC, resulting in a shared identity (Chapter 7). Consequently, a forum core community evolves through its members' attachment to the group, not through interpersonal trust. Findings from the previous chapters indicate that the social context in a MOOC setting is comprised of the interactions of information exchange and negotiations of an emergent group identity.

Indicators of the social context can demonstrate the differences in the development of forum cores across analysed MOOC. Such indicators are derived in this chapter through the modelling of four forum cores' network structures. A forum core's structure is hypothesised through empirical insights about informal identity-based communities. The formation of informal online communities has been previously explained through individuals' altruistic motivation to contribute to the collective benefit (Wasko & Faraj, 2005). Moreover, prior research examined motivation and network structure in Q&A forums (Nam, Ackerman, & Adamic, 2009; Oh, 2012; P. Wu & Korfiatis, 2013; Yang, Morris, Teevan, Adamic, & Ackerman, 2011), in communities of open source coders (Baytiyeh & Pfaffman, 2010; Hars & Ou, 2001; C. Wu, Gerlach, & Young, 2007), and within Wikipedia's editing practices (Kuznetsov, 2006; Prasarnphanich & Wagner,

2009). These studies suggested that altruistic participation is at the basis of identity-based communities, and can be reflected through a range of local network forms (structures). For instance, altruistic participation can be captured by network configurations representing different types of reciprocity. Therefore, in this chapter, forum core networks are modelled through the network configurations of direct, indirect and triadic-level reciprocity. Theories of social exchange and collective action (Blau, 1964; Ekeh, 1974; Molm, 2010; Molm, Collett, & Schaefer, 2007) are applied to interpret these network forms.

This chapter reports on the results of network modelling and interprets the differences in the indicators across four MOOCs. Prior to the network modelling, Section 8.2 justifies theoretical considerations around this chapter's analyses, and reviews related empirical studies. First, formalist network analyses reported in this chapter are distinguished from the previously implemented relational network analyses (Chapter 6–7). Such integration of theoretically distinct traditions in social network analyses is due to their capacity to offer complementary insights about the social context in MOOCs. Then, prior research on informal online identity-based communities is presented to explain the selection of modelled network patterns. The insights from this research inform the selection of network configurations to be included in the model. Section 8.2 concludes with the research question for the empirical examination. To illuminate the details around the method of analysis and ensure further replication of the study, Section 8.3 explains the foundations of exponential random graph modelling (ERGM) and elaborates the modelled configurations. Section 8.4 presents the results of network modelling and Section 8.5 discusses the study's implications and limitations, while section 8.6 outlines the conclusions.

8.2 Analysis of Network Structure

Previous chapters suggested that so-called identity-based communities can evolve within the MOOC contexts. These types of communities are characterised by individual trust in the group and interest in a common goal or intent (Prentice, Miller, & Lightdale, 1994). To derive such conceptualisation of a MOOC's forum core, relational network analyses were applied in Chapters 5–7. That is, the context of the forum cores' networks was inductively investigated through the properties of network ties and the type of content exchanged through them. To complement these relational social network

analyses, this chapter investigates the forum cores from a formalist perspective to networks. The chapter examines the structure of the forum cores (its form). Such analyses contrast the studies reported in Chapters 6 and 7 where the forum cores were investigated through the content exchanged through the networks. From a formalist perspective, specific network configurations manifest social processes underlying network formation. Formalist analyses offer another dimension of understanding to a forum core by quantifying network patterns linked to social processes. This section explains the formalist approach to network analysis and justifies the patterns included in the modelling through a literature review of informal online identity-based communities.

8.2.1 Formalist and Relational Network Analysis

Thus far, this Doctoral thesis employed relational approaches to network analysis: MOOC forum data have been inductively analysed through courses' contexts. Chapter 5 evaluated the intensity and quality of interactions between individuals as regular, occasional or regular with timely replies. Chapters 6 and 7 analysed the kinds of exchanges occurring through these interactions. Investigating the context, agency and content flowing through the ties of a network carries a socio-cultural orientation (Emirbayer & Goodwin, 1994; for review see Erikson, 2013; Mische, 2011). Simply put, within the relational socio-cultural tradition, the context of who is participating, what is being exchanged, and what kinds of ties underpin the exchanges explain a network structure.

In contrast to the relational network analyses, a formalist approach examines a network purely as a form, rather than through its content. Simmel was the pioneer of formalist network research seeking to investigate network structures as building blocks of society (Erikson, 2013). Simmel considered that social relations existed a priori contextual considerations, and hence were generalisable across different contexts. Such a philosophical premise resulted from the expansion of Kant's philosophy of natural sciences to the social sciences. Similar to Kant's quest of natural laws explaining the world, Simmel inquired about the mechanisms and rules that made society possible (Erikson, 2013, p. 225). Simmel (2009/1858–1918) aimed to uncover the laws of social relations through the forms society consists of (p. 40). The detachment of a network structure from the content exchanged through its ties is well-reflected in the methods of

data collection used in formalist network analyses. For instance, one of the questions on a general social survey (GSS), a widely used instrument deriving information about societal structure (Smith, Marsden, & Hout, 2015) asks respondents to list the individuals with whom they discuss "important matters". This survey has been in use since 1985, and has never asked to clarify the subjective meaning of "important matters". Recently Bearman and Parigi (2004) demonstrated that these 'important matters' represent various conversational topics, and only some of them could be perceived as important. However, from the formalist perspective, the content constituting 'important matters' is simply not relevant. The focus is on the structures where people are embedded rather than the differences in what important matters may constitute.

Such a focus on the structure a priori the context differentiates the formalist approach to network analysis from the relational one. Erikson (2013) explained that relational network researchers reject a priori categories and insist on the emergent intersubjectivity of experience and meaning, as well as the content of interactions. In relationalism, agency, culture and the values of the social are emphasised, and thus each analysis is highly contextual (Emirbayer & Goodwin, 1994). In contrast, formalism aims to identify a priori categories of relational types and patterns operating independently from the context or the setting. The context, according to Simmel, is defined by the form itself. Simmel conceived social forms as not constituted by transactions, but as giving "form to transactions because they [social forms] posit discrete, pre-given, and fixed entities" (as cited in Erikson, 2013, p. 225). Because formalist hypotheses are not situated in a particular setting, the patterns of the social forms can be generalisable and replicable across different contexts.

Although the philosophical disagreements between formalism and relationalism may not be resolved, the combination of the two approaches can develop social network science, and stimulate empirical research. Erikson (2013) argued that "relationalism and formalism should be considered two distinct and well-developed theoretical foundations for social network research" (p. 236). Despite the underlying inconsistencies between formalism and relationalism, Erikson (2013) asserted that the two approaches illuminate different parts of an examined phenomenon, content (i.e. culture) and form (i.e. structure). Due to their complementary insights, Erikson urged that both research traditions should be encouraged. Conforming to such a research agenda, this chapter

employs formalist network analysis to complement insights derived from the previously undertaken relational network analyses.

8.2.2 Network Exchange Patterns in Online Communities

Prior research on informal online communities based on group attachment offered a foundation for inferences about the potential structure of a forum core's networks. The literature review of this subsection discusses a potential forum core structure, and frames the research question posed at the end of the section. Hypotheses about the forum core structure were drawn from the research of a particular case of informal online identity-based communities represented by the electronic networks of practice (ENP). Since the forum core may evolve similarly to informal online identity-based communities, this subsection reviews the essence of ENP and its social orientation. Prior research presented altruism as a dominant behaviour in an ENP and different types of reciprocity as its building blocks. The subsection also reviews the types of reciprocity and correspondent network forms previously identified in informal identity-based communities and MOOC networks for all posters. The network models discussed in the subsequent section include the network patterns justified within this subsection.

Electronic Networks of Practice (ENP). Prior research has investigated a particular instance of identity-based communities, so-called ENPs. An ENP represents "a self-organizing, open activity system that focuses on a shared interest or practice and exists primarily through computer-mediated communication" (Faraj, Wasko, & Johnson, 2008, p. 270). Wasko and Faraj (2005) conceptualised such networks as sustained by voluntary participation and knowledge exchange. Participants of these informal online communities share knowledge about particular interests and have been observed in Flickr photography groups, among Wikipedia contributors, and open source software developers. Research has shown that a stronger identification with the ENP is positively associated with trust in the group as a whole (Wasko, Teigland, & Faraj, 2009; W. Zhang, 2007). Such features of informal identity-based communities resonate with the assumed nature of a forum core, i.e. driven by knowledge exchange, underpinned by group attachment.

In addition to identity-based attachment and knowledge exchange around a shared practice, informal identity-based communities have been associated with altruistic behaviour. That is, in online communities based on shared purpose or interest, people

seek information by relying on the "kindness of strangers" (Constant, Sproull, & Kiesler, 1996). For example, Constant et al. (1996) found that the employees of a global computer manufacturer participating in an ENP gave useful advice and solutions to the problems of others despite the lack of personal connection. Some 81% of the providers did not know the seekers, and 10% were barely acquainted. Hence, replies did not directly benefit those who offered solutions. Altruistic motivation explained such contributions were offered for the sake of the community and public good, without an expectation of a reply in return. Faraj and Johnson (2011) suggested that the information exchanges within an ENP have a social, i.e. group-oriented, motivation. The authors contended that interactions among the participants were more than just information queries, and had a social purpose. Their contributions were socially-oriented towards a broader community benefit. Simply put, individuals participating in an ENP were aware that the larger group would benefit from their reply to an individual information-seeker. They provided the information in ways accessible to others reading it later. The underlying altruistic motives were also observed in a range of contexts demonstrating a group-oriented attitude among dominant motivations of ENP members. Examples can be found in online question and answer forums (Nam et al., 2009; Oh, 2012; P. Wu & Korfiatis, 2013; Yang et al., 2011), in communities of software developers (Baytiyeh & Pfaffman, 2010; Hars & Ou, 2001; C. Wu et al., 2007), and within Wikipedia editing practices (Kuznetsov, 2006; Prasarnphanich & Wagner, 2009). Given the prominence of group-oriented motivation in identity-based communities, hypothetically, a forum core could also be described by the presence of altruistic behaviour.

Reciprocity: Network Configurations and Interpretation. The altruistic 'pay-it-forward' behaviour of providing a public good without an expected return is one of the underlying forms of reciprocity. Molm et al. (2007) distinguished between unilateral and bilateral, as well as direct and indirect forms of reciprocity. The reciprocity types differ as to when an action is expected to be reciprocated, if at all. For instance, when both parties negotiate the terms of a bilateral exchange as in "if you scratch my back, I'll scratch yours", such reciprocity is considered "negotiated". According to the collective action theory, such an exchange indicates a non-altruistic motivation of self-interest (Ekeh, 1974). When the direct reciprocation of one's action could be delayed, or not offered at all, Molm (2003) described such a form of exchange as unilateral.

Unilateral reciprocity implies that although the possibility that the other person may reciprocate exists, there is a level of uncertainty as to when or if this may occur. Indirect reciprocity, a type of unilateral reciprocity also referred to as generalised exchange (Molm, Peterson, & Takahashi, 1999; Takahashi, 2000), is represented by the network form where individual "A may give to B on one occasion and to C on a different occasion". Indirect reciprocity can also be represented by the following patterns "A gives to B, B gives to C" as well as "A gives to B, B gives to C, and C gives to A" (Molm et al., 2007, p. 208). Examples of such forms of collective exchange can be found, for instance, in voluntary blood-giving activity or academic peer review practices.

The network configurations associated with indirect reciprocity are more than just structural representations. The forms of reciprocity carry meaning about the socio-emotional quality of the exchange they represent. One such quality is the underpinning trust that characterises the forms of unilateral reciprocity. In a unilateral exchange when a service is reciprocated directly or indirectly, the actor provides material or symbolic goods without knowing "whether, when, or to what extent the other will reciprocate" (Molm, 2003, p. 35). Since an individual is unsure about the return of the service, an increased possibility of risk is another characteristic of a unilateral exchange (direct and indirect reciprocation). Due to such embedded risk, such reciprocal forms are also associated with a higher emotional value. By reciprocating without an expectation of return, individuals demonstrate their readiness to pay forward a service and show a willingness to continue the relationship. According to Molm (2003), the forms of a unilateral exchange (direct and indirect reciprocity) take place in settings where conflict is of low salience. That is, if conflict is minimal, individuals are more likely to offer services unilaterally. Molm et al. (2007) further validated that 'paying it forward' to other group members, represented by the indirect reciprocity forms, occurred in contexts with higher levels of social solidarity, than those described by direct or negotiated reciprocity (p.236).

Network Patterns of Reciprocity in ENP and MOOCs. Prior research identified the presence of reciprocity forms in network structures of ENPs and all-poster MOOC networks. Both direct and indirect reciprocity configurations were found pertinent to identity-based communities. In modelling the ENP, Wasko et al. (2009) described their structure through the dominant features of direct reciprocity over indirect reciprocity.

Similarly, the features of direct reciprocity were found prominent in describing the structure of all-poster MOOC networks (Joksimović, Manataki, Gašević, Dawson, Kovanović, & De Kereki, 2016; Kellogg, Booth, & Oliver, 2014; J. Zhang, Skryabin, & Song, 2016). Besides network forms associated with direct and indirect reciprocity, Joksimović et al. (2015) demonstrated that triadic closure (Figure 8.1c) can in part reflect the structure of an all-poster MOOC networks. Specifically, the authors explained the structure of an all-poster's MOOC network through a more prominent direct reciprocity network configuration, as well as less prominent simmelian ties network configuration. In a MOOC forum context, a simmelian tie network form can be associated with the mutual exchange of interactions at the triadic level, and is characterised by the following structure: A → B, B → A, B → C, C → B, C → A, A → C. In this chapter, such form of triadic closure will be referred to as the reciprocal exchange at the triad level. Such a form of reciprocal closure is one of the local structures reflecting the clique-like clustering in a network. Finally, besides the similarities in their reciprocity patterns, research demonstrated that both ENP and all-poster MOOC networks were not described by a preferential attachment pattern (i.e. network form where A sends the tie to B, because B receives many ties from other actors) (Faraj & Johnson, 2011; Joksimović, Manataki, Gašević, Dawson, Kovanović, & de Kereki, 2016; Kellogg, Booth, & Oliver, 2014b; P. F. Wu & Korfiatis, 2013). To conclude, based on the prior insights, forum core structures, similarly to ENP, could hypothetically be comprised of such as features of direct and indirect reciprocity, including reciprocal exchange at the triad level.

By modelling forum core network structures through the features of direct and indirect reciprocity as observed in informal online identity-based communities, the chapter addresses the following research question:

RQ: To what extent are the participatory patterns observed in a forum core's networks typical of identity-based communities?

8.3 Methods

To address the research question, the networks were constructed under the assumption that forum replies were offered to a group, rather than to individuals (explained in Section 9.3.1). Upon this, the forum core network data in four MOOCs were modelled using ERGM. Four final models that converged and did not show signs of degeneracy

are presented in this chapter. This section elaborates on: (1) the network construction; (2) ERGM methodology and modelled network configurations; and (3) explanations of the measures of node-level activity that control for forum participation levels in the models.

8.3.1 Network Construction

For the analyses in this study, forum core networks were constructed differently than in the previous chapters. Network ties were non-valued and directed as representative of the direction of the service offered. Furthermore, networks were constructed to represent the collective orientation of communication in identity-based communities; that is, posting contributions were considered as offered to the group, rather than to individuals. Consequently, network ties were directed to all the posters in a single thread who preceded the actor based on timestamps. If A posted, B replied, and C commented, and D replied, each of the subsequent actors would have a directed tie to everybody else prior to them. That is, B→A, C→B, C→A, D→C, D→B, D→A. Such approach to network construction has not been widely adopted as MOOC analyses (specifically of Coursera networks) have used person-to-person post-reply structure (see Section 2.4.3). Yet, for all poster MOOC forum networks in edX, Brown et al. (2015) constructed learner networks with a similar, one-to-many structure, though without a conceptual justification for it. In this study, representing interactions through one-to-many ties was based on the principle of collective reciprocity where an individual contributed information to the posting group, rather than provided a service to an identified individual.

Network boundaries were set around the course delivery times: between the week the first video lecture was released, and the week when the last lecture in the course was released. As communication between actors can potentially continue well past a course's lectures due to exams or assignments, any established closing date for the forum analysis would have been arbitrary. The discrete temporal limits applied in this study made the comparison of courses more feasible.

8.3.2 Exponential Random Graph Modelling

Exponential random graph modelling (ERGM) was used to analyse the network properties describing the structure of forum cores' networks. ERGM is a probability

model (p*) for network analysis (Lusher, Koskinen, & Robins, 2012; Robins, Pattison, Kalish, & Lusher, 2007) that predicts the presence of the ties and local structures in networks. In ERGMs, network structure is understood as emergent from (1) endogenous structural parameters associated with social processes, such as, for example, preferential attachment or reciprocity, and (2) exogeneous parameters, such as individual actor attributes or the impact of interpersonal relations in another network. ERGMs assume that individual characteristics influence actors in sending to or receiving ties from others (Robins, Elliott, & Pattison, 2001). The role of these parameters in the network formation is theoretically hypothesised, and tested within the ERGM to estimate the probability of each parameter to form ties in the analysed network.

The logic of the ERGM can be explained as follows. The ERGM estimates the probability of a network G, given the sum of network statistics weighted, just as in regression, by parameters inside an exponential (Lusher, 2013. p. 10). The network statistics used to reproduce network structure represent social processes expressed in sets of two actors (dyads), three actors (triads) or larger combinations, such as several triads sharing a tie (social circuits). An ERG model estimates the likelihood that each selected configuration, associated with a theoretically formulated social process, will occur beyond chance in a random network generated based on the modelled constraints. Such probability is derived from generating a distribution of networks similar to the one that is examined, and estimating if any of the modelled configurations occurs in their structure beyond chance. ERGM applies different classes of statistical models to conduct such estimations (e.g., Frank & Strauss, 1986; Holland & Leinhardt, 1981; Robins, Snijders, Wang, Handcock, & Pattison, 2007; Snijders, Pattison, Robins, & Handcock, 2006; Wasserman & Pattison, 1996). Different parameters associated with distinct and even conflicting processes can be modelled simultaneously (Contractor, Wasserman & Faust, 2006; Lusher, et al., 2013) since an ERGM controls for them against one another, due to their dependency within the network structure.

Estimation of whether new models produce a better fit are done in reference to the null model. A baseline model in ERGM examines only the network density to identify that the network does not occur at random. Modelling beyond the null ERGM includes subsequent addition of network configurations representing endogenous and exogenous parameters hypothesised to form the network. If the modelled network converges, the model is checked for the goodness of fit to examine how well the modelled data

represents the analysed data. The model fit is represented by Akaike information criterion (AIC, Akaike, 1976), as well as the comparisons of modelled network statistics (degree distribution, distance, etc.) versus the examined ones (Harris, 2013; Lusher et al., 2012). In case of the acceptable goodness of fit, the ERGM estimations are considered to reflect the structure of the analysed network. The modelled output included a parameter estimate where zero indicates that the modelled effect was no different from random. These log-odds of a parameter can also be transformed into the odds of ties occurring.

ERGMs are appropriate for the analysis envisioned in this chapter if compared to other statistical network models. Stochastic Actor Oriented Models (SAOMs) and Relational Event Models (REM) are alternative network models that could capture the network structure through statistical modelling. However, both these methodologies incorporate temporal analysis of network formation. Although temporal element of how the networks of learners develop are of utmost relevance, this study first intends to model the overall structure of the network as it is observed by the end of the course. Being able to compare the different network structures, as formed through varying social processes allows for further hypothesising of how these networks may develop. In other words, ERGMs capture and explain group-level collective learning outcomes by the end of the course. Further analysis may use extrapolations derived from ERGMs to model and understand temporal dimension of how these structures have evolved.

8.3.3 Modelled Configurations and Effects

To model forum core networks, structural configurations representing social processes and individual level of activity were modelled through exponential random graph modelling. The modelled configurations are depicted in Figure 8.1.

Structural Configurations. The study examined the propensity of three structural configurations to form the structure of forum core's networks, listed along with the correspondent statnet label (Handcock, Hunter, Butts, Goodreau, & Morris, 2008; Hunter, Goodreau, & Handcock, 2013):

- direct reciprocity characterised by mutual *statnet* statistic (Figure 8.1a);
- indirect reciprocity characterised by the cyclicalties *statnet* statistic (Figure 8.1b);

- reciprocal exchange at the triad level characterised by the simmelianties *statnet* statistic (Figure 8.1c).

The simmelianties statistic representing the mutual exchange at the triadic level (Figure 8.1c) controlled for the network closure. Although social circuit configurations capturing network closure (gwesp and gwidegree) were fitted on some of the networks, they were not characteristic of all the networks, and thus were not used to maintain comparability in the modelled outputs. Table 8.1 presents the counts of modelled network configurations for each of the networks.

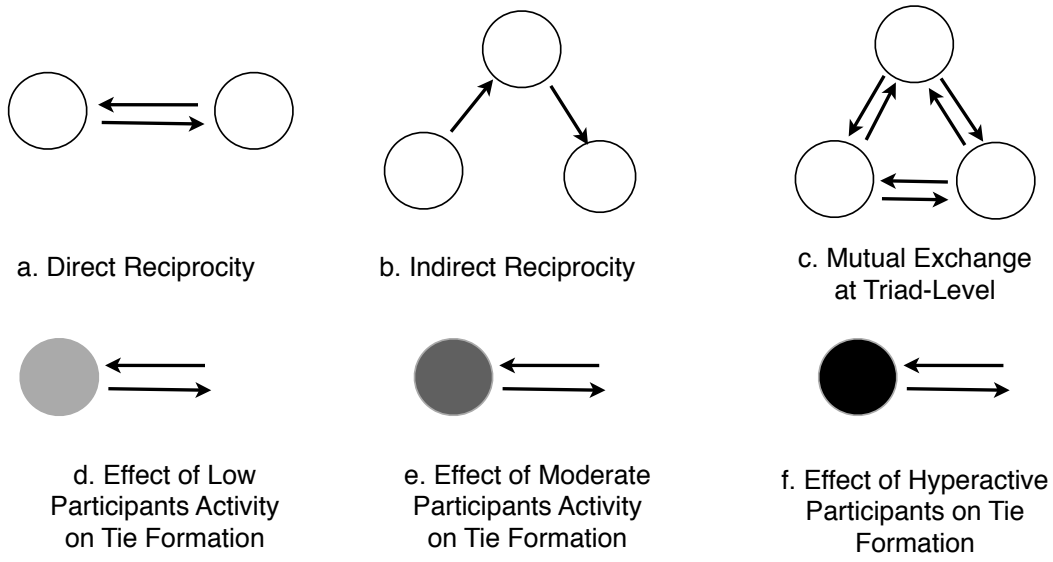


Figure 8.1 Visualisation of the configurations for the ERGM

Table 8.1 Descriptive statistics of structural network configurations

Network Statistic	**Course Name**			
	Water	**Solar**	**FP**	**Excel**
Nodes	340	254	161	220
Edges, Total	4574	3059	2978	2259
Direct Reciprocity, N (Figure 8.1a)	725	629	880	516
Indirect Reciprocity, N (Figure 8.1b)	3442	2626	1734	2084
Mutual Reciprocity at Triad-Level, N (Figure 8.1c)	1164	1060	2963	882

Node Attributes. Besides controlling for the structural parameters of reciprocity, the models were also controlled for the effect of learner superposting activity on the propensity to send and receive ties. The distinct levels of activity across online networks have been previously supported in numerous empirical studies. In particular, studies of Internet communities captured the very small group of answering-people who actively take on a large mass of overall activity (Turner, Smith, Fisher, & Welser, 2005). In MOOCs, several studies have shown the presence of hyper-posting activity (Huang, Dasgupta, Ghosh, Manning, & Sanders, 2014; Skrypnyk, Joksimović, Kovanović, Gašević, & Dawson, 2014). For example, Skrypnyk et al. (2014) demonstrated that hyper-posters in cMOOCs emerged both early and late in courses, and their activity approached the same level as that of the designated teacher. Huang et al. (2014) examined the health of forums at a group level in 44 Coursera MOOCs in relation to the superposting behaviours of its learners. The authors found that an overall higher activity of superposters is positively correlated with a higher activity from other forum users, concluding that superposters do not "drown out the silent majority" (p. 118). Papadopolous et al. (2014) also emphasised the role of community teaching assistants recruited from the active student body in the maintenance of forum spirit. As these learners made voluminous forum contributions they served the in-between role of bridging students with the staff, and can influence forum perceptions in a positive manner. Poquet and Dawson (2015) further demonstrated that teacher assistants and staff with higher activity levels occupied brokerage positions, in part securing the flow of information between learners. For this reason, modelling a forum core needed to account for a very active participation level of a select few, and a relatively low level for the overall majority. Figure 8.1 (d, e, f) demonstrates this configuration, though it should be noted in the ERGM no distinction made if the learner sent or received ties. The nodefactor statnet ERGM-term was used for modelling the effect of participation level.

To identify the overall level of activity, this study used measures of learner activity (1) as devised by Hecking, Hoppe and Harrer (2015) rather than relying on the number of messages posted. Hecking and colleagues (Hecking et al., 2015; Boroujeni, Hecking, Hoppe, & Dillenbourg, 2017) established in-reach and out-reach metrics of individual activity occurring in a MOOC network. These measures (below) calculate the diversity of outgoing and ingoing relations for a node i, where w(ei,j) represent the weight of the

tie between them, od (i,j) and id (i,j) represent out-degree and in-degree respectively. As the authors explained, the measures are useful for detecting differences in the activity of users that are connected to by many actors in a network. The measure has been applied in the MOOC context to capture in-reach and out-reach help-seeking and help-giving behaviour.

$$outreach(i) = od(i) - \sum_{j \in outneigh(i)} w(e_{i,j}) * log\left(\frac{w(e_{i,j})}{od(i)}\right)$$
$$inreach(i) = id(i) - \sum_{j \in inneigh(i)} w(e_{i,j}) * log\left(\frac{w(e_{i,j})}{id(i)}\right) \quad (1)$$

The present study replicated this measure in order to identify the in- and out-reach of an individual's activity across all-posters in MOOC forums. By using k-means clustering of in-reach and out-reach measures per person, all forum posters (see Chapter 5) were divided into three groups: to represent highest, moderate and low forum participation activity in the entire network. These attributes of posting activity were used to control for tie propensity formation in forum core ERGMs. Table 8.2 presents the descriptive statistics of actors within each cluster (counts), along with their average message posting activity. Cluster 1 refers to learners with low posting activity. Cluster 2 refers to the posters with moderate posting activity, and Cluster 3 represents hyper-posters.

As presented in Table 8.2, the labelling of activity levels is relative to the course, and heterogeneous in measures across the courses. That is, a hyper-poster in one course may have similar activity numerically as a moderate poster in another course. For example (Table 8.2), the Water course had the lowest level of moderate and hyperposting activity (40 posts per person and 218 posts per person respectively) and FP had rather high level of posting activity for low level and moderate level of participation within its network (20 posts per person, and 125 posts per person respectively). However, on average, the proportion of activity across the three clusters within the courses appeared to be relatively comparable. Appendix F presents the measures for cluster representatives.

Goodness of Fit. Several steps ensuring the goodness of fit were conducted, as emphasised in technical literature (Goodreau, Handcock, Hunter, Butts, & Morris, 2008; Harris, 2013; Lusher et al., 2012; Morris, Handcock, & Hunter, 2008). That is, converged networks with the best fit were checked for degeneracy. For this study's analysis, R packages *statnet* and *ergm* has been employed (Handcock et al., 2015; Handcock, Hunter, Butts, Goodreau, & Morris, 2008). The statnet package embeds a

Monte Carlo Markov Chain (MCMC) algorithm that stochastically approximates the Maximum Likelihood Estimation (MLE). These algorithms were therefore used to check for network degeneracy and to examine the goodness of fit. Estimated models demonstrated reasonable goodness of fit: (1) as the mean number of networked configurations was similar to configurations in the observed network; (2) the MLE estimation produced non-degenerate networks; and (3) the AIC coefficient demonstrated better fit that the null model (Table 8.3). Goodness of fit plots are included in the appendices (Appendix G).

Table 8.2 Descriptive statistics of the three clusters representing forum core posting activity levels in forums

Course	Cluster Characteristics	Poster Clusters (based on the in-reach and out-reach of their posting activity)		
		Cluster 1 (Low Participation)	Cluster 2 (Moderate Participation)	Cluster 3 (High Participation)
Water	Posters, N			
	• All Poster Network	2998	273	6
	• Forum Core	238	96	6
	Posts per Poster, *Mean (SD)*	15 (9)	40 (27)	218 (45)
Solar	Posters, N			
	• All Poster Network	4082	40	7
	• Forum Core	221	28	5
	Posts per Poster, *Mean (SD)*	18 (14)	129 (95)	586 (238)
Excel	Posters, N			
	• All Poster Network	1994	38	3
	• Forum Core	201	16	3
	Posts per Poster, *Mean (SD)*	13 (12)	115 (108)	534 (286)
FP	Posters, N			
	• All Poster Network	1317	41	4
	• Forum Core	130	27	4
	Posts per Poster, *Mean (SD)*	20 (17)	125 (178)	459 (188)

8.4 Results

This study examined if a forum core network structure resembled the structure of informal online communities based on group attachment and common interest. Forum core networks were investigated using the ERGM methodology. The models included the baseline density of networks, direct reciprocity of ties, indirect reciprocity of ties, and reciprocal exchange at the triadic level. The models also controlled for the

propensity of the formation of ties sent and/or received by the posters with different levels of network in-reach and out-reach, reflective of their overall posting activity. The levels were distinguished between individuals with low, moderate and high forum participation behaviour. All fitted ERGMs have converged, were not degenerate, and showed acceptable goodness of fit. All models were fitted using the described network configurations.

ERGM Outputs. The estimates for the network parameters are outlined in Table 8.3. There was a significant and positive effect for direct reciprocity in all four forum core networks, namely [Water: 1.27(.009), Solar: 1.55(.12), Excel: 2(.15), FP: 1.02(.28)], with estimates reported for each course and the standard errors in parentheses. There was a significant and positive effect for indirect reciprocity in all four forum core networks [Water: 0.2 (.03), Solar: .7(.003), Excel: .81(.006), FP: 1.54(.24)]. There was also a positive and significant effect for reciprocal exchange at triad-level [Water: 0.53 (.004), Solar: .54(.05), Excel: .22(.05), FP: .83(.13)]. All the estimates in four models were very significant, with a p value of <.001.

The results confirmed that the structure of forum core networks can be considered analogous to those found in online public electronic groups. The forum cores' networks in the three analysed courses were more defined by a higher propensity for direct dyadic exchange [Water: 1.27(.009), Solar: 1.55(.12), Excel: 2(.15)] in contrast to a lower propensity over tie formation of indirect exchanges [Water: 0.2 (.03), Solar: .7(.003), Excel: .81(.006)] or triadic-level exchanges [Water: 0.53 (.004), Solar: .54(.05), Excel: .22(.05)]. In sum, this reflects the expected network structure observed in informal online communities and all-poster MOOC networks (Joksimović, Manataki, Gašević, Dawson, Kovanović, & de Kereki, 2016; Faraj & Johnson, 2011; Kellogg et al., 2014; Wasko et al., 2009). Such findings generally support the presence of hypothesised social processes of reciprocity in a forum core. In general terms, a lower indirect reciprocity indicates a lower social solidarity, and potentially a lower level of maturity for an identity-based community within a forum core.

In the FP course, however, such dynamics were not replicated. The FP course had a higher propensity of indirect exchanges [1.54(.24)] to form ties than a direct reciprocity configuration [1.02 (.28)]. Also, FP's estimate for mutual exchanges at a triadic-level was higher than in other courses [Water: 0.53 (.004), Solar: .54(.05), Excel: .22(.05),

FP: .83(.13)]. FP's higher propensity for network closure can be interpreted as a more evolved network development and FP's higher level of indirect reciprocity can be interpreted as a network form of social solidarity. The propensity for these two levels of reciprocity to describe the forum core's structure in the FP MOOC suggests that this social group had evolved into a more developed identity-based community.

Table 8.3 Outputs for final ERG models in four edX MOOCs

	Course Name			
	Water	**Solar**	**Excel**	**FP**
Forum Moderation	None	Moderate (with CTAs)	High (TAs and staff)	High (TAs and staff)
Nodes, N	340	254	220	161
Edges, N	4574	3059	2260	2978
1. Structural Effects				
Modelled configuration (*statnet*)	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
Density (*edges*)	−4.34*** (0.03)	−4.38*** (0.05)	−4.69*** (0.07)	−4.76*** (0.24)
Reciprocity (*mutual*)	1.27*** (0.09)	1.55*** (0.12)	2*** (0.15)	1.02*** (0.28)
Indirect Reciprocity (*cyclicalties*)	0.2*** (0.03)	0.7*** (0.03)	0.81*** (0.06)	1.54*** (0.24)
Reciprocal Exchange at Triad-Level (*simmelianties*)	0.53*** (0.04)	0.54*** (0.05)	0.22*** (0.05)	0.83*** (0.13)
2. Effects of Node Participation Level (low level activity as intercept)				
	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
Effect of Moderate Forum Participation (*nodefactor*)	0.77*** (0.02)	0.68*** (0.02)	1.17*** (0.04)	0.73*** (0.02)
Effect of Hyper-posting Forum Participation (*nodefactor*)	1.5*** (0.05)	0.87*** (0.04)	1.7*** (0.07)	1.17*** (0.05)
AIC Null	38485	24601	18244	18450
AIC Final	32875	20262	13690	14308
BIC Final	32933	20317	13743	14357

Note: Each course network was modelled separately, only the final models are presented. *** stands for *p* value of <.001. *SE* – standard error, CTA – community teaching assistants, TA – teaching assistants, AIC – Akaike information criterion, BIC - Bayesian information criterion.

Interpretations of the ERGM Outputs. The results suggest relationships between the network configurations of direct reciprocity (interpersonal reciprocation), indirect reciprocity (social solidarity), triadic-level reciprocity (shifts to cliques and amplifications of information exchanges) and higher levels of hyper-posting activity. To demonstrate these relationships, Figure 8.2 illustrates the odds, i.e. probability of network configurations and participation effects on tie formation across the four MOOCs.

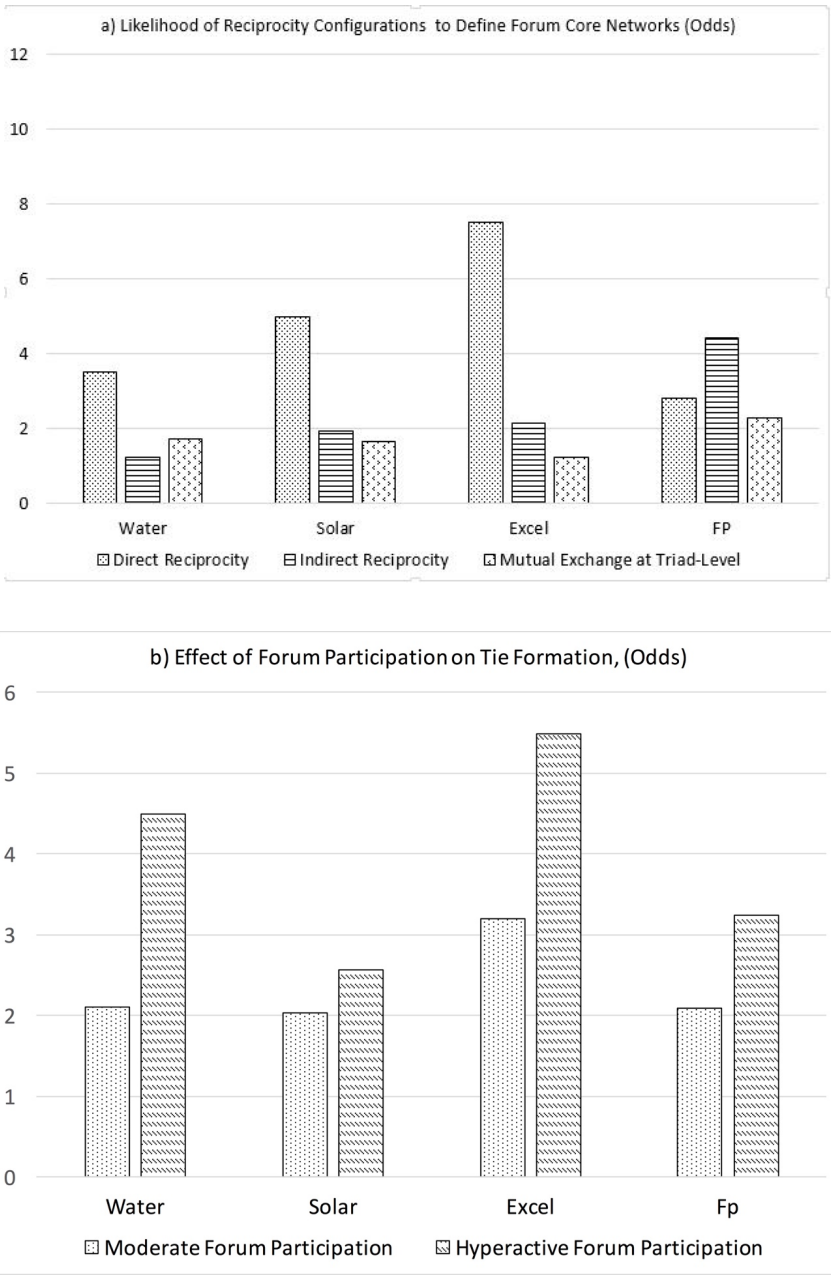


Figure 8.2 Cross-course comparison of the social processes driving the formation of a forum core.

First, the propensity for forum core networks to be based on direct and indirect reciprocity appears to be negatively associated. Such dynamics are best demonstrated in the two courses with opposing strategies for forum moderation, namely the FP course that had an intensive moderation strategy and the Water MOOC that had no moderation. To explain, the FP MOOC was a highly moderated course with the strongest presence of social processes (Chapter 6). In the ERGM outputs, FP was less characterised by direct reciprocity, while indirect reciprocity appeared as its most defining network configuration. A high indicator of indirect reciprocity is associated with social solidarity (Molm et al., 2007). On the other end of the spectrum, the unmoderated Water course, with fewer signs of social processes and no social negotiations central to its forum core, demonstrated high direct reciprocity and the lowest indirect reciprocity. Such network dynamics have been replicated in informal online communities such as ENPs (Wasko et al., 2009). The relationship between indirect and direct reciprocity is also negative in two moderated courses, Solar and Excel – much more complex to interpret, this is addressed below.

Similar to the relationship between direct and indirect reciprocity, the propensity of forum core networks to form based on direct and triadic-level reciprocity appears negatively related. Triadic-level reciprocity represents network closure, amplification and the shift of reciprocal relationships to a group-level. In a networked context, the prominence of such features signal interconnectedness between participants. Interconnectedness is a desirable outcome from an educational point of view, as a network shifts from being centred on a teacher or knowledgeable others. Such an outcome is also desirable as embeddedness in a network of interactions has also been associated with a sense of belonging and commitment. Hence, a higher propensity of generalised exchanges demonstrates a generally higher relational quality of network interactions. A negative relationship between triadic-level reciprocity and direct reciprocity was most prominent in the Excel course. Chapter 6 analyses also singled out the Excel MOOC as having few non-curriculum discussions and no centrality of social negotiation in its forum core. Together, this suggests a lack of reciprocal closure that would otherwise enable a shift to the triadic, i.e. group-level processes. Instead, a very high level of dyadic reciprocity defines a focused information exchange without a higher quality of relatedness. In Excel this can be interpreted as the result of a particularly high impact of moderators' activity, as depicted in Figure 8.2b (see further

discussion in this section). It may appear that the propensity of triadic-level reciprocity was somewhat counter-intuitive in the Water course – as it was relatively higher than that in moderated courses. This result is difficult to interpret as the Water course was both unmoderated, with few non-curriculum discussions and a lack of social negotiation conversations in the forum core. Therefore, triadic-level closure in this forum core is surprising, as its propensity is higher than that of some moderated networks. The most plausible conclusion is to treat these results with caution. The Water MOOC had the lowest indicators of posting activity (Table 8.2) and the lowest co-occurrence between its forum core members (Chapter 5, Table 5.3). In Water, an average forum core poster co-occurred twice, and in FP, this was 6-9 times. In other words, while triadic-closure may represent some clustering in a network, it's less plausible that it should be interpreted with a certain relational quality due to the differences in the intensity of reciprocation.

Data analysis also revealed a positive relationship between indirect reciprocity and triadic-level reciprocity. The higher indicators of these two features propelling a forum core's formation took place in courses where the social context for the becoming an identity-based community was most evolved. The FP course had both a high frequency of interpersonal co-occurrences (Chapter 5) and higher levels of non-curriculum discussion threads (Chapter 6), some of which were central to its forum core's formation (Chapter 6). Accordingly, FP can be described as the course with highest level of social context development among the four analysed MOOCs. Insights from the previous chapters align with the suggested interpretation of the forms of indirect reciprocity and reciprocal closure. Molm et al. (2003) proposed three conditions that describe environments with indirect reciprocity. First, the risk of vulnerability is heightened, that is individuals are vulnerable in offering a service that holds no guarantee of reciprocation. Second, the expressive value of giving is higher, as the act of 'paying forward' demonstrates a commitment to the shared practice by contributing without expecting anything back. Third, indirect reciprocity takes place when a conflict is less salient, that is the social tensions do not prevent contributors to 'pay forward'. Reciprocal closure, or reciprocation at the triadic-level is more representative of group-like formations that are taking over a network. An amplification of information flow, less control over information within a selected few and a higher integration of participants are aspects associated with such a structure. In sum, the combination of

higher indirect reciprocity and triadic reciprocity could be interpreted as social solidarity and group cohesion.

Analyses also demonstrate that there may be a positive relationship between hyper-posting activity and direct reciprocity, as well as a negative relationship between hyper-posting activity and indirect reciprocity. For the former, the pattern of a high propensity of hyper-posters to send and receive ties, high direct exchanges and low triad-level exchanges is observed in three of the four examined courses (Water, Solar and Excel). For the latter, two courses (Water and Excel) with a higher direct reciprocation and higher hyper-posting impact on tie formation have low indirect reciprocation. Both relationships can be interpreted as the impact of hyper-posters impeding group formation within the networks, despite hyper-posters' effort to stimulate discussions, or help those seeking answers. Higher levels of indirect reciprocity would indicate that learners feel compelled to reciprocate a favour they may have received in the past. However, if hyper-posters take on the responsibility to actively take on questions, the overall group may not feel compelled to reciprocate to unknown strangers, or develop a common identity to learn and help others learn. In other words, hyper-posting individuals may be taking away from opportunities associated with "pay-it-forward" behaviour.

Yet, what seems to be a negative relationship between social solidarity and high hyper-posting impact does not necessarily indicate that moderation is detrimental for an evolving relational quality of interpersonal interactions. In line with prior research and practice of group formation, the role of a teacher, a moderator or a leader is critical at early stages of group formation. In initial course stages, moderation (e.g. Excel, Solar, FP) could stimulate direct reciprocity, as well as indirect reciprocity. The trend of a higher propensity to reciprocate is observed across all moderated courses that were examined in this study, as opposed to the unmoderated one (i.e. Water). However, while a critical mass of moderate posters may be willing to pay effort and actions forward (in their posting activity), they still require opportunities to do so. If a moderator continues addressing all possible queries posted in a MOOC forum, e.g. in the Excel course, at an incredibly high propensity, affordances for network closure are limited. In other words, as much as moderation is needed at the onset of a course, a moderator's activity may need to be less involved further along a course's timeline, in order to allow for reciprocal closure.

Summary of results. The overall patterns of reciprocity, i.e. direct, indirect and reciprocal closure, appear to describe different forum core networks, and replicate some of the baseline indicators found in informal online communities. Furthermore, a high social solidarity indicator, represented by indirect reciprocity configurations, and a high group cohesion indicator, represented by triadic-level closure, both define the FP moderated forum core, where social context appeared to have evolved the most. However, a course adopting a similar moderation strategy, i.e. Excel, did not exhibit the same indicators of social context. In contrast, the discrepancy was explained by a high impact of hyper-poster involvement in network formation. Analyses presented in the chapter do not offer sufficient evidence of the relationships between these varying indicators. However, they lead to some interesting hypotheses that should investigate the relationships between: (1) moderation strategies and indicators of indirect reciprocity; and (2) direct reciprocity and the balance of hyper-posting and moderate posting activity.

8.5 Discussion

This chapter presented an analysis of the network configurations describing the forum cores and underlying social processes in four MOOC courses. Previous chapters argued that a forum core matures towards an identity-based community and that such a social entity is a group based on attachment, characterised by a commitment to the group's purpose and shared practice. Analyses of Chapter 6 and 7 demonstrated that forum cores develop into informational exchange groups based on an attachment to a shared practice of learning together in a MOOC. Consequently, the relational quality of such interactions demonstrates the low affectivity of interpersonal relations and high group cohesion. Chapter 7 argued that group attachment occurred through the negotiation of situational norms and an identification with the enactments of social learning practices. This chapter continued empirical investigations into forum cores as identity-based communities. To capture the distinct contexts of forum cores across four MOOCs through measurable indicators, ERGM was applied. The networks' structures were modelled based on previous findings from informal online contexts and all-poster MOOC networks. More specifically, configurations representing direct and indirect reciprocity, reciprocal closure and the effect of poster activity were used to model the forum core networks.

The chapter hypothesised that baseline processes of identity-based communities in a non-formal educational setting would resemble those observed in informal online networks. Results supported the hypothesis that the nature of forum cores in MOOCs resembles informal online identity-based communities. That is, chosen indicators characteristic of informal online identity-based communities could capture the social processes underpinning a forum core's formation, and were capable of reproducing forum cores' networks. More specifically, the forum core's structure in three courses was described by somewhat similar patterns of direct, indirect and triadic-level reciprocal exchanges that took place. Across all courses, the dynamics of the networks was similar to ENPs but courses with a higher presence of social processes pertinent to identity-based communities seemed to be characterised by higher indirect reciprocity and reciprocal closure. These findings align with the network configurations found in informal online communities, and all-poster MOOC networks. All of these are defined by a high propensity for direct reciprocity and a low propensity for indirect reciprocity. As such, forum cores align with what Wasko et al. (2004) termed ENPs.

However, in one highly moderated network examined in this study – the FP MOOC – network amplification and indirect reciprocity features were more prominent than direct reciprocity. This pattern differed from the conventional identity-based communities observed in informal online environments. Given the course's dynamics, as investigated in prior chapters, and its moderation, these indicators were hypothesised as indicative of an identity-based community that was more evolved than the forum cores of the other courses examined that were not as moderated or were unmoderated. In other words, using the distinctions by Brown and Duguid (2001), the identity-based community has moved from what can be described as an ENP to what can be described as an electronic CoP. The indicators characterising this forum core are associated with social solidarity and group cohesion. The indicators of interpersonal co-occurrence (Chapter 5) also singled-out this course as offering the most dyadic co-occurrences. In sum, indicators of higher indirect reciprocity and network amplification along with lower direct reciprocity could be associated with a higher relational quality of group attachment.

Results suggest a number of relationships between the observed network configurations that are to be further investigated. That is, the ERGM outputs indicated: (1) a negative relationship between direct and triadic-level reciprocity; (2) a positive relationship between indirect reciprocity and triadic-level reciprocity; (3) a negative relationship

between a high participation effect and the extent of triadic-level reciprocity; and (4) a negative relationship between indirect reciprocity and a high participation effect. An estimation of the strength of these relationships was not feasible, given the scope of the analyses.

This study offered additional insights into the roles, played by the MOOCs' staff, teaching and community assistants. MOOC research claimed their importance in supporting the spirit of forums. This chapter further demonstrated that moderated forums appear to transition from their baseline properties, as their networks' structure shift from distributed to more hierarchical. The results of the present study also supported Huang et al.'s (2014) conclusion that superposters do not drown-out the silent majority. More specifically, the level of moderate posting and reciprocity configurations are still high in courses with pronounced hyper-posting activity. Yet, this study demonstrated that the health of a forum may still be affected. In courses where hyper-posting activity was more distinct than that of moderate posters', features of network amplification and indirect reciprocity were not as pronounced as expected. Such dynamics may indicate that hyper-posters are limiting engagement opportunities for learners who are moderately active. In other words, if there is a moderator who always replies, and no teacher to stimulate dialogue, then a community has less of a reason to engage amongst themselves beyond their direct information needs. It does seem that hyper-posters, who are staff and teaching assistants at early stages, help networks develop, as their presence results in higher participation levels in general. However, it appears that they may not be always skilful at deciphering the optimal moments and strategies for them to diminish their own posting activity or sustain conversations between posters, rather than closing conversations.

A direct reciprocity of knowledge exchange seems to be inherent in identity-based communities, such as forum cores. As a forum core is based on information exchange, posters seek information, and receive information, possibly with a thank you or a clarification. The feature of direct reciprocity was prominent and present across all four examined courses. Over time and with facilitative efforts, the network structures came to be defined by their core and periphery: with reciprocation and subsequent network closure based on a reciprocation of services, likely at the core. This clustering may also be interpreted as clique formation, or polarisation of power and access that takes place as a network shifts from distributed to amplified. Facilitated forum core networks

appear to be characterised by a higher degree of clustering of reciprocated triads, and a lower level of random reciprocation. This could mean that random direct exchanges within a group decrease as information flow amplifies across more and more members, shifting from a cliquish core and distributed periphery to an amplified interconnected cluster.

This study's analyses yielded many promising results. However, much of the interpretations and their associated implications remain speculative. Future work should seek to account for the temporal aspect of how networks unfold and interact, including forum core interactions and comparisons with all-forum posters in MOOCs. An extrapolation of the findings relates to the time-based dynamics of network formation. Observations draw out connections between facilitation efforts within MOOCs and shifts in forum cores' network structure, from distributed to cliquish to egalitarian structures, where an egalitarian structure is the product of combined direct and indirect reciprocity. A forum core is in turn, situated within the network exchanges of intermittent posters, and the relationship of a forum core to those posters is based on the features of indirect reciprocity. More precisely, if a forum core member gets reciprocated by a staff member at an early stage, she would be more likely to offer answers to an intermittent poster later on in the same course. In other words, measuring the extent of network amplification over time offers partial explanations as to the nature of social exchanges, while indirect reciprocations between a forum core and other posters within the same network explains the remaining dynamics. These speculations form the basis for future research.

Several direct implications stem from this study's analyses. First, empirical validation of forum core networks as identity-based communities is valuable as MOOC forum research can build on established research agendas, both in the social sciences and in studies of human-computer interaction. For instance, Ren et al. (2012) developed a set of design principles for the facilitation of identity-based communities, and these could be applied in MOOCs. Yet, as the dataset of courses was limited, more diverse sets of courses need to be analysed to see if the insights demonstrate consistent patterns across disciplines, student cohorts and MOOC sizes.

While this study did not include text mining and an analysis of learners' posts, identity-based online communities have been researched using text analysis along with social

network analysis. In particular, Grabowiecz et al. (2013) suggested that measures of entropy capture the lack of topical diversity in identity-based communities, what they refer to as 'topical groups', and the lack of divergence of discussion themes in what they refer to as 'social groups'. Furthermore, granular text mining for the concepts related to power and authoritativeness (Howley, Mayfield, & Rosé, 2013; Howley, Mayfield, Rosé, & Strijbos, 2013) may offer insights as to where the shifts in network structure are reflected in the text features of the various learner posts. On a final note, the indicators proposed from this work should be validated in formal online education environments, so as to assess for comparability between formal and open online settings.

8.6 Conclusions

The study presented in this chapter concludes the research propositions put forward in the literature review of the thesis. The study validated a set of indicators that define the social processes underpinning a forum core's formation. Indicators of direct reciprocity represent dyadic information exchange, indirect reciprocity represents social solidarity, and reciprocal closure indicates amplification of a network and a shift to a more egalitarian structure from a centralised one. The study demonstrated that courses where a designated moderator was involved and more evolved social process were observed in the previous chapters exhibited higher indirect reciprocity than direct reciprocity in network formation. The formation of a forum core network was also mostly defined by an amplification of information flow reciprocally, at the triadic level. The study also revealed that hyper-posting activity may impede network shifts from centralised information flow to egalitarian.

9 Conclusion

This thesis has investigated the development of the social context in MOOCs in order to understand the differences that may exist between formal and informal online course settings. Despite the unique properties of open online educational settings, such as the scale of diversity, idiosyncratic learner intentions, and open participation, MOOCs have been largely examined as if they were formal online courses (see Chapter 2). The thesis has argued that the distinct features of MOOC environments impact the process of participation and affect how learning should be designed. As a result, the patterns of forum activity were examined to elucidate the social context emergent from forum participation, leading to insights into the dynamics of learning in networks. Thus, this final chapter brings together the analyses presented in the individual empirical studies (Chapters 4–8) that have explored MOOCs' networks so as to discuss the role of forum core posters in creating a social learning context in MOOCs.

As established in this thesis, distinct MOOC features may alter the development of the social context from that expected in formal online courses. A social context emergent from interpersonal interactions is at the foundation of contemporary learning theories (see Chapter 2). In formal online education, the social context is defined by the developing learning community which has been linked to deeper knowledge construction (e.g. Garrison & Akyol, 2013), enhanced motivation and emotional support (Martin & Dowson, 2009) and learner satisfaction (e.g. Jung, Choi, Lim, & Leem, 2002). An evolving relational quality of interpersonal interactions is also embedded within socio-constructivist and socio-cultural learning theories. However, a dearth of research has examined if interpersonal interactions in MOOCs develop relational quality as in formal education settings (see Chapter 2). The lack of opportunity for a learner to develop such a relational quality through social contexts in MOOCs may discredit the use of formal educational theories and practices in open online educational settings.

In light of this research gap, this study examined the social context emergent in platform-based MOOCs through the overarching research question:

RQ: What are the properties of social context development in MOOC settings, and what indicators can capture it?

To address this research question, the thesis examined the patterns and quality of group-level interpersonal interactions in four edX MOOC forums. The investigation was structured around the research propositions identified through the literature review (see Chapter 2). These research propositions inquired: 1) if interpersonal relationships could form between learners in MOOC settings; 2) whether these relationships may lead up to community development; and 3) what network structures could describe such a community. In addressing these propositions, Chapter 4 framed the thesis' analyses through a typology of forum use that included posters with varying participation levels. Chapter 5 analysed the potential for the formation of interpersonal relationships between these different MOOC forum posters. Chapter 6 examined the nature of exchanges between the forum posters who may be forming personal relationships. Chapter 7 in particular focused on interpreting the off-task communication through social-task, social non-task and mixed discussions occurring in the examined MOOC forums. The analysis was implemented through the lens of the CoP theory (Lave & Wenger, 1991; Wenger, 1998). Chapter 8 investigated specific structural forms for the networks of posters in the four MOOCs.

This chapter now briefly presents and synthesises the findings (Section 9.1), and specifies the theoretical and practical contributions made within the thesis (Section 9.2). Thereafter, the chapter notes the limitations of the research presented in the thesis, and outlines a futu**re research agenda**. A final note summarising the value of the contribution concludes the thesis.

9.1 Summary of the Findings

Exploratory data analysis (Chapter 4).

A working typology of forum users was adopted to frame the analyses of forum activity. Given the lack of existing typologies that are unique to forum use in MOOCs, the types of forum users were differentiated based on the regularity of their forum access and posting activity. An exploration of the data showed that regular posters generated the largest volume of clickstream activity in the forums. To consider the impact of the social environment on learner interactions, forum core posters were delineated within

the group of regular posters. Forum core posters were defined as learners who post in forums in any three or more weeks throughout their courses, and receive timely replies to their posts. For the investigation of the social context in MOOCs, the forum core, regular posters, and occasional posters were determined as the focus of the Doctoral thesis.

Study 1 (Chapter 5)

What types of forum participation are associated with opportunities for relationship formation?

The frequency of co-occurrences between the forum core, regular posters and occasional posters was examined to identify opportunities for relationship formation. The findings suggested that forum core posters are the strongest sub-group of forum posters associated with the potential for relationship formation. Such potential exists between forum core posters since these posters co-occur with one another most often. Their co-occurrence also has a temporal dimension, as it accumulates throughout a course's duration. To reveal how the forum core posters utilised their potential for relationship formation, the content of their interactions required examination. Since the average frequency of their co-occurrence was rather low, the relational quality of their interactions remained questionable.

Study 2 (Chapter 6)

What is the role of on-task and off-task communication in the formation of a forum core?

The content of forum core interactions was examined so as to gain insight into their relational quality. Results showed that both on-task and off-task communication played a role in forum core formation. Forum core posters predominantly engaged in the social activity (discussion forums) to address curriculum-related issues (74% of the co-occurrences). Furthermore, forum core engagement in curriculum-related discussions increased throughout the courses' progression. Despite the lower volume of social non-task communication, social non-task exchanges were also found to play a central role in a forum core's network formation (in network centrality) in two MOOCs (Solar and FP), suggesting that their forum cores engaged in social processes. Hence, in contrast to formal online groups, the development of a forum core is unlikely to be based on

interpersonal affectivity, although, while limited, group-oriented social processes were also present.

Study 3 (Chapter 7)

What were the forum core posters talking about within social-task, social non-task and mixed discussion threads?

Social-task, social non-task and mixed discussion threads where the forum core co-occurred were further interpreted through the lens of the CoP theory (Wenger, 1998). Thematic analysis of the discussions suggested that forum core posters in some courses engaged in collective negotiation around the shared practices of forum participation and situational norms of learning through MOOC forums. Participation in negotiating course assessment, forum rules, and course organisation was interpreted as the mechanism for self-identification with the emergent forum core identity and the development of group-level attachment. From this perspective, in contrast to learning groups in formal online education, a forum core can be perceived as evolving into an identity-based community based on attachment to their group's identity, without accompanying interpersonal familiarity and trust at the dyadic-level.

Study 4 (Chapter 8)

To what extent are the participatory patterns observed in a forum core's networks typical of identity-based communities?

The forum core was conceptualised as an identity-based community where group-level attachment emerges through negotiation of situational norms in MOOC forums. The structure of forum core networks was reproduced and quantified through the forms of direct and indirect reciprocity, as previously observed in informal online identity-based communities. The forum core nework structures in three MOOCs were found to resemble informal online identity-based communities, defined by high direct, low indirect and low triadic-level reciprocal exchanges. In contrast to unmoderated forum core networks, the structure of the moderated forum core of the FP MOOC differed, as defined by high indirect reciprocity and high triadic-level reciprocal exchanges. These structural differences were interpreted as indicative of more evolved social processes, accompanied by the amplification of information flow and more community-like structure.

9.2 The Contribution of the Thesis

The first research proposition formulated in this thesis set to examine affordances for the formation of interpersonal relationships in MOOCs, distinguishing between different levels of participation. Chapter 5 demonstrated that forum core posters are able to form relationships as they co-occur with one another the most often throughout a course's progression.

The second research proposition inquired if group-level processes, similar to that of community formation, could take place through the more frequent co-occurrences within a forum core. Analyses in Chapters 6 and 7 contend that communities can evolve in MOOCs. However, the mechanisms for their formation and the kind of emotional attachment that motivates further engagement are conceptually different to those theorised in the learning contexts of formal online education. To explain, the formation of communities based on dyadic affectivity and trust – an emotional person-to-person attachment underpinning community formation in formal online courses appeared unlikely. The majority of forum core exchanges were curriculum-driven. The low levels of on-task interpersonal communication containing self-disclosure further suggested the lack of interpersonal emotional underpinnings of forum cores' interactions. The analyses of on-task and off-task communication of a social nature, central to the forum core's formation in two examined MOOCs, despite generally low in volume, demonstrated that forum cores may undergo community formation processes. These social processes are likely to be based on group-level attachment rather than dyadic bond-based attachment. The results suggest that community formation mechanisms in the analysed MOOCs were enacted through shared group-level negotiations related to the situational norms around the forums and learning in MOOCs.

The third research proposition emphasised the need to theoretically ground the social network analysis of MOOC forums for meaningful insights about their network structure. To address such a proposition, forum core networks were represented and modelled as identity-based communities. Analyses found that the indicators of reciprocity and information exchange associated with the altruistic motivation to contribute can quantify the network structure of forum cores. The structure of forum cores resembled the network patterns of informal identity-based online communities in courses that were unmoderated and/or were less evolved as a social entity. However, the

structure of a moderated forum core with more pronounced community-formation processes differed.

Conceptualisation of the Social Context in MOOCs. In addition to addressing the three research propositions, this study contributes a novel conceptualisation of social context development in MOOC settings. The social context of a MOOC forum evolves through a multi-level structure with different dynamics and relational processes underpinning different parts of the network (Figure 9.1). The interactions between occasional posters are not underpinned by emotional connections, attachment or a shared history. These individual did not engage in relational processes (Figure 9.1; light grey nodes and dotted-line interactions). In contrast, a forum core's interactions may develop a relational quality throughout a MOOC. However, the mechanism of the evolution of a forum core into a community is not bond-based, as in formal classrooms. The level of socio-emotional discussions in a forum core is low, resulting in an overall low affectivity of the emergent context (see Chapter 6). Another important distinction is that the attachment between members is not based on interpersonal relations, but rather on self-identification with the group learners relate to. In the analysed courses, the source of identification was values espoused during the negotiation of situational norms and personal views of their respective MOOC and forum as a space for learning (Chapter 7, Figure 9.1, dark grey nodes with solid-line interactions).

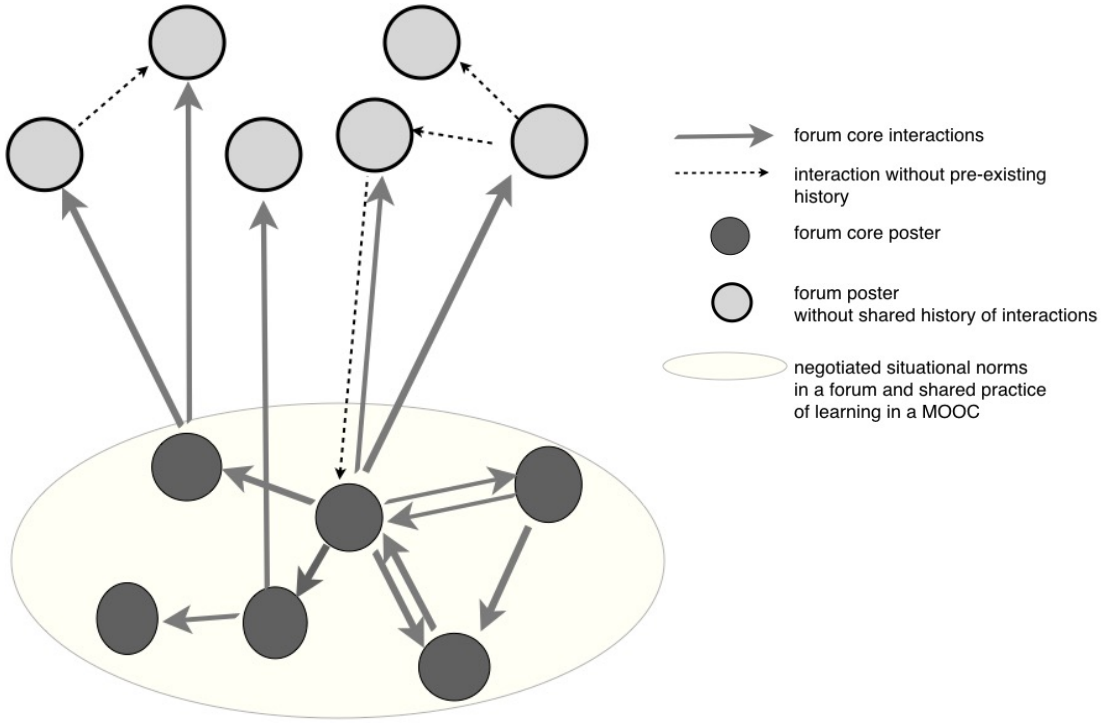


Figure 9.1 Multi-level relationships in a MOOC Forum.

The social context in MOOC forums is somewhat similar to that of informal online communities. The structure of the networked relations resembles online informal networks, with few people having a high number of connections (see Chapter 5). Similar to open online communities, most of the content is generated by a small group of people (see Chapter 4). The underlying network patterns of unmoderated forum cores resemble those of informal online communities (see Chapter 8). Similar to the voluntary behaviour observed in informal open online communities, attitudes to forum participation differ from lightweight, i.e. occasional posting, to heavyweight, i.e. negotiation of the shared space (see Chapters 5 and 7). Such a self-organised way of interacting in open online settings, based on learner autonomy, is similar to informal open online environments.

The social context in MOOCs also appears to share some similarities with the dynamics of formal online classrooms. Like groups in formal classrooms, forum core members accumulate interpersonal encounters throughout the course they undertake (see Chapter 5). These encounters are largely defined by the course's curriculum and tasks (see Chapter 6). Despite the different scale of communication, forum core learners experience growing self-disclosure as their course progresses, while the social norms are negotiated early on in the course (see Chapter 6) – reflective of the pattern in formal online courses. Fluctuations of the discussion threads shared by a forum core, as well as learner participation are influenced by course design and course deadlines (see Chapter 6). Finally, in highly moderated forums with a strong teacher presence, the network structure of a forum core shifts from that resembling informal online communities to a more amplified information flow, suggesting the emergence of a learning community (Chapter 8). To sum up, the formal educational set-up provides a structure within which learners self-organise towards the development of a community.

Indicators of Social Context. The findings from the thesis serve to contribute to a range of indicators that can effectively capture the social context in a MOOC forum. Table 9.1 provides a summary of the forum indicators applied in this thesis. Of critical relevance is the premise that these indicators should be applied to the conceptually different groups of forum users, representing different levels of social context (Figure 9.1). As indicated in Table 9.1, the number of dyadic co-occurrences (Chapter 5) between forum posters can indicate their potential for relationship formation. Another indicator reflecting the development of the social context is the centrality of different

types of discussions, in particular those containing social negotiations. The centrality of non-curriculum related discussions of a social nature to the forum core network signifies that learners have engaged in group-related processes (Chapter 6). The modelling of forum cores' structures as identity-based communities through the elements of direct and indirect reciprocity, including triadic-level reciprocity, reveals the extent of group cohesion development (Chapter 8). Modelling the network structure can also demonstrate the influence of certain forum participants on the network's formation.

Evaluations based on these indicators offer insights into the development of the social context in MOOC forums. In particular, analyses in the thesis demonstrate that the following serve as indicators of a MOOC forum (FP) where a social context is most evolved: 1) a high level of interpersonal co-occurrence; 2) a centrality of social negotiation threads to the forum core; and 3) prominence of network features of indirect reciprocity over direct reciprocity and high triadic-level closure.

Table 9.1 Indicators of the social context applied in the thesis

Indicator	Explanation	Interpretation
Interpersonal co-occurrence	Number of interpersonal co-occurrence	Potential for relationships
Centrality of socio-emotional threads	Network position of discussion threads in relation to the forum core	Uptake of the potential for relationships
Network structure	Direct reciprocity, indirect reciprocity, triadic reciprocity, effect of superposters	Reciprocation, social solidarity, group cohesion, facilitating cliques, impediments to network amplification

To explain, the FP MOOC presented the highest average number of interpersonal co-occurrences as compared to the other courses, indicating the highest potential for its forum core posters to establish interpersonal relationships (Table 9.2). In order of highest to lowest levels of co-occurrence were Solar, closely followed by Excel, and then Water, suggesting that these MOOC forums garnered varying stages of social context development. Other indicators listed in Table 9.2 supported such an interpretation, complementing it in insight. FP and Solar were the two courses where social-task and social non-task communication was found central to the forum core's

communication. Hence, the position of the discussions of a social nature in these two MOOCs revealed that their forum cores had taken-up relational opportunities to engage in negotiations on social matters.

Table 9.2 Overview of indicators in the analysed MOOCs

	FP	Solar	Excel	Water
Range of times for interpersonal co-occurrence (forum core)	[6.68, 9.09]	[3.32, 3.89]	[2.93, 3.52]	[2.52, 2.87]
Centrality of socio-emotional threads	Yes	Yes	No	No
Prominent network structures	Indirect reciprocity; triadic-level reciprocity	Direct reciprocity	Direct reciprocity; effect of superposter	Direct reciprocity, effect of superposter

The indicators of the network structure derived through modelling the forum cores as identity-based communities further captured the development of the social context in FP. The forum core network in this MOOC was characterised as having the highest propensity for group cohesion (triadic-level reciprocity) and social solidarity (indirect reciprocity), along with the lowest direct reciprocity as compared to the other analysed courses (Chapter 8). Solar had comparatively high indicators of group cohesion and indirect reciprocity but was lower than the direct reciprocity indicators that appear to have captured baseline activity in informal identity-based online communities (Chapter 8). The Water MOOC had a low frequency of forum core co-occurrence, had no social off-task communication threads central to many forum core posters. The course was also defined by prominent direct reciprocity. These indicators suggest that the social group-level processes underpinning the interactions in the Water MOOC were less evolved.

9.3 Implications

A novel conceptualisation of the MOOC forum as a multi-level entity with a forum core that evolves into an identity-based community has theoretical, methodological and practical implications.

9.3.1 Transfer of Theories and Models

Group-level attachment forging the evolution of the social context in MOOC settings represents a mechanism that conceptually differs from the person-to-person attachment that sits at the foundation of theories of online education. The overwhelming bulk of existing approaches to social context development in formal courses favours interpersonal relations as facilitative of the social context. In formal online education theory, constructs related to the social context are dependent on an interpersonal dyadic level of interaction and salience of the self. Despite references to the collective level, social context and identity within the overall conceptualisations of social presence in both Gunawardena and Zittle (1997) and Garrison (Arbaugh et al., 2008), measure dyadic-level attachment. More recently, Kreijns et al. (2014) revisited the social presence construct to argue that the original theorisation of social presence by Short et al. included two elements: the element of social presence as a dyadic-level construct that involves interpersonal impression and the element of the social space as a web of relationships. Kreijns et al. (2014) described the construct of the social space in a manner resembling an identity-based community; however, they theorised dyadic social presence as a mediating factor for social space. Given the lack of salience of the self in a MOOC setting and the group-level attachment as mechanism for community formation, these conceptualisations and the models of online education associated with them should be applied with caution.

Instead of replanting the analytical tools across education settings, educators and researchers need to explicitly address the assumptions about the social context embedded within the theories and models they are applying, as well as justify why the transfer of the educational tool is valid. For instance, the social capital of learners has been increasingly acknowledged as an important learning outcome from social interactions (Carceller, Dawson, & Lockyer, 2015; Gašević, Zouaq, & Janzen, 2013; Haythornthwaite & de Laat, 2012). Building a network where as many learners engage with as many learners as possible has been a pedagogical strategy addressing the learner's in social capital development (e.g. Skrypnyk, Joksimović, Kovanović, Gašević, & Dawson, 2014). At the end of the course, social capital has been quantified through one's number interlocutors in online forums (Carceller et al., 2015; Joksimović et al., 2015). Among adopted theoretical underpinning of such social capital is that of Lin (2008) who refers to the social capital as an individual-level outcome within a

network, as a perceived benefit from another learner. When, re-applying Lin's concept to a MOOC context, the evident concern is that, as the thesis suggested, the perceived benefit from another learner is mediated by group-level attachment. In other words, individuals may expect a return from another person mostly if trust into the learning community has evolved, unless the individual is the teacher or a superposter with a salient personality. If perceived benefit is mediated through an individual's attachment to the group, the number of interactions per se is insufficient to reflect one's social capital because the social capital is drawn from an identification associated with and developed from the group, rather than individual connections. Furthermore, if a teacher's intention is to help learners build social capital as conceptualised in Lin's understanding of the construct, MOOC forums are probably not the right kind of technology to facilitate this process. Expressed differently, although a learner can draw on a group's social capital throughout a course, communication will dissolve at the end of the course. Thus, helping learners build networks outside forums, through Twitter, Github, Facebook, or other, may offer a pedagogical strategy that can serve to build Lin's type of social capital, as any networks created are more sustainable.

This example allows for a demonstration of the complexity in the transfer of educational tools across contexts. As highlighted through the studies in the thesis, such decision-making by both researchers and educationalists requires additional caution.

9.3.2 Methodological Implications

The conceptualisation of the social context in MOOCs posed in the thesis carries methodological implications. Methodological justification of the boundaries imposed unto analysed entities in open educational settings is crucial for deeper and more granular analyses of learning. The hybrid nature of the social relationships underpinning interactions in MOOC forums should be captured through a network representation that differentiates between the ties defined by distinct motivations and relational quality. The actors sending these ties of distinct quality, should also be approached as acting in different modes, or belonging to different sub-groups. To explain, the network ties spanning interactions between individual MOOC posters are qualitatively different between forum core posters and other posters. Ties sent by occasional posters (Figure 9.1) do not have a quality of temporal familiarity between posters, while ties between forum core posters represent repeated co-occurrences accumulated throughout a given

course (Chapter 5). Hence, analysing these ties and these nodes from a network perspective needs to take place at different levels. Social network analysis of MOOC forums needs to adopt more elaborate analysis methods in order to represent the complexity of the emergent social context – multi-level network analysis (Lazega & Snijders, 2015) is just one example of this type of methodology.

9.3.2 Implications for Designing Technology and Evaluating Interventions

While pedagogical implications of how to facilitate identity-based communities still need further investigation, the findings hold much importance for the design and use of technology in MOOCs. First, by proposing that communities in MOOCs evolve through group-level attachment as identity-based communities, the thesis links a solid body of empirical research that discussed elements of design for informal online identity-based communities for the design of web interfaces in MOOCs (Ren et al., 2012; Ren, Kraut, & Kiesler, 2007). Secondly, the findings raise interesting questions about the use of recommender systems in MOOCs. Chapter 8's results suggest that indirect reciprocity features prompt the amplification of information flow in forum core networks, while prolonged superposting activity constrains it. Therefore, a question remains as to the integration of automated question and answer recommendation systems that help learners to navigate a forum, and if these would diminish the effect of community formation, much like an over-active moderator. Finally, as research has experimented with the integration of chatrooms that facilitate dyadic exchanges in MOOC forums, the results of this study suggest that to further evaluate the effects of these interventions, researchers may consider distinguishing between the impact on the dyadic perceptions and relationships and group-level sense of belonging.

9.3.4 Implications for Pedagogy

In regards to the application of the existing theories and models of teaching online, their effectiveness is yet to be empirically validated. The results of this thesis imply that forum facilitation models that rely only on the facilitation of interpersonal relationships may not have the intended impact on the MOOC context. Rather, the findings suggested that facilitation of the group culture or social space would be relevant to enhance social context in the open online educational setting

The thesis also links the social context emergent in the MOOC platforms to the pedagogical strategies used for teaching to large groups. Dron and Anderson (2014) distinguished four types of social entities operating at scale online: dyads, groups, nets, and sets. Dyads rely on interpersonal familiarity and trust; groups also depend on dyadic relationships that could alternatively be underpinned by the common unity (as an in emergent CoP). Nets represent loose connections underpinned by information exchange. Sets related to interest-based networks, that operate around a shared identity based on the topic. Dron and Anderson hypothesise that in open online settings these entities overlap and form the so-called collectives.

The social context emergent on the MOOC forum can be characterised by what Dron and Anderson (2014) referred to as a collective. Depending on the evolution of the social context, this collective could comprise the net represented by the occasional posters and a set represented by the forum core. It could be argued that if forum core transitions through negotiations to lower transactional distance, as hypothesised in the previous section, a set of posters could evolve towards a group. Dron and Anderson (2014) described pedagogical considerations relevant to facilitating these different social entities. These considerations could be used to guide pedagogical practice, as well as validate impact of design and pedagogy on the development of a particular social entity in an experimental manner.

9.4 Limitations

The findings of the study are promising and contribute to the emergent area of practice where research is limited. The claims drawn from the empirical analysis, however, do need to be further validated. The limitations of the study largely stem from the scope of the thesis, the data, and the fact that technology aiding social interactions in MOOCs is constantly evolving.

Scope as Limitation. First, findings reported and interpreted within this thesis carry limited generalisability as a result of **the scope of this study. Analyse**s were conducted only on four MOOC courses, and findings may not fully replicate across other MOOCs. Furthermore, the findings are limited to the MOOC courses run through the same platform and by the same institution. The courses were also within similar STEM disciplinary group. Although these similarities enabled cross-course comparison presented in the thesis, they challenge the generalisability of the insights, mostly across

the different MOOC platforms. Generalisability of the findings is also challenged by the constant evolution and change of MOOC platform technology. Since the findings in the thesis were derived from the basic edX forum arrangement, their generalisability to the courses is compromised within the forums where perhaps additional social learning plug-ins/ features have been applied.

In other words, theorisations postulated within this thesis require replication in the future work to strengthen or challenge the claims around reported findings.

Sub-population Focus as Limitation. The thesis has examined only one part of the MOOC network, the forum core, leaving out the other posters and types of forum participation outlined in Chapter 4. The thesis contends that occasional posters differ in the relational quality of their interactions. Regular posters as well as occasional posters were not described as a part of the evolving identity-based community of practice. The quality of their interactions remains unclear. The role these participants play in the formation of the core posters community also has not been addressed. It is plausible that regular posters and occasional posters that regularly view the forum, also develop some sense of belonging into the core learner community and become driven to contribute as the course progresses. However, the theory of the communities of practice is limited in explaining their role in the social context formation. Future research is necessary to fully understand how various learners using the forums contribute and experience the evolution of its social context.

Limited Focus on Learning Design. Another element that was not given a stronger focus within the analysis, due to the scope of the study, was the role pedagogical design plays on the social context. Although the findings suggested the role of assessment, and the course structure was accounted for in the analyses, the thesis did not undertake a deep analysis of the links between the intended learning design and forum activity beyond the stated course structure and official moderation strategies.

9.5 Future Research

Despite the limitations associated with the thesis findings, it offers a foundation for further research dialogue around the social context within open online environments. In particular, the limitations of the findings can be addressed and corroborated within future research. This section details the future research that would address the study's

limitations and advance the understanding of as well as the indicators for collective learning in open online education settings.

9.5.1 Relational Model of the Components of Social Learning in a MOOC Setting

To address issues of alignment of the learning outcomes with the process of learning and the presage stage of learning, as described in the opening of the thesis, the study proposes a relational model of the components of social learning in MOOC settings (Figure 9.2).

As previously noted (see Chapter 1), designing for learning involves three stages (presage, process and product) that need to reach equilibrium for teaching and learning to evolve effectively (Biggs, 1993, Biggs & Tang, 2011; Biggs & Telfer, 1987). Given the different motivations of MOOC learners, their self-reported learning intentions may be either directly associated with the learning gains, or mediated by the social activity, if the learners engage in it. Other learning outcomes may include acquired competence and identity, or social capital. Both these learning outcomes can only be available to the learners who engaged in the social activity in a MOOC.

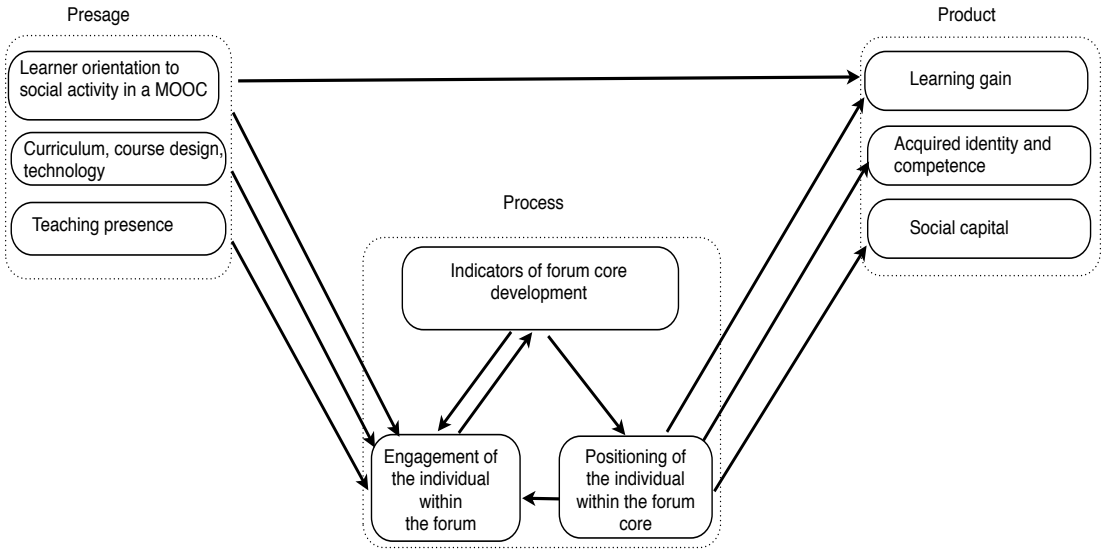


Figure 9.2 Relational model of the components of social learning in a MOOC setting

In relation to the process of learning, the model focuses on the social aspects of learning (forum participation). Learning through interpersonal interactions in a MOOC can be conceptually represented by the relationships between individual engagement within the

forum, the formation of the community that impacts both the further individual engagement levels. Individual engagement could, for instance, be represented by the number of forum posts, as well as regularity of participation. Individual engagement overall, links to the evolving social context. Here, the collective learning outcomes mediate the gains associated with the individual positioning within the forum core, i.e. an evolving community. The indicators of the forum core development could include average dyadic co-occurrence of an individual, a variable indicating the centrality of relational discussion threads in the forum core network, or variables capturing the likelihood of reciprocity features to describe the formation of forum core (Table 9.1). These indicators of the level of development of the context should be taken into account when individual network positions are interpreted. The three boxes represent the indicators that capture the varying levels of individual learning mediated by the collective-level state of the social context.

Finally, the embeddedness of an individual within the forum core is linked to the different types of learning outcomes. First, learning gains, as the learning of forum users has been mediated by their social activity. Second, their positioning is linked to the outcome representing if they have acquired competence in social learning at scale, or a changed identity.

By conceptualising learning activity in MOOCs' forums through such a relational model, future research can investigate the relationship between the evolving social context and learning outcomes, as well as the impact of different learning design on either of the elements of the learning process. For example, future research can examine the relationship between one's cognitive engagement and embeddedness within the social context; or, the role of dyadic attachment and group-level attachment on the cognitive and socioe-emotional benefits obtained by MOOC learners.

9.5.2 Indicators of Identity-Based Communities

Another area of learning analytics that could strengthen the presented research is derived from the analysis of the forum posts. In particular, there are two areas where discourse analytics can make a significant advance to support or challenge the analyses presented in the thesis.

First, discourse and content analysis can offer indicators for the emergent identities. For instance, automated analysis of text can address forum's sentiment, entropy, style, collective orientation, etc. to bring additional highlights into how communities in open online courses develop shared identity. Alternatively, epistemic networks can represent shared group identity, and capture its change throughout the course. Secondly, discourse analysis has potential to define the boundaries of the emergent community. The thesis suggested that the membership in forum core is defined by participation in the negotiation of the situational norms. Such can be extrapolated as a network boundary that can delineate forum core, along with the behavioural indicators of participation, such as continuity and reciprocity and timeliness of replies. Approaching network boundaries as emergent from the discourse is an novel direction in social science research, and suggestion to delineate network boundaries from the discourse are both empirically-driven (Ackland & O'Neil, 2011), as well as theoretically-grounded (Levina & Arriaga, 2014).

9.5.3 Temporal Analysis of the Social Context Development

One of the areas that was not addressed in the thesis analysis related to the temporal development of the communities of learners in a MOOC forum. The temporal dimension of learner interactions was only present in the initial stage of analysis (Chapter 5) to highlight the differences in the temporal activity among different forum participation types. Yet, network analysis (Chapter 8) and qualitative discourse analysis (Chapter 7) theorise a number of claims around how the identity-based community of MOOC learners evolves with the course progression. Future research should further investigate the hypothesised dynamics of the evolving MOOC context.

Particularly, future research could address the evolution of group-level network structures through the application of stochastic actor oriented models (SAOMs) and relational event modelling (REMs). These methodologies can aid in illuminating whether the different reciprocity indicators evolve from direct to indirect to triadic-level reciprocity, as hypothesised in Chapter 7. Such analyses could also address the role of facilitators and instructors in forming the reciprocal social processes – a reason as to why these elements may be instrumental to the temporal network development.

In addition, discourse indicators of the conversations shared by the forum core (as well as other forum participants) if analysed over the course duration, can complement the insights drawn from the network-level temporal analysis.

Adopting such research foci in the future will help to explain social processes in open online environment fully integrating the role temporal dimension plays in the community development.

9.6 Final Note

In conclusion, the results of the thesis have significant implications for the research and practice of massive open online courses. The thesis contended that MOOCs can develop social context conducive to richer learning experiences, yet the development and form of the social context in MOOCs does differ from formal online education. These differences need to be addressed within the analytical frameworks, research methods, and adopted learning design to capitalise on the affordances that exist within open education settings.

Most importantly, the thesis contributes to the position that learning with others can and should be facilitated in the learning settings regardless of their scale. In doing so, the ambition of the research was to emphasise that learning is a holistic process, beyond studying the curriculum; and the formation of an identity and positioning oneself within informal learning communities is an integral and sometimes undermined part of the learning process.

References

Ackland, R. & O'Neil, M. (2011). Online collective identity: The case of the environmental movement. *Social Networks, 33*(3), 177–190. http://dx.doi.org/10.1016/j.socnet.2011.03.001

Adamic, L., Huberman, B., Barabási, A., Albert, R., Jeong, H., & Bianconi, G. (2000). Power-Law Distribution of the World Wide Web. Science, 287(5461), 2115. http://dx.doi.org/10.1126/science.287.5461.2115a

Akaike, H. (1976). An information criterion (AIC). Math Sci, 14(153), 5–9.

Akyol, Z., & Garrison, D. R. (2008). The development of a community of inquiry over time in an online course: Understanding the progression and integration of social, cognitive and teaching presence. Journal of Asynchronous Learning Networks, 12(3–4), 3–22.

Alario-Hoyos, C., Pérez-Sanagustín, M., Cormier, D., & Delgado-Kloos, C. (2014). Proposal for a conceptual framework for educators to describe and design MOOCs. Journal of Universal Computer Science, 20(1), 6–23

Allen, I.E., & Seaman, J. (2003). Sizing the opportunity. The quality and extent of online education in the United States, 2002 and 2003. Babson Park, MA: The Sloan Consortium.

Allen, I. E., & Seaman, J. (2010). *Learning on demand: Online education in the United States, 2009*. Babson Park, MA: The Sloan Consortium

Allen, I. E., & Seaman, J. (2013). *Changing course: Ten years of tracking online education in the United States*. Babson Park, MA: The Sloan Consortium

Allen, I. E., & Seaman, J. (2014). *Grade change: Tracking online education in the United States, 2013*. Babson Park, MA: The Sloan Consortium

Allen, I. E., Seaman, J., & Garrett, R. (2007). *Blending in. The extent and promise of blended education in the United States*. Babson Park, MA: The Sloan Consortium

Allen, I. E., Seaman, J., Poulin, R., & Straut, T. T. (2016). *Online report card: Tracking online education in the United States*. Babson Park, MA: The Sloan Consortium

Allport, F. H. (1924). The group fallacy in relation to social science. *American Journal of Sociology, 29*(6), 688–706. http://dx.doi.org/10.1086/213647

Andersen, R., & Ponti, M. (2014). Participatory pedagogy in an open educational course: challenges and opportunities. *Distance Education*, *35*(2), 234–249. https://dx.doi.org/10.1080/01587919.2014.917703

Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2014). Engaging with massive online courses. *Proceedings of the 23rd International World Wide Web Conference (WWW 2014)*, 687–697. https://dx.doi.org/10.1145/2566486.2568042

Anderson, T. (2003). Getting the mix right again: An updated and theoretical rationale for interaction. *International Review of Research in Open and Distance Learning*, *4*(2), 126–141. http://dx.doi.org/10.19173/irrodl.v4i2.149

Anderson, T. (Ed.). (2011). *Theory and practice of online learning (2nd edition)*. Athabasca: Athabasca University Press.

Anderson, T., & Dron, J. (2012). Learning technology through three generations of technology enhanced distance education pedagogy. European Journal of Open, Distance and E-Learning, 15(2). Retrieved from http://www.eurodl.org/index.php?p=current&sp=full&article=523

Anderson, T. & Kanuka, H. (1999). Using constructivism in technology-mediated learning: Constructing order out the chaos in the literature. International Journal of Radical Pedagogy, 1(2). Retrieved from http://auspace.athabascau.ca/handle/2149/728

Anderson, T., Rourke, L., Garrison, D. R., & Archer, R. (2001). Assessing teaching presence in computer conferencing transcripts. Journal of the Asynchronous Learning Network, 5(2). Retrieved from http://auspace.athabascau.ca/handle/2149/725

Andrews, R., & Haythornthwaite, C. (2007). Introduction to E-learning Research. In R. Andrews & C. Haythornthwaite (Eds.), The SAGE handbook of e-learning research (pp. 1–53). UK: SAGE Publications.

Anh, J., Butler, B., & Alam, A. (2013). Learner participation and engagement in open online courses: Insights from the Peer 2 Peer University. MERLOT JOLT, 9(2). Retrieved from http://jolt.merlot.org/vol9no2/ahn_0613.htm

Arbaugh, J. B., Cleveland-Innes, M., Diaz, S. R., Garrison, D. R., Ice, P., Richardson, J. C., & Swan, K. P. (2008). Developing a community of inquiry instrument: Testing a measure of the Community of Inquiry framework using a multi-institutional sample. The Internet and Higher Education, 11(3–4), 133–136. https://dx.doi.org/10.1016/j.iheduc.2008.06.003

Archer, W., Garrison, D. R., Anderson, T., & Rourke, L. (2001). *A framework for analyzing critical thinking in computer conferences.* Paper presented at EURO-CSCL, Maastricht, the Netherlands.

Arguello, J., & Shaffer, K. (2015). Predicting speech acts in MOOC forum posts. In *Proceedings of the Ninth International AAAI Conference on Web and Social Media* (pp. 2–11), University of Oxford, Oxford, UK, May 26–29, 2015. Palo Alto, California: AAAI Press.

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd international conference on learning analytics and knowledge (LAK'12),* 267–270. http://dx.doi.org/ 10.1145/2330601.2330666

Atapattu, T., & Falkner, K. (2016). A framework for topic generation and labeling from MOOC discussions. In *Proceedings of the Third (2016) ACM Conference on*

Learning@ Scale (L@S'16), 201–204. http://dx.doi.org/10.1145/2876034.2893414

Baggaley, J. (2013). MOOC rampant. *Distance Education*, *34*(3), 368–378. http://dx.doi.org/10.1080/01587919.2013.835768

Baggaley, J. (2014a). MOOCS: digesting the facts. *Distance Education*, *35*(2), 159–163. https://dx.doi.org/10.1080/01587919.2014.919710

Baggaley, J. (2014b). Reflection: MOOC postscript. *Distance Education*, *35*(4),126–132. https://dx.doi.org/10.1080/01587919.2013.876142

Baker, R., Evans, B., Greenberg, E., & Dee, T. (2014). Understanding persistence in moocs (massive open online courses): Descriptive & experimental evidence. In U. Cress & C. Delgado-Kloos (Eds.). *Proceedings of the European MOOC Stakeholder Summit (eMOOCs'14)* (pp. 5–10). Retrieved from https://www.emoocs2014.eu/

Baker, R.S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J.A Larusson & B. White (Eds). *Learning analytics* (pp. 61–75). New York, US: Springer. http://dx.doi.org/10.1007/978-1-4614-3305-7_4

Bakharia, A. (2016). Towards Cross-domain MOOC Forum Post Classification. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale (L@S'16),* 253–256. http://dx.doi.org/10.1145/2876034.2893427

Bali, M. (2014). Going beyond the good MOOC/bad MOOC debate (An observation on the special issue). *Journal of Global Literacies, Technologies, and Emerging Pedagogies, 2(3),*261–266. Retrieved from http://www.academia.edu/download/34660656/11._Going_beyond_the_good_MOOC.pdf

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, *37*(2), 122–147.http://dx.doi.org/10.1037/0003-066X.37.2.122

Bates, A. (1988). Television, learning and distance education. *Journal of Educational Television*, *14*(3), 213–225. http://dx.doi.org/10.1080/0260741880140305

Bates, A. (2005). *Technology, e-learning and distance education*. London, UK: Routledge.

Bates, A. (2014). MOOCs: getting to know you better. *Distance Education*, *35*(2), 145–148. https://dx.doi.org/10.1080/01587919.2014.926803

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*(3), 497–529. https://dx.doi.org/10.1037/0033-2909.117.3.497

Bayne, S., & Ross, J. (2014). *The pedagogy of the Massive Open Online Course (MOOC): the UK view* (Report No. HEER000427). UK: Higher Education Academy. Retrieved from https://www.heacademy.ac.uk/resources/detail/elt/the_pedagogy_of_the_MOOC_UK_view

Baytiyeh, H., & Pfaffman, J. (2010). Open source software: A community of altruists. *Computers in Human Behavior*, *26*(6), 1345–1354. https://dx.doi.org/10.1016/j.chb.2010.04.008

Bearman, P. S., & Parigi, P. (2004). Cloning headless frogs and other important matters: Conversation topics and network structure. *Social Forces*, *83*(2), 535–557. http://dx.doi.org/10.1353/sof.2005.0001

Bergner, Y., Kerr, D., & Pritchard, D. E. (2015). Methodological challenges in the analysis of MOOC data for exploring the relationship between discussion forum views and learning outcomes. In O. Santos, J. Boticario, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, J.M. Luna, C. Mihaescu, P. Moreno, A.Hershkovitz, S.Ventura, and M. Desmarais (Eds.) Proceedings of the 8th International Conference on Educational Data Mining (EDM'15), Madrid, Spain (pp.234–241). Retrieved from http://www.educationaldatamining.org/EDM2015/uploads/papers/paper_61.pdf

Bernard, R., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A., & Bethel, E. C. (2009). A meta-analysis of three types of interaction treatments in distance education. *Review of Educational Research*, *79*(3), 1243–1289. https://dx.doi.org/10.3102/0034654309333844

Bernard, R., Abrami, P. C., Wade, A., Borokhovski, E., & Lou, Y. (2004). The effects of synchronous and asynchronous distance education: A meta-analytical assessment of Simonson's "Equivalency Theory". Presented at the Association for Educational Communications and Technology, Chicago, IL. Retrieved from http://eric.ed.gov/?id=ED485078

Biesta, G. (2009). Good education in an age of measurement: On the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability (Formerly: Journal of Personnel Evaluation in Education)*, *21*(1), 33–46. http://dx.doi.org/10.1007/s11092-008-9064-9

Biesta, G., & Burbules, N. C. (2003). *Pragmatism and educational research*. Oxford, UK: Rowman & Littlefield Lanham, MD.

Biggs, J. (1993). What do inventories of students' learning processes really measure? A theoretical review and clarification. *British Journal of Educational Psychology*, *63*(1), 3–19. http://dx.doi.org/10.1111/j.2044-8279.1993.tb01038.x

Biggs, J. and Tang, C. (2011). *Teaching for quality learning at university: What the student does (4th ed.)*. New York, USA: McGraw-Hill Education.

Biggs, J., & Telfer, R. (1987). *The process of learning*. McGraw-Hill/Appleton & Lange.

Blaikie, N. (2007). *Approaches to social enquiry: Advancing knowledge*. Malden, MA: Polity Press.

Blau, P. M. (1964). *Exchange and power in social life*. New Brunswick, USA/London, UK: Transaction Publishers.

Blom, J., Verna, H., Li, N., Skevi, A., & Dillenbourg, P. (2013). MOOCs are more social than you believe. *E-Learning Papers* [Online], 33. Retrieved from https://oerknowledgecloud.org/sites/oerknowledgecloud.org/files/Fro m-field_33_1.pdf

Borgatti, S. P., & Halgin, D. S. (2011). Analyzing affiliation networks. In J. Scott & P.J.Carrington (Eds.) *The SAGE handbook of social network analysis* (pp.417–433). SAGE Publications Ltd

Borokhovski, E., Tamim, R., Bernard, R., Abrami, P. C., & Sokolovskaya, A. (2012). Are contextual and designed student–student interaction treatments equally effective in distance education? *Distance Education*, *33*(3), 311–329. https://dx.doi.org/10.1080/01587919.2012.723162

Boroujeni M.S., Sharma K., Kidziński Ł., Lucignano L., Dillenbourg P. (2016) How to Quantify Student's Regularity?. In: Verbert K., Sharples M., Klobučar T. (Eds.) *Proceedings of the Conference on Adaptive and Adaptable Learning EC-TEL 2016. Lecture Notes in Computer Science* (vol. 9891). Springer, Cham. http://dx.doi.org/10.1007/978-3-319-45153-4_21

Brandes, U., Robins, G., McCranie, A., & Wasserman, S. (2013). What is network science? *Network Science*, *1*(1), 1–15. https://dx.doi.org/10.1017/nws.2013.2

Bransford, J., Brown, A., & Cocking, R. (2000). *How people learn*. Washington, DC: National Academy Press.

Boroujeni, M., Hecking, T., Hoppe, H. U., & Dillenbourg, P. (2017). Dynamics of MOOC discussion forums. To be presented at the *7th International Learning Analytics and Knowledge Conference (LAK17)*, March, 2017, Vancouver, Canada. Retrieved from https://infoscience.epfl.ch/record/223718

Breslow, L., & Pritchard, D. (2013). Studying Learning in the Worldwide Classroom - Research into edX's First MOOC. *Research & Practice in Assessment*, *8*(Summer 2013), 13–25.

Brinton, C. G., Chiang, M., Jain, S., Lam, H., Liu, Z., & Wong, F. M. F. (2013). Learning about social learning in MOOCs: From statistical analysis to generative model. *IEEE Transactions on Learning Technologies, 7*(4), 346–359. http://dx.doi.org/10.1109/TLT.2014.2337900

Brooks, C., Thompson, C., & Teasley, S. (2015). Who you are or what you do: Comparing the predictive power of demographics vs. activity patterns in Massive Open Online Courses (MOOCs). *Proceedings of the Second (2015) ACM Conference on Learning @ Scale (L@S'15)*, 245–248. https://dx.doi.org/10.1145/2724660.2728668

Brown, J., & Duguid, P. (2000). *The social life of information*. Boston, MA: Harvard Business Press.

Brown, J., & Duguid, P. (2001). Knowledge and organization: A social-practice perspective. *Organization Science*, *12*(2), 198–213. http://dx.doi.org/10.1287/orsc.12.2.198.10116

Brown, R., Lynch, C. F., Eagle, M., Albert, J., Barnes, T., Baker, R., … McNamara, D. (2015). Good communities and bad communities: Does membership affect performance? In O. Santos, J. Boticario, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, J.M. Luna, C. Mihaescu, P. Moreno, A.Hershkovitz, S.Ventura, and M. Desmarais (Eds.) *Proceedings of the 8th International Conference on Educational Data Mining (EDM'15*), Madrid, Spain (pp.612–613).

Brown, R., Lynch, C. F., Wang, Y., Eagle, M., Albert, J., Barnes, T., … McNamara, D. (2015). Communities of performance & communities of preference. *In CEUR Workshop Proceedings of the 8th International Conference on Educational Data Mining (EDM'15)*, Madrid, Spain.

Bruner, J. S. (1996). *The culture of education*. Cambridge, MA: Harvard University Press.

Budhathoki, N. R., & Haythornthwaite, C. (2012). Motivation for open collaboration: Crowd and community models and the case of OpenStreetMap. *American Behavioral Scientist*, *57*(5), 548–575. https://dx.doi.org/10.1177/0002764212469364

Bulfin, S., Pangrazio, L., & Selwyn, N. (2014). Making "MOOCs": The construction of a new digital higher education within news media discourse. *The International Review of Research in Open and Distributed Learning*, *15*(5). http://dx.doi.org/10.19173/irrodl.v15i5.1856

Burnett, K., Bonnici, L. J., Miksa, S. D., & Kim, J. (2007). Frequency, intensity and topicality in online learning: An exploration of the interaction dimensions that contribute to student satisfaction in online learning. *Journal of Education for Library and Information Science*, 21–35.

Calhoun, C. (2006). The university as the public good. *Thesis Eleven*, *84*(1), 7–43. https://dx.doi.org/10.1177/0725513606060516

Campbell, P. (2006). *Affect as a presence in the Community of Inquiry model* (Doctoral dissertation, Athabasca University, Athabasca). Retrieved from http://hdl.handle.net/2149/525

Cao, M. (2014). *Understanding learners' experience in MOOCs: a review of literature* (Master's thesis, The University of Texas, Austin, Austin). Retrieved from http://repositories.lib.utexas.edu/handle/2152/26325

Carceller, C., Dawson, S. P., & Lockyer, L. (2015). Social capital from online discussion forums: Differences between online and blended modes of delivery. *Australasian Journal of Educational Technology, 31*(2). http://dx.doi.org/10.14742/ajet.1448

Carlen, U., & Jobring, O. (2005). The rationale of online learning communities. *International Journal of Web Based Communities*, *1*(3), 272–295. http://dx.doi.org/10.1504/IJWBC.2005.006927

Caswell, T., Henson, S., Jensen, M., & Wiley, D. (2008). Open Content and Open Educational Resources: Enabling universal education. *The International Review*

of Research in Open and Distributed Learning, *9*(1). http://dx.doi.org/10.19173/irrodl.v9i1.469

Champaign, J., Colvin, K. F., Liu, A., Fredericks, C., Seaton, D., & Pritchard, D. E. (2014). Correlating Skill and Improvement in 2 MOOCs with a Student's Time on Tasks. *Proceedings of the First ACM Conference on Learning @ Scale Conference (L@S'15),* 11–20. https://dx.doi.org/10.1145/2556325.2566250

Cheung, E. (2014). *Analyzing student engagement and retention in Georgetown's first MOOC: Globalization's winners and losers: Challenges for developed and developing Countries*. (Master's thesis, Georgetown University, Washington, DC). Retrieved from http://hdl.handle.net/10822/709735

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*(1), 73–105. http://dx.doi.org/10.1111/j.1756-8765.2008.01005.x

Chong, S. (2016, August). *The millennial adult learner in higher education.* Paper presented at the European Educational Research Association Conference "Learning Education: The Distinct Contributions of Educational Research and Researchers" (ECER'16). Retrieved from http://www.eera-ecer.de/ecer-programmes/conference/21/contribution/37120/

Chung, I., & Ho, A. (2016). *HarvardX and MITx: Four years of open online courses - Fall 2012-Summer 2016*. (Report for the Research Committee for the Vice Provost for Advances in Learning at Harvard University and the Office of Digital Learning at MIT). http://dx.doi.org/10.2139/ssrn.2889436

Cleveland-Innes, M., & Campbell, P. (2012). Emotional presence, learning, and the online learning environment. *The International Review of Research in Open and Distributed Learning*, *13*(4), 269–292. http://dx.doi.org/10.19173/irrodl.v13i4.1234

Clow, D. (2013a). An overview of learning analytics. *Teaching in Higher Education*, 18(6), 683–695. https://doi.org/10.1080/13562517.2013.827653

Clow, D. (2013b). MOOCs and the funnel of participation. *Proceedings of the Third International Conference on Learning Analytics and Knowledge (LAK '13)*, 185–189. https://dx.doi.org/10.1145/2460296.2460332

Coetzee, D., Fox, A., Hearst, M. A., & Hartmann, B. (2014). Should your MOOC forum use a reputation system? *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing,* 1176–1187. https://doi.org/10.1145/2531602.2531657

Coetzee, D., & Hearst, M. A. (2014). Chatrooms in MOOCs: All talk and no action, *Proceedings of the First ACM Conference on Learning @ Scale Conference (L@S'14),* 127–136. http://dx.doi.org/10.1145/2556325.2566242

Coffrin, C., Corrin, L., de Barba, P., & Kennedy, G. (2014). Visualizing patterns of student engagement and performance in MOOCs. *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge (LAK'14)*, 83–92. https://dx.doi.org/10.1145/2567574.2567586

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://dx.doi.org/10.1177/001316446002000104

Comer, D., Baker, R., & Wang, Y. (2015). Negativity in Massive Online Open Courses: Impacts on learning and teaching and how instructional teams may be able to address it. *InSight: A Journal of Scholarly Teaching*, *10*, 92–113.

CEC (2000). Commission staff working paper. *A memorandum on lifelong learning.* Brussels: European Commission.

Connell, J. P., & Wellborn, J. G. (1991). Competence, autonomy, and relatedness: A motivational analysis of self-system processes. In M. Gunnar & A. Sroufe (Eds). *Self processes and development*, (pp. 43-77). Hills*d*ale, NJ, US: Lawrence Erlbaum Associates, Inc.

Connell, R. (2013). The neoliberal cascade and education: an essay on the market agenda and its consequences. *Critical Studies in Education*, *54*(2), 99–112. https://dx.doi.org/10.1080/17508487.2013.776990

Constant, D., Sproull, L., & Kiesler, S. (1996). The kindness of strangers: The usefulness of electronic weak ties for technical advice. *Organization Science*, *7*(2), 119–135. http://dx.doi.org/10.1287/orsc.7.2.119

Contractor, N. S., Wasserman, S., & Faust, K. (2006). Testing multitheoretical, multilevel hypotheses about organizational networks: An analytic framework and empirical example. *Academy of Management Review,* 31(3), 681-703. http://dx.doi.org/ 10.5465/AMR.2006.21318925

Creanor, L., & Walker, S. (2010). Interpreting Complexity: a case for the sociotechnical interaction framework as an analytical lens for learning technology research. In L. Dirckinck-Holmfeld, V. Hodgson, C. Jones, M. de Laat, D. McConnell, T. Ryberg (Eds.) *Proceedings of the 7th International Conference on Networked Learning* (pp.517–518), Aalborg, Denmark.

Creswell, J. (2009). *Research design: Qualitative, quantitative, mixed methods approaches*. SAGE Publications, Inc.

Creswell, J., & Clark, V. P. (2011). *Designing and conducting mixed methods research*. SAGE Publications, Inc.

Crossley, S., Danielle, S., Baker, R., Wang, Y., Paquette, L., Barnes, T., & Bergner, Y. (2015). Language to completion: Success in an educational data mining Massive Open Online Class. In O. Santos, J. Boticario, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, J.M. Luna, C. Mihaescu, P. Moreno, A.Hershkovitz, S.Ventura, and M. Desmarais (Eds*.) Proceedings of the 8th International Conference on Educational Data Mining (EDM'15)* (pp.388–391), Madrid, Spain.

Crosslin, M. & Dellinger, J. (2015). Lessons learned while esigning and Imiplementing a multiple pathways xMOOC + cMOOC. In D. Rutledge & D. Slykhuis (Eds.), *Proceedings of Society for Information Technology & Teacher Education*

International Conference 2015 (pp. 250-255). Chesapeake, VA: Association for the Advancement of Computing in Education (AACE).

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, *Complex Systems*, 1695.

Cui, Y., & Wise, A. (2015). Identifying content-related threads in MOOC discussion forums. In *Proceedings of the Second (2015) ACM Conference on Learning @ Scale*, 299–303. https://doi.org/10.1145/2724660.2728679

Czerniewicz, L; Deacon, A; Fife, M; Small, J; Walji, S (2015). *CILT position paper: MOOCs*. CILT, University of Cape Town. Retrieved from https://open.uct.ac.za/handle/11427/14041

Daniel, J. (2012). Making sense of MOOCs: Musings in a maze of myth, paradox and possibility. *Journal of Interactive Media in Education*, (3), Art. 18. http://dx./doi.org/10.5334/2012-18

Daniel, S. J. (2014). MOOCs – Evolution or revolution? Foreword to the special section on Massive Open Online Courses. *MERLOT Journal of Online Learning and Teaching, 10*(1). Retrieved from http://jolt.merlot.org/vol10no1/daniel_foreword_0314.pdf

Darabi, A., Liang, X., Suryavanshi, R., & Yurekli, H. (2013). Effectiveness of online discussion strategies: A meta-analysis. *American Journal of Distance Education*, *27*(4), 228–241. https://dx.doi.org/10.1080/08923647.2013.837651

Davis, A., Gardner, B. B., & Gardner, M. R. (1941). *Deep South: A social anthropological study of caste and class*. Columbia, South Carolina: The University of South Carolina Press.

Dawson, S. (2006). Online forum discussion interactions as an indicator of student community. *Australasian Journal of Educational Technology*, *22*(4), 495–510.

Dawson, S., Gašević, D., Siemens, G., & Joksimović, S. (2014). Current state and future trends: A citation network analysis of the Learning Analytics field. *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge (LAK'14)*, 231–240. http://dx.doi.org/10.1145/2567574.2567585

Dawson, S., Joksimović, S., Kovanović, V., Gašević, D., & Siemens, G. (2015). *Recognising learner autonomy: Lessons and reflections from a joint x/c MOOC.* Paper presented at the Higher Education Research and Development Society of Australasia (HERDSA), Melbourne, Australia.

Dawson, S., Macfadyen, L., & Lockyer, L. (2009). Learning or performance : predicting drivers of student motivation. In R. Atkinson & C. McBeath (Eds.), *Ascilite 2009: Same places, different spaces* (pp. 184–193). Auckland, NZ: Ascilite.

DBIS (Department for Business Innovation and Skills, Great Britain). (2013) *The Maturing of the MOOC: Literature Review of Massive Open Online Courses and Other Forms of Online Distance Learning*. (BIS Research paper 130). London, UK: Department for Business, Innovation and Skills.

de Laat, M. (2006). *Networked learning* (Doctoral dissertation, Utrecht University, the Netherlands). Retrieved from http://www.e-learning.nl/files/dissertatie maarten.pdf

de Laat, M., Lally, V., Lipponen, L., & Simons, R.-J. (2007a). Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for social network analysis. *International Journal of Computer-Supported Collaborative Learning*, *2*(1), 87–103. https://dx.doi.org/10.1007/s11412-007-9006-4

de Laat, M., Lally, V., Lipponen, L., & Simons, R.-J. (2007b). Online teaching in networked learning communities: A multi-method approach to studying the role of the teacher. *Instructional Science*, *35*(3), 257–286. https://dx.doi.org/10.1007/s11251-006-9007-0

De Vries, P., Hennis, T., Skrypnyk, A., (2015). *TU Delft Online Learning Research Working Paper #1* (DelftX MOOC Course Report ET3034Tux Solar Energy (2013)). Delft, the Netherlands: Delft University of Technology.

De Wever, B., Schellens, T., Valcke, M., & Van Keer, H. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers & Education*, *46*(1), 6–28. https://dx.doi.org/10.1016/j.compedu.2005.04.005

DeBoer, J., Ho, A. D., Stump, G. S., & Breslow, L. (2014). Changing "Course": Reconceptualizing educational variables for Massive Open Online Courses. *Educational Researcher*, *43*(2), 74–84. https://dx.doi.org/10.3102/0013189X14523038

Deboer, J., Stump, G. S., Pritchard, D. E., Seaton, D., & Breslow, L. (2013). Bringing student backgrounds online: MOOC user demographics, site usage, and online learning. In S. K. D'Mello, R.A. Calvo, and Olney, A. (Eds.) *Proceedings of the 6th International Conference on Educational Data Mining (EDM'13)*(pp.312–313), Memphis, Tennessee, USA. Retrieved from http://www.educationaldatamining.org/IEDMS/EDM2013

DeBoer, J., Stump, G. S., Seaton, D., & Breslow, L. (2013). Diversity in MOOC students' backgrounds and behaviors in relationship to performance in 6.002 x. In *Proceedings of the Sixth Learning International Networks Consortium Conference* (Vol. 4). Retrieved from http://tll.mit.edu/sites/default/files/library/LINC '13.pdf

Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process*. Boston: DC Heath and Company.

Dewey, J. (1958). *Experience and nature*. NY, USA: Dover Publications, Inc.

Dewey, J., & Bentley, A. (1949). *Knowing and the known*. Boston, MA: Beacon Press.

Diehl, W. C. (2011). *Learning at the back door. Charles A. Wedemeyer and the evolution of open and distance education* (Doctoral dissertation, The Pennsylvania State University, USA). Retrieved from https://etda.libraries.psu.edu/catalog/11795

Dillahunt, T. R., Wang, B. Z., & Teasley, S. (2014). Democratizing higher education: Exploring MOOC use among those who cannot afford a formal education. *The International Review of Research in Open and Distance Learning*, *15*(5). http://dx.doi.org/10.19173/irrodl.v15i5.1841

Dillenbourg, P., Baker, M., Blaye, A., & O'Malley, C. (1996). The evolution of research on collaborative learning. In E. Spada & P. Reiman (Eds.), *Learning in Humans and Machine: Towards an interdisciplinary learning science* (pp. 189-211). Oxford, England: Elsevier.

Dowell N.,Skrypnyk O., Joksimović S., Graesser, A, Dawson S., Gašević D. … Kovanović, V. (2015). Modeling learners' social centrality and performance through language and discourse. In O. Santos, J. Boticario, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, J.M. Luna, C. Mihaescu, P. Moreno, A.Hershkovitz, S.Ventura, and M. Desmarais (Eds.) *Proceedings of the 8th International Conference on Educational Data Mining (EDM'15)* (pp.250–257), Madrid, Spain. Retrieved from http://www.educationaldatamining.org/EDM2015/proceedings/full250-257.pdf

Dringus, L. P., & Ellis, T. (2010). Temporal transitions in participation flow in an asynchronous discussion forum. *Computers & Education*, *54*(2), 340–349. http://dx.doi.org/10.1016/j.compedu.2009.08.011

Dron, J. (2007). Designing the undesignable: Social software and control. *Educational Technology & Society*, *10*(3), 60–71. Retrieved from http://www.ifets.info/others/abstract.php?art_id=769

Dron, J., & Anderson, T. (2014). *Teaching Crowds. Learning and social media*. Athabasca, Canada: AU Press.

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development*. Lillington, NC: Edwards Brothers.

Ebben, M., & Murphy, J. S. (2014). Unpacking MOOC scholarly discourse: a review of nascent MOOC scholarship. *Learning, Media and Technology*, *39*(3), 328–345. http://dx.doi.org/10.1080/17439884.2013.878352

University of Edinburgh (2013). *MOOCs @ Edinburgh 2013 – Report # 1.* Edinburgh, Scotland: University of Edinburgh. Retrieved from https://www.era.lib.ed.ac.uk/bitstream/handle/1842/6683/Edinburgh_MOOCs_Report2013_no1.pdf?sequence=1&isAllowed=y

Ekeh, P. P. (1974). *Social exchange theory: The two traditions*. Cambridge, USA: Harvard University Press.

Emirbayer, M., & Goodwin, J. (1994). Network analysis, culture, and the problem of agency. *American Journal of Sociology*, *99*(6), 1411–1454. http://dx.doi.org/10.1086/230450

Engeström, Y. (2005). *Developmental work research: Expanding activity theory in practice* (Vol. 12). Berlin, Germany: Lehmanns Media.

Engle, D., Mankoff, C., & Carbrey, J. (2015). Coursera's introductory human physiology course: Factors that characterize successful completion of a MOOC. *The International Review of Research in Open and Distributed Learning*, *16*(2). http://dx.doi.org/10.19173/irrodl.v16i2.2010

Eraut, M. (2000). Non-formal learning, implicit learning and tacit knowledge. *British Journal of Educational Psychology, 70,* 113–136.

Erikson, E. (2013). Formalist and relationalist theory in social network analysis. *Sociological Theory*, *31*(3), 219–242. https://dx.doi.org/10.1177/0735275113501998

Ernst and Young. (2012). *University of the future: A thousand year old industry on the cusp of profound change* (Research report by Ernst and Young, Australia). Retrieved from http://www.ey.com/Publication/vwLUAssets/University_of_the_future/$FILE/University_of_the_future_2012.pdf

Eurostat. (2016, June). Lifelong learning statistics [Web page]. Retrieved from http://ec.europa.eu/eurostat/statistics-explained/index.php/Lifelong_learning_statistics

Eynon, R., Gillani, N., Hjorth, I., and Yasseri, T. (2014). *Conceptualising interaction and learning in MOOCs*. MOOC Research Initiative. Final Report. Retrieved in September, 2014, from http://www.moocresearch.com/wp-content/uploads/2014/06/C9146_EYNON_MOOC-Research-InitiativeEynon9146Final.pdf

Ezen-Can, A., Boyer, K. E., Kellogg, S., & Booth, S. (2015). Unsupervised modeling for understanding MOOC discussion forums: A learning analytics approach. *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge,* 146–150. https://dx.doi.org/10.1145/2723576.2723589

Fahy, P. J., Crawford, G., & Ally, M. (2001). Patterns of interaction in a computer conference transcript. *The International Review of Research in Open and Distributed Learning*, *2*(1). http://dx.doi.org/10.19173/irrodl.v2i1.36

Faraj, S., & Johnson, S. L. (2011). Network exchange patterns in online communities. *Organization Science*, *22*(6), 1464–1480. http://dx.doi.org/10.1287/orsc.1100.0600

Faraj, S., Wasko, M., and S. L. Johnson (2008). Electronic Knowledge Networks: Processes and structure. In I. Becerra-Fernandez and D. Leidner (eds.), *Knowledge management: An evolutionary view of the field* (pp. 270–291). Armonk, NY: M.E. Sharpe, Inc.

Ferguson, R., & Clow, D. (2015). Examining engagement: Analysing learner subpopulations in Massive Open Online Courses (MOOCs). *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 51–58). https://dx.doi.org/10.1145/2723576.2723606

Festinger, L., Back, K. W., & Schachter, S. (1950). *Social pressures in informal groups: A study of human factors in housing* (Vol. 3). Stanford, California: Stanford University Press.

Filvà, D. A., Guerrero, M. J. C., & Forment, M. A. (2014). The effects of massiveness on the participation in social technologies: a MOOC in secondary education. In *Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality* (pp. 397–402). http://dx.doi.org/10.1145/2669711.2669930

Frank, O., & Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, *81*(395), 832–842. http://dx.doi.org/10.1080/01621459.1986.10478342

Freedman, D. H. (2010, December). Why scientific studies are so often wrong: The streetlight effect. *Discover Magazine*, *26*. Retrieved from http://www.sjsu.edu/people/fred.prochaska/courses/ScWk240/s3/Freedman-Week-15.pdf

FutureLearn. (2016, November 15). FutureLearn Reached 5 million registered learners, turns 3 and runs the world's largest MOOC [Online Press Release]. Retrieved from https://about.futurelearn.com/press-releases/futurelearn-reaches-5-million-registered-learners-turns-3-runs-worlds-largest-mooc/

Garrison, D. R. (1985). Three generations of technological innovations in distance education. *Distance Education*, *6*(2), 235–241. http://dx.doi.org/10.1080/0158791850060208

Garrison, D. R., & Akyol, Z. (2013). The community of inquiry theoretical framework. In M.G. Moore (Ed.), *Handbook of Distance Education* (pp. 104–119). New York, NY: Routledge.

Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical Inquiry in a Text-Based Environment: Computer Conferencing in Higher Education. *The Internet and Higher Education*, *2*(2–3), 87–105. https://dx.doi.org/10.1016/S1096-7516(00)00016-6

Gašević, D., Dawson, S., Rogers, T., & Gašević, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, *28*, 68–84. https://dx.doi.org/10.1016/j.iheduc.2015.10.002

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64–71. http://dx.doi.org/10.1007/s11528-014-0822-x

Gašević, D., Kovanović, V., Joksimović, S., & Siemens, G. (2014). Where is research on Massive Open Online Courses headed? A data analysis of the MOOC Research Initiative. *The International Review of Research in Open and Distance Learning, 15*(5). http://dx.doi.org/10.19173/irrodl.v15i5.1954

Gašević, D., Zouaq, A., & Janzen, R. (2013). "Choose Your Classmates, Your GPA Is at Stake!": The Association of cross-class social ties and academic performance.

American Behavioral Scientist, *57*(10), 1460–1479. https://dx.doi.org/10.1177/0002764213479362

Gelo, O., Braakman, D., & Benetka, G. (2008). Quantitative and Qualitative Research: Beyond the Debate. *Integrative Psychological and Behavioral Science*, *42*, 266–290, http://dx.doi.org/10.1007/s12124-008-9078-3

Gillani, N., & Eynon, R. (2014). Communication patterns in massively open online courses. *The Internet and Higher Education*, *23*, 18–26. https://dx.doi.org/10.1016/j.iheduc.2014.05.004

Gillani, N., Eynon, R., Osborne, M., Hjorth, I., & Roberts, S. (2014). Communication communities in MOOCs. *arXiv preprint arXiv*:1403.4640.

Gillani, N., Yasseri, T., Eynon, R., & Hjorth, I. (2014). Structural limitations of learning in a crowd: communication vulnerability and information diffusion in MOOCs. *Scientific Reports*, *4*, Art.6447. http://dx.doi.org/10.1038/srep06447

Glass, C. R., Shiokawa-Baklan, M. S., & Saltarelli, A. J. (2016). Who takes MOOCs? *New Directions for Institutional Research*, *2015*(167), 41–55. https://dx.doi.org/10.1002/ir.20153

Goggins, S. P., Galyen, K., & Laffey, J. (2010). Network analysis of trace data for the support of group work: Activity patterns in a completely online course. *Proceedings of the 16th ACM International Conference on Supporting Group Work* (pp. 107–116). https://dx.doi.org/10.1145/1880071.1880089

Goldberg, L. R., Bell, E., King, C., O'Mara, C., McInerney, F., Robinson, A., & Vickers, J. (2015). Relationship between participants' level of education and engagement in their completion of the Understanding Dementia Massive Open Online Course. *BMC Medical Education*, *15*(1), 1–7. https://dx.doi.org/10.1186/s12909-015-0344-z

Goodreau, S. M., Handcock, M. S., Hunter, D. R., Butts, C. T., & Morris, M. (2008). A statnet tutorial. *Journal of Statistical Software*, *24*(9), 1–27.

Goodyear, P., Salmon, G., Spector, J. M., Steeples, C., & Tickner, S. (2001). Competences for online teaching: A special report. *Educational Technology Research and Development*, *49*(1), 65–72. https://dx.doi.org/10.1007/BF02504508

Grabowicz, P. A., Aiello, L. M., Eguíluz, V. M., & Jaimes, A. (2013). Distinguishing topical and social groups based on common identity and bond theory. *Proceedings of the sixth ACM international conference on Web search and data mining*, 627–636. http://dx.doi.org/10.1145/2433396.2433475

Grainger, B. (2013). *Massive Open Online Course (MOOC) Report 2013*. London, UK: University of London International Academy. Retrieved from http://www.londoninternational.ac.uk/sites/default/files/documents/mooc_report-2013.pdf

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 1360–1380. http://dx.doi.org/10.1086/225469

Greene, J. A., Oswald, C. A., & Pomerantz, J. (2015). Predictors of retention and achievement in a Massive Open Online Course. *American Educational Research Journal*. https://dx.doi.org/10.3102/0002831215584621

Greller, W., & Drachsler, H. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics. *Educational Technology & Society*, *15*(3), 42–57.

Grunspan, D. Z., Wiggins, B. L., & Goodreau, S. M. (2014). Understanding Classrooms through Social Network Analysis: A Primer for Social Network Analysis in Education Research. *CBE-Life Sciences Education*, *13*(2), 167–178. http://dx.doi/org/10.1187/cbe.13-08-0162

Gruzd, A. & Haythornthwaite, C. (2008). *Automated discovery and analysis of social networks from threaded discussions.* Paper presented at the International Network of Social Network Analysis, St. Pete Beach, FL, USA. Retrieved from http://hdl.handle.net/10150/105081

Gruzd, A., & Haythornthwaite, C. (2013). Enabling community through social media. *Journal of Medical Internet Research, 15*(10), e248. https://dx.doi.org/10.2196/jmir.2796

Gunawardena, C. N. (1995). Social presence theory and implications for interaction and collaborative learning in computer conferences. *International Journal of Educational Telecommunications*, *1*(2), 147–166.

Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1997). Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. *Journal of Educational Computing Research*, *17*(4), 397–431. https://dx.doi.org/10.2190/7MQV-X9UJ-C7Q3-NRAG

Gunawardena, C. N., & Zittle, F. J. (1997). Social presence as a predictor of satisfaction within a computer-mediated conferencing environment. *American Journal of Distance Education*, *11*(3), 8–26. http://dx.doi.org/10.1080/08923649709526970

Guo, P. J., & Reinecke, K. (2014). Demographic differences in how students navigate through MOOCs. *Proceedings of the first ACM conference on Learning @ scale conference (L@S' 14)*, 21 – 30. http://dx.doi.org/10.1145/2556325.2566247

Handcock, M., Hunter, D., Butts, C., Goodreau, S., Krivistky, P., & Morris, M. (2015). ergm: Fit, Simulate and Diagnose Exponential-Family Models for Networks (Version R package version 3.5.1). [Software, The Statnet Project, http://www.statnet.org]. Retrieved from http://CRAN.R-project.org/package=ergm

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software*, *24*(1), 1548.

Hanneman, Robert A. and Mark Riddle. 2005. Introduction to social network methods. Riverside, CA: University of California, Riverside.

Hansen, J. D., & Reich, J. (2015). Socioeconomic status and MOOC enrollment: Enriching demographic information with external datasets. *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge,* 59–63. https://dx.doi.org/10.1145/2723576.2723615

Hara, N., Bonk, C. J., & Angeli, C. (2000). Content analysis of online discussion in an applied educational psychology course. *Instructional Science*, *28*(2), 115–152. https://dx.doi.org/10.1023/A:1003764722829

Harris, J. K. (2013). *An introduction to exponential random graph modeling* (Quantitative applications in the social sciences series, vol. 173). Sage Publications.

Harrison, L. (2013). *Open UToronto MOOC Initiative: Report on first year of activity*. Retrieved from http://www.ocw.utoronto.ca/open-utoronto-mooc-initiative-year-one/

Hars, A., Shaosong O. (2001). Working for free? Motivations of participating in open source projects. *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*. http://dx.doi.org/10.1109/HICSS.2001.927045

Haythornthwaite, C. (1996). Social Network Analysis : An Approach and Technique for the Study of Information Exchange. *Library and Information Science Research, 18*(4). http://dx.doi.org/10.1016/S0740-8188(96)90003-1

Haythornthwaite, C. (1999). Collaborative work networks among distributed learners. *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. HICSS-32. Abstracts and CD-ROM of Full Papers,* Maui, HI, USA. https://dx.doi.org/10.1109/HICSS.1999.772707

Haythornthwaite, C. (2000). Online Personal Networks Size, Composition and Media Use among Distance Learners. *New Media & Society*, *2*(2), 195–226. https://dx.doi.org/10.1177/14614440022225779

Haythornthwaite, C. (2001). Exploring Multiplexity: Social Network Structures in a Computer-Supported Distance Learning Class. *The Information Society*, *17*(3), 211–226. https://dx.doi.org/10.1080/01972240152493065

Haythornthwaite, C. (2002). Strong, Weak, and Latent Ties and the Impact of New Media. *The Information Society*, (18), 385–401. http://dx.doi.org/10.1080/01972240290108195

Haythornthwaite, C. (2005). Social network methods and measures for examining e-learning. *Social Networks*, 1–22.

Haythornthwaite, C., & de Laat, M. (2010). Social networks and learning networks: Using social network perspectives to understand social learning. In L. Dirckinck-Holmfeld, V. Hodgson, C. Jones, M. de Laat, D. McConnell, T. Ryberg (Eds.), *Proceedings of the 7th International Conference on Networked Learning* (pp.517–518), Aalborg, Denmark. Retrieved from http://celstec.org.uk/system/files/file/conference_proceedings/NLC2010_Proceedings/abstracts/PDFs/Haythornwaite.pdf

Haythornthwaite, C., & de Laat, M. (2012). Social network informed design for learning with educational technology. In A. D. Olofsson, J. O. Lindberg, K. Klinger, & C. Shearer (Eds.), *Informed Design of Educational Technologies in Higher Education : Enhanced Learning and Teaching* (pp. 352–374). IGI Global.

Haythornthwaite, C., de Laat, M., & Dawson, S. (2013). Introduction to the special issue on learning analytics. *American Behavioral Scientist*, *57*(10), 1371–1379. https://dx.doi.org/10.1177/0002764213498850

Haythornthwaite, C., de Laat, M., & Schreurs, B. (2016). A social network analytic perspective on e-learning. In C. Haythornthwaite, R. Andrews, J. Fransman, E. Meyers (Eds.), *The SAGE handbook of e-learning research* (2nd ed.) (pp. 251–271). Sage Publications Ltd.

He, J., Bailey, J., Rubinstein, B. I., & Zhang, R. (2015). Identifying at-risk students in Massive Open Online Courses. *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI'15)* (pp. 1749–1755). Retrieved from http://dl.acm.org/citation.cfm?id=2886563

Hecking, T., Göhnert, T., Zeini, S., & Hoppe, U. (2013). Task and time aware community detection in dynamically evolving social networks. *Procedia Computer Science, 18*, 2066–2075. http://dx.doi.org/10.1016/j.procs.2013.05.376

Hecking, T., Chounta, I.-A., & Hoppe, H. U. (2016). Investigating social and semantic user roles in MOOC discussion forums. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK'16),* 198–207. http://dx.doi.org/10.1145/2883851.2883924

Hecking, T., Hoppe, H. U., & Harrer, A. (2015). Uncovering the structure of knowledge exchange in a MOOC discussion forum. *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015* (ASONAM'15), 1614–1615. http://dx.doi.org/10.1145/2808797.2809359

Henri F. (1992) Computer conferencing and content analysis. In: Kaye A.R. (Ed.), *Collaborative learning through computer conferencing* (NATO ASI Series, Series F: Computer and Systems Sciences, vol 90). Berlin, Heidelberg: Springer. http://dx.doi.org/10.1007/978-3-642-77684-7_8

Hesse, B. W., Werner, C. M., & Altman, I. (1988). Temporal aspects of computer-mediated communication. *Computers in Human Behavior*, *4*(2), 147–165. http://dx.doi.org/10.1016/0747-5632(88)90023-4

Hew, K. F., Cheung, W. S., & Ng, C. S. L. (2010). Student contribution in asynchronous online discussion: A review of the research and empirical exploration. *Instructional Science*, *38*(6), 571–606. http://dx.doi.org/10.1007/s11251-008-9087-0

Hill, P. (2013, March 2). The Four Student Archetypes Emerging in MOOCs. [Web log post]. Retrieved from http://mfeldstein.com/the-four-student-archetypes-emerging-in-moocs/

Hiltz, S. R., & Wellman, B. (1997). Asynchronous learning networks as a virtual classroom. *Communications of the ACM,* 40(9). http://dx.doi.org/10.1145/260750.260764

Holland, P. W., & Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, *76*(373), 33–50. http://dx.doi.org/10.1080/01621459.1981.10477598

Hollands, F. M., & Tirthali, D. (2014). Why do institutions offer MOOCs? *Online Learning: Official Journal of the Online Learning Consortium*, *18*(3). Retrieved from https://olj.onlinelearningconsortium.org/index.php/olj/article/view/464

Hood, N., Littlejohn, A., & Milligan, C. (2015). Context counts: How learners' contexts influence learning in a MOOC. *Computers & Education*, *91*, 83–91. http://dx.doi.org/10.1016/j.compedu.2015.10.019

Horn, M., & Christensen, C. (2013, February 20). Beyond the buzz, where are MOOCs really going? *Wired*. Retrieved from https://www.wired.com/2013/02/beyond-the-mooc-buzz-where-are-they-going-really/

Howley, I., Mayfield, E., & Rosé, C. P. (2013). Linguistic analysis methods for studying small groups. In C.E. Hmelo-Silver, C.A. Chinn, C. Chan, & A. O'Donnell (Eds.), *The international handbook of collaborative learning* (pp. 184–203). New York/London: Routledge.

Howley, I., Mayfield, E., Rosé, C. P., & Strijbos, J.-W. (2013). A multivocal process analysis of social positioning in study groups. In D. Suthers, K. Lund, C.P. Rosé, C. Teplovs, N. Law (Eds.), *Productive multivocality in the analysis of group interactions* (pp. 205–223). Springer US. http://dx.doi.org/10.1007/978-1-4614-8960-3_11

Hrastinski, S. (2008). What is online learner participation? A literature review. *Computers & Education*, *51*(4), 1755–1765. http://dx.doi.org/10.1016/j.compedu.2008.05.005

Hrastinski, S. (2009). A theory of online learning as online participation. *Computers & Education*, *52*(1), 78–82. http://dx.doi.org/10.1016/j.compedu.2008.06.009

Hrastinski, S., & Keller, C. (2007). Computer-mediated Communication in Education: A review of recent research. *Educational Media International*, *44*(1), 61–77. https://dx.doi.org/10.1080/09523980600922746

Huang, J., Dasgupta, A., Ghosh, A., Manning, J., & Sanders, M. (2014). Superposter behavior in MOOC forums. *Proceedings of the First ACM Conference on Learning @ Scale (L@S '14)*, 117–126. https://dx.doi.org/10.1145/2556325.2566249

Hunter, D. R., Goodreau, S. M., & Handcock, M. S. (2013). ergm. userterms: A Template Package for Extending statnet. *Journal of Statistical Software*, *52*(2), i02.

Järvelä, S., & Häkkinen, P. (2002). Web-based cases in teaching and learning–the quality of discussions and a stage of perspective taking in asynchronous communication. *Interactive Learning Environments*, *10*(1), 1–22. http://dx.doi.org/10.1076/ilee.10.1.1.3613

Jayaprakash, S. M., Moody, E. W., Lauría, E. J., Regan, J. R., & Baron, J. D. (2014). Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics*, *1*(1), 6–47. http://dx.doi.org/10.18608/jla.2014.11.3

Jiang, S., Fitzhugh, S. M., & Warschauer, M. (2014). Social positioning and performance in MOOCs. In *G-EDM workshop (Lynch and Barnes) in EDM 2014 Extended Proceedings (Gutierrez-Santos and Santos)*, 4-7July (pp.55-58), Institute of Education, London, UK.

Jiang, S., Warschauer, M., Williams, A. E., O'Dowd, D., & Schenke, K. (2014). Predicting MOOC performance with week 1 behavior. In J. Stamper, J., Z. Pardos, M. Mavrikis, B. McLaren (Eds.) *Proceedings of the 7th International Conference on Educational Data Mining (EDM'14)* (pp. 273–275). Retrieved from http://educationaldatamining.org/EDM2014/uploads/procs2014/shortpapers/273_EDM-2014-Short.pdf

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, *1*(2), 112–133. https://dx.doi.org/10.1177/1558689806298224

Joksimović, S., Dowell, N., Skrypnyk, O., Kovanović, V., Gašević, D., Dawson, S., & Graesser, A. (2015). How do you connect? Analysis of social capital accummulation in connectivist MOOCs. *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge (LAK'15*), 64-68. http://dx.doi.org/10.1145/2723576.2723604

Joksimović, S., Gašević, D., Loughin, T. M., Kovanović, V., & Hatala, M. (2015). Learning at distance: Effects of interaction traces on academic achievement. *Computers & Education*, *87*, 204–217. http://dx.doi.org/10.1016/j.compedu.2015.07.002

Joksimović, S., Manataki, A., Gašević, D., Dawson, S., Kovanović, V., & de Kereki, I. F. (2016). Translating network position into performance: importance of centrality in different network configurations. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK'16)* (pp. 314–323). https://dx.doi.org/10.1145/2883851.2883928

Joksimović, S., Poquet, O., Kovanović, V., Dowell, N., Mills, C., Gašević, D., … Brooks, C. (2017). *Modelling learning at scale.* Manuscript submitted for publication.

Jona, K., & Naidu, S. (Eds.) (2014). MOOCs: emerging research. *Distance Education, 35*(2). Retrieved from http://www.tandfonline.com/toc/cdie20/35/2

Jones, C. (2004a). Networks and learning: communities, practices and the metaphor of networks. *Research in Learning Technology*, *12*(1), 81–93. Retrieved from http://journals.co-action.net/index.php/rlt/article/view/11227

Jones, C. (2004b). The conditions of learning in networks. In *Kaleidoscope CSCL SIG First Symposium, October 7th-9th 2004, Switzerland*. Retrieved from https://telearn.archives-ouvertes.fr/hal-00190649/

Jones, C. (2015). *Networked learning: An educational paradigm for the age of digital networks*. Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-01934-5

Jones, C., & Esnault, L. (2004). *The metaphor of networks in learning: Communities, collaboration and practice*. Paper presented at the 4th International Conference on Networked Learning 2004, Lancaster University, UK. Retrieved from https://telearn.archives-ouvertes.fr/hal-00190280/

Jones, C., Ferreday, D., & Hodgson, V. (2008). Networked learning a relational approach: weak and strong ties. *Journal of Computer Assisted Learning*, *24*(2), 90–102. http://dx.doi.org/10.1111/j.1365-2729.2007.00271.x

Jordan, K. (2014). Initial trends in enrolment and completion of massive open online courses Massive Open Online Courses. *International Review of Research in Open and Distance Learning, 15*(1), 133–160. http://dx.doi.org/10.19173/irrodl.v15i1.1651

Jung, I., Choi, S., Lim, C., & Leem, J. (2002). Effects of different types of interaction on learning achievement, satisfaction and participation in web-based instruction. *Innovations in Education and Teaching International*, *39*(2), 153–162. http://dx.doi.org/10.1080/14703290252934603

Justice, T., & Jamieson, D. W. (2012). *The facilitator's fieldbook (3rd ed.)*. USA: HRD Press.

Kadushin, C. (2012). *Understanding social networks*: *Theories, concepts, findings*. New York, USA: Oxford University Press.

Kasworm, C. E., Sandmann, L. R., & Sissel, P. A. (2000). Adult learners in higher education. In A. Wilson, E. Hayes (Eds.), *Handbook of adult and continuing education* (pp. 449–463). San Francisco, US: Jossey-Bass.

Keegan, D. (1996). *Foundations of distance education (3rd ed.)*. London, UK: Routledge.

Kellogg, S., Booth, S., & Oliver, K. (2014). A social network perspective on peer supported learning in MOOCs for educators. *The International Review of Research in Open and Distributed Learning*, *15*(5). http://dx.doi.org/10.19173/irrodl.v15i5.1852

Kennedy, G., Coffrin, C., de Barba, P., & Corrin, L. (2015). Predicting success: How learners' prior knowledge, skills and activities predict MOOC performance. *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge*, 136–140. https://dx.doi.org/10.1145/2723576.2723593

Kilgore, W., & Lowenthal, P. (2015). The Human Element MOOC: An experiment in social presence. In R. Wright (Ed.), *Student-Teacher Interaction in Online Learning Environments* (pp. 373–391). IGI Global.

Kirschner, P., & Kester, L. (Eds.) (2016). Special issue: Learning analytics in Massively Multi-User Virtual Environments and Courses. *Journal of Computer Assisted Learning, 32(*3). http://dx.doi.org/10.1111/jcal.v32.3

Kizilcec, R., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in Massive Open Online Courses. *Proceedings of the Third International Conference on Learning Analytics and Knowledge (LAK '13)*, 170–179. https://doi.org/10.1145/2460296.2460330

Kizilcec, R., & Schneider, E. (2015). Motivation as a lens to understand online learners: toward data-driven design with the OLEI scale. *ACM Transactions on Computer-Human Interaction (TOCHI) - Special Issue on Online Learning at Scale*, *22*(2), Art.6, 1–24. https://dx.doi.org/10.1145/2699735

Kizilcec, R., Schneider, E., Cohen, G. L., & McFarland, D. A. (2014). Encouraging forum participation in online courses with collectivist, individualist and neutral motivational framings. *eLearning Papers, 37*, 13 – 22. Retrieved from https://www.openeducationeuropa.eu/sites/default/files/legacy_files/asset/In_depth_37_2.pdf

Knoke, D., & Yang, S. (2008). *Social network analysis* (Quantitative Applications in the Social Sciences Series, vol. 154). Sage Publications, Inc.

Knowledge@Wharton, D. (2015, January 5). The Hype is Dead but MOOCs Are Marching On [Web log post]. Retrieved from http://knowledge.wharton.upenn.edu/article/moocs-making-progress-hype-died/

Knox, J. (2014). Digital culture clash: "massive" education in the E-learning and Digital Cultures MOOC. *Distance Education*, *35*(2), 164–177. https://dx.doi.org/10.1080/01587919.2014.917704

Knox, J. (2016). *Posthumanism and the Massive Open Online Course: Contaminating the subject of global education*. New York/London: Routledge.

Koedinger, K. R., Kim, J., Jia, J. Z., McLaughlin, E. A., & Bier, N. L. (2015). Learning is not a spectator sport: doing is better than watching for learning from a MOOC. *Proceedings of the Second ACM Conference on Learning @ Scale* (L@S'15), 111–12. https://dx.doi.org/10.1145/2724660.2724681

Koller, D., Ng, A., Do, C., & Chen, Z. (2013, June). Retention and intention in massive open online courses: In depth. *Educause Review*, *48*(3), 62–63. Retrieved from http://er.educause.edu/articles/2013/6/retention-and-intention-in-massive-open-online-courses-in-depth

Kolowich, S. (2013, November 20). MOOCs are largely reaching privileged learners, survey finds. *The Chronicle of Higher Education*. Retrieved from http://www.chronicle.com/blogs/wiredcampus/moocs-are-reaching-only-privileged-learners-survey-finds/48567

Kop, R., Fournier, H., & Mak, S. F. J. (2011). A pedagogy of abundance or a pedagogy to support human beings? Participant support on Massive Open Online Courses. *International Review of Research in Open and Distance Learning*, *12*(7). http://dx.doi.org/10.19173/irrodl.v12i7.1041

Koschmann, T. (1996). Paradigm shifts and instructional technology. In T. Koschmann (Ed.), *CSCL: Theory and practice of an emerging paradigm (*pp. 1–23*)*. Mahwah, NJ: Lawrence Erlbaum Associates.

Koschmann, T. (2001). Revisiting the paradigms of instructional technology. In *Meeting at the crossroads.* In *G. Kennedy, M. Keppell, C. McNaught & T. Petrovic (Eds.), Meeting at the Crossroads. Proceedings of the 18th Annual Conference of the Australian Society for Computers in Learning in Tertiary Education, Melbourne: Biomedical Multimedia Unit, The University of Melbourne* (pp. 15–22).

Koschmann, T. (2002). Dewey's contribution to the foundations of CSCL research. In G. Stahl (Ed.), *Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community (CSCL'02)* (pp. 17–22), Boulder, Colorado, USA. New Jersey, US: Lawrence Erlbaum Associates, Inc.

Kovanović, V., Joksimović, S., Gašević, D., Siemens, G., & Hatala, M. (2014). What public media reveals about MOOCs? *British Journal of Educational Technology, 46*(3), 510–527. http://dx.doi.org/10.1111/bjet.12277

Kovanović, V., Joksimović, S., Gašević, D., Siemens, G., & Hatala, M. (2015). What public media reveals about MOOCs: A systematic analysis of news reports. *British Journal of Educational Technology*, *46*(3), 510–527. http://dx.doi.org/10.1111/bjet.12277

Kovanović, V., Joksimović, S., Poquet, O., Hennis, T., Cukic, I., de Vries, P., … Gašević, D. (2017). *Exploring communities of inquiry in Massive Open Online Courses*. Manuscript under review.

Kozma, R. B. (1991). Learning with media. *Review of Educational Research*, *61*(2), 179–211. https://dx.doi.org/10.3102/00346543061002179

Kozma, R. B. (1994). The influence of media on learning: The debate continues. *School Library Media Research*, *22*(4).

Krackhardt, D. (1998). Super Strong and Sticky. In R. Kramer, M. Neale (Eds.), *Power and influence in organizations* (pp. 21–38). SAGE Publications International.

Kreijns, K., Kirschner, P. A., Jochems, W., & Buuren, H. (2011). Measuring perceived social presence in distributed learning groups. *Education and Information Technologies*, *16*(4), 365–381. https://dx.doi.org/10.1007/s10639-010-9135-7

Kreijns, K., Kirschner, P., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: a review of the research. *Computers in Human Behavior*, *19*(3), 335–353. https://dx.doi.org/10.1016/S0747-5632(02)00057-2

Kreijns, K., Kirschner, P., & Vermeulen, M. (2013). Social Aspects of CSCL Environments: A Research Framework. *Educational Psychologist, 48*(4), 229–242. https://dx.doi.org/10.1080/00461520.2012.750225

Kreijns, K., Van Acker, F., Vermeulen, M., & Van Buuren, H. (2014). Community of Inquiry: social presence revisited. *E-Learning and Digital Media*, *11*(1), 5–18. https://dx.doi.org/10.2304/elea.2014.11.1.5

Kuznetsov, S. (2006). Motivations of contributors to Wikipedia. *ACM SIGCAS Computers and Society*, *36*(2), Art.1. http://dx.doi.org/10.1145/1215942.1215943

Kvålseth, T. O. (1989). Note on Cohen's kappa. *Psychological Reports*, *65*(1), 223–226. https://dx.doi.org/10.2466/pr0.1989.65.1.223

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159–174.

Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge, UK: Cambridge University Press.

Lazega, E., & Snijders, T. A. (Eds.) (2015). *Multilevel network analysis for the social sciences: theory, methods and applications* (Methodos Series, Vol. 12). Switzerland: Springer International Publishing.

Levina, N., & Arriaga, M. (2014). Distinction and status production on user-generated content platforms: Using Bourdieu's theory of cultural production to understand social dynamics in online fields. *Information Systems Research*, *25*(3), 468–488. http://dx.doi.org/10.1287/isre.2014.0535

Lewis, L., Snow, K., Farris, E., & Levin, D. (1999). *Distance education at postsecondary education institutions: 1997-98. (*National Center for Education Statistics Report No. 2000-013*)*. Washington, DC: US Department of Education. Retrieved from http://nces.ed.gov/pubs2000/2000013.pdf

Lin, N. (2008). A network theory of social capital. In D. Castiglione, J.W. van G. Woleb (Eds.) *The handbook of social capital* (*pp. 50–69)*. Oxford: Oxford University Press.

Liu W., Kidziński Ł., Dillenbourg P. (2016) Semiautomatic Annotation of MOOC Forum Posts. In: Li Y. et al. (Eds.) *State-of-the-art and future directions of smart learning (*Lecture notes in educational Technology Series*)*. Singapore. Springer. http://dx.doi.org/10.1007/978-981-287-868-7_48

Livingstone, D. W. (2001). *Adults' informal learning: Definitions, findings, gaps, and future research* (NALL Working Paper# 21). Canada: Ontario Inst. for Studies in Education, Toronto. Retrieved from http://eric.ed.gov/?id=ED452390

Liyanagunawardena, T. (2014). MOOC experience: A participant's reflection. *ACM SIGCAS Computers and Society Newsletter*, *44*(1), 9–14. https://dx.doi.org/10.1145/2602147.2602149

Liyanagunawardena, T., Adams, A. A., & Williams, S. A. (2013). MOOCs: A systematic study of the published literature 2008-2012. *The International Review of Research in Open and Distributed Learning, 14*(3). http://dx.doi.org/10.19173/irrodl.v14i3.1455

Lowenthal, P. (2012). *Social presence: What is it? How do we measure it?* (Doctoral dissertation, University of Colorado at Denver) Retrieved from http://gradworks.umi.com/35/06/3506428.html

Lusher, D., Koskinen, J., & Robins, G. (2012). *Exponential random graph models for social networks: Theory, methods, and applications*. New York, USA: Cambridge University Press.

Macfadyen, L., & Dawson, S. (2010). Mining LMS data to develop an "early wanting system" for educators: A proof of concept. *Computers & Education*, *54*(2), 588–599. http://dx.doi.org/10.1016/j.compedu.2009.09.008

Macleod, H., Haywood, J., Woodgate, A., & Alkhatnai, M. (2015). Emerging patterns in MOOCs: Learners, course designs and directions. *TechTrends*, *59*(1), 56–63.

Margaryan, A., Bianco, M., & Littlejohn, A. (2014). Instructional Quality of Massive Open Online Courses. *Computers & Education*. http://dx.doi.org/10.1016/j.compedu.2014.08.005

Marginson, S. (2012, July 8). *The problem of public good (s) in higher education*. Paper presented at 41st Australian Conference of Economists, Melbourne, Australia.

Martin, A. J., & Dowson, M. (2009). Interpersonal relationships, motivation, engagement, and achievement: Yields for theory, current issues, and educational practice. *Review of Educational Research*, *79*(1), 327–365. https://doi.org/10.3102/0034654308325583

Mathis, W. J. (2017). The Effectiveness of Class Size Reduction. *The Education Digest*, *82*(5), 60.

McAuley, A. A., Stewart, B., Siemens, G., & Cormier, D. (2010). The MOOC Model for Digital Practice. Retrieved from http://www.elearnspace.org/Articles/MOOC_Final.pdf

McDonald, J. (1998). *Interpersonal aspects of group dynamics and development in computer conferencing* (Doctoral dissertation, The University of Wisconsin – Madison). Retrieved from Proquest Database.

McIntyre, J. (2005). *Adult learning and Australia's ageing population: A policy briefing paper*. Canberra, ACT: Adult Learning Australia.

McMillan, D. W., & Chavis, D. M. (1986). Sense of community: A definition and theory. *Journal of Community Psychology*, *14*(1), 6–23.

Meister, J. (2016, January). *The Future of Leadership Webinar*. Retrieved from https://blessingwhite.com/wp-content/uploads/2016/01/Future-Leader_BlessingWhite_Jan2016_Webinar.pdf

Meyer, D. K., & Turner, J. C. (2002). Discovering emotion in classroom motivation research. *Educational Psychologist*, *37*(2), 107–114. http://dx.doi.org/10.1207/S15326985EP3702_5

Milana, M. (2012). Political globalization and the shift from adult education to lifelong learning. *European Journal for Research on the Education and Learning of Adults*, *3*(2), 103–117. http://dx.doi.org/10.3384/rela.2000-7426.rela0070

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage Publications.

Milligan, C., Littlejohn, A., & Hood, N. (2016). Learning in MOOCs: a comparison study. In M. Khalil, M. Ebner, M. Kopp, A. Lorenz, & M. Kalz (Eds.). *Proceedings of the European Stakeholder Summit on experiences and best practices in and around MOOCs (eMOOCs'16)* (pp. 15–26). Retrieved from https://www.emoocs2016.eu/

Milligan, C., Littlejohn, A., & Margaryan, A. (2013). Patterns of engagement in connectivist MOOCs. *MERLOT Journal of Online Learning and Teaching*, *9*(2). Retrieved from http://jolt.merlot.org/vol9no2/milligan_0613.htm

Milligan, S. (2015). Crowd-sourced learning in MOOCs: Learning analytics meets measurement theory. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge (LAK'15)* (pp. 151–155). https://dx.doi.org/10.1145/2723576.2723596

Milligan, S., & Griffin, P. (2016). Understanding learning and learning design in MOOCs: A measurement-based interpretation. *Journal of Learning Analytics*, *3*(2), 88–115. https://dx.doi.org/10.18608/jla.2016.32.5

Mische, A. (2011). Relational sociology, culture, and agency. In J. Scott and P. Carrington (Eds.), *The Sage handbook of social network analysis* (pp. 80–97). London: SAGE Publications.

Miyazoe, T., & Anderson, T. (2013). Interaction equivalency in an OER, MOOCS and informal learning era. *Journal of Interactive Media in Education, 2,* Art. 9. 1–15. http://dx.doi.org/10.5334/2013-09

Moe, R. (2014). *The evolution and impact of the massive open online course*. (Doctoral dissertation, Pepperdine University). Retrieved from http://gradworks.umi.com/36/25/3625105.html

Moe, R. (2016). The Phenomenal MOOC. In Veletsianos, G. (Ed.), *Emergence and Innovation in Digital Learning: Foundations and Applications (pp.163–178)*. Athabasca: AU Press.

Molm, L. D. (2003). Theoretical comparisons of forms of exchange. *Sociological Theory*, *21*(1), 1–17. http://dx.doi.org/10.1111/1467-9558.00171

Molm, L. D. (2010). The structure of reciprocity. *Social Psychology Quarterly*, *73*(2), 119–131.https://dx.doi.org/10.1177/0190272510369079

Molm, L. D., Collett, J. L., & Schaefer, D. R. (2007). Building solidarity through generalized exchange: A theory of reciprocity 1. *American Journal of Sociology*, *113*(1), 205–242. http://dx.doi.org/10.1086/517900

Molm, L. D., Peterson, G., & Takahashi, N. (1999). Power in negotiated and reciprocal exchange. *American Sociological Review*, 876–890.

Monge, P. R., & Contractor, N. S. (2003). *Theories of communication networks*. USA: Oxford University Press.

Mood, T. A. (1995). *Distance education: An annotated bibliography*. Retrieved from http://eric.ed.gov/?id=ED380113

Moore, M. (1989). Three Types of Interaction. *The American Journal of Distance Education*, *3*(2), 1–7.

Moore, M. G. (1997). Theory of transactional distance. In D. Keegan (Ed.), *Theoretical principles of distance education* (pp. 22–38). New York: Routledge.

Morgan, D. L. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research*, *1*(1), 48–76. https://dx.doi.org/10.1177/2345678906292462

Morris, M., Handcock, M. S., & Hunter, D. R. (2008). Specification of Exponential-Family Random Graph Models: Terms and Computational Aspects. *Journal of Statistical Software*, *24*(4), 1548–7660.

Moutsios, S. (2011). Global governance and education policy. *Unge Paedagoger*, *3*, 62–68

Mulder, A. (2016, December 9). *MOOCs for Credit .. goes live* [Web log post]. Retrieved from https://ankamulder.weblog.tudelft.nl/tag/credits-for-moocs/

Mullaney, T. (2014). Making sense of Moocs: A reconceptualization of HarvardX Courses and their students. (Essay presented for a B.A. degree, Harvard College). *Available at SSRN 2463736*. http://dx.doi.org/10.2139/ssrn.2463736

Muller, M., Shami, N. S., Millen, D. R., & Feinberg, J. (2010). We are all lurkers: consuming behaviors among authors and readers in an enterprise file-sharing service. In *Proceedings of the 16th ACM international conference on Supporting group work,* 201–210. http://dx.doi.org/10.1145/1880071.1880106

Mustafaraj, E., & Bu, J. (2015). The visible and invisible in a MOOC discussion forum. In *Proceedings of the Second ACM Conference on Learning @ Scale (L@S'15)*, 351–354. https://dx.doi.org/10.1145/2724660.2728691

Myers, K., Conte, N., & Rubenson, K. (2014). *Adult Learning Typology: Adult Learning and Returns to Training Project* (Report for the Social Research and Demonstration Corporation) Retrieved from http://www.srdc.org/media/199755/adult-learning-typology.pdf

Naidu, S., & Järvelä, S. (2006). Analyzing CMC content for what? *Computers & Education*, *46*(1), 96–103. http://dx.doi.org/10.1016/j.compedu.2005.04.001

Nam, K. K., Ackerman, M. S., & Adamic, L. A. (2009). Questions in, knowledge in?: a study of naver's question answering community. *Proceedings of the SIGCHI conference on human factors in computing systems (CHI'09)*, 779–788. http://dx.doi.org/10.1145/1518701.1518821

Newman, D. R., Webb, B., & Cochrane, C. (1995). A content analysis method to measure critical thinking in face-to-face and computer supported group learning. *Interpersonal Computing and Technology*, *3*(2), 56–77.

Newman, M. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, *98*(2), 404–409. http://dx.doi.org/10.1073/pnas.98.2.404

Newman, M. (2004). Analysis of weighted networks. *Physical Review E*, *70*(5), 56131. https://doi.org/10.1103/PhysRevE.70.056131

Ng'ambi, N., & Bozalek, V. (Eds.) (2015). Special Issue: Massive Open Online Courses (MOOCs): 'disrupting' teaching and learning practices in higher education, *46*(3), 451–678. http://dx.doi.org/10.1111/bjet.2015.46.issue-3

Nipper, S. (1989). Third generation distance learning and computer conferencing. *Mindweave: Communication, Computers and Distance Education*, 63–73.

Nonnecke, B., & Preece, J. (2001). Why lurkers lurk. *AMCIS 2001 Proceedings*, 294, 1521–1530. Retrieved from http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1733&context=amcis2001

Ochoa, X., & Duval, E. (2008). Quantitative analysis of user-generated content on the web. In *Proceedings of webevolve2008: web science workshop at WWW2008,* 1–8. Retrieved from https://lirias.kuleuven.be/handle/123456789/200042

Ochoa, X., & Worsley, M. (2016). Editorial: Augmenting Learning Analytics with Multimodal Sensory Data. *Journal of Learning Analytics*, *3*(2), 213–219. https://dx.doi.org/10.18608/jla.2016.32.10

OECD (Organisation for Economic Co-operation and Development). (1996). *The Knowledge-based economy*. Paris, France: OECD.

OECD (Organisation for Economic Co-operation and Development). (2007). *Participative Web: User-generated Content* (Report prepared by the OECD Committee for Information, Computer and Communications Policy). Retrieved from https://www.oecd.org/sti/38393115.pdf

Oh, S. (2012). The characteristics and motivations of health answerers for sharing information, knowledge, and experiences in online environments. *Journal of the American Society for Information Science and Technology*, *63*(3), 543–557. https://dx.doi.org/10.1002/asi.21676

Opsahl, T. (2013). Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. *Social Networks*, *35*(2), 159–167. http://dx.doi.org/10.1016/j.socnet.2011.07.001

Orellana, A. (2006). Class size and interaction in online courses. *Quarterly Review of Distance Education*, *7*(3), 229–248.

Oren, A., Mioduser, D., & Nachmia, R. (2002). The sevelopment of social climate in virtual learning discussion groups. *The International Review of Research in Open and Distributed Learning*, *3*(1). http://dx.doi.org/10.19173/irrodl.v3i1.80

Palincsar, A. S. (1998). Social constructivist perspectives on teaching and learning. *Annual Review of Psychology*, *49*(1), 345–375. http://dx.doi.org/10.1146/annurev.psych.49.1.345

Papadopoulos, K., Sritanyaratana, L., & Klemmer, S. (2014). Community TAs scale high-touch learning, provide student-staff brokering, and build esprit de corps. *In Proceedings of the first ACM conference on Learning @ scale conference (L@S'14)*, 163–164. http://dx.doi.org/10.1145/2556325.2567860

Pappano, L. (2012, November 2). The year of the MOOC. *The New York Times*. Retrieved from http://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html?pagewanted=all&_r=0

Parr, C. (2014, October 14). Moocs "Will Not Transform Education", Says FutureLearn Chief. *Times Higher Education (THE)*. Retrieved from https://www.timeshighereducation.com/news/moocs-will-not-transform-education-says-futurelearn-chief/2016379.article

Pea, R. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *The Journal of the Learning Sciences*, *13*(3), 423–451. http://dx.doi.org/10.1207/s15327809jls1303_6

Pena-Shaff, J. B., & Nicholls, C. (2004). Analyzing student interactions and meaning construction in computer bulletin board discussions. *Computers & Education*, *42*(3), 243–265. http://dx.doi.org/10.1016/j.compedu.2003.08.003

Piaget, J. (1954). Language and thought from a genetic perspective. *Acta Psychologica*, *10*, 51–60.

Poole, M. S., Hollingshead, A. B., McGrath, J. E., Moreland, R. L., & Rohrbaugh, J. (2004). Interdisciplinary perspectives on small groups. *Small Group Research*, *35*(1), 3–16. https://dx.doi.org/10.1177/1046496403259753

Poquet (Skrypnyk), O & Dawson, S. (2015). Analysis of MOOC forum participation. In *Globally connected, digitally enabled (Ascilite 2015)* (pp. 224–234). Perth: Australasian Society for Computers in Learning in Tertiary Education.

Poquet, O., Kovanović, V., de Vries, P., Hennis, T., Joksimović, S., Gašević, D., & Dawson, S. (2017). *Social presence in Massive Open Online Courses*. Manuscript under review.

Postmes, T., & Spears, R. (1998). Deindividuation and antinormative behavior: A meta-analysis. *Psychological Bulletin*, *123*(3), 238–259. http://dx.doi.org/10.1037/0033-2909.123.3.238

Postmes, T., Spears, R., & Lea, M. (1999). Social identity, group norms, and "deindividuation": Lessons from computer-mediated communication for social influence in the group. In N. Ellemers, R. Spears, & B. Doosje (Eds.), S*ocial identity: Context, commitment, content* (pp.164–184). Oxford: Blackwell.

Prasarnphanich, P., & Wagner, C. (2009). The role of wiki technology and altruism in collaborative knowledge creation. *Journal of Computer Information Systems*, *49*(4), 33–41.

Preece, J., Nonnecke, B., & Andrews, D. (2004). The top five reasons for lurking: improving community experiences for everyone. *Computers in Human Behavior*, *20*(2), 201–223. http://dx.doi.org/10.1016/j.chb.2003.10.015

Prensky, M. (2001). Digital natives, digital immigrants part 1. *On the Horizon*, *9*(5), 1–6. http://dx.doi.org/10.1108/10748120110424816

Prentice, D. A., Miller, D. T., & Lightdale, J. R. (1994). Asymmetries in attachments to groups and to their members: Distinguishing between common-identity and common-bond groups. *Personality and Social Psychology Bulletin*, *20*(5), 484–493. https://doi.org/10.1177/0146167294205005

R Core Team. (2014). *R: A language and environment for statistical computing*. [Software]. R Foundation for Statistical Computing, Vienna, Austria. 2013.

Raffaghelli, J. E., Cucchiara, S., & Persico, D. (2015). Methodological approaches in MOOC research: Retracing the myth of Proteus. *British Journal of Educational Technology*, *46*(3), 488–509. https://dx.doi.org/10.1111/bjet.12279

Ramesh, A., Goldwasser, D., Huang, B., Daume, H., & Getoor, L. (2014). Uncovering hidden engagement patterns for predicting learner performance in MOOCs. In *Proceedings of the First ACM Conference on Learning @ Scale Conference (L@S' 14),* 157–158. https://dx.doi.org/10.1145/2556325.2567857

Ravenna, G., Foster, C., & Bishop, C. (2012). Increasing student interaction online : A Review of the literature. *Journal of Technology and Teacher Education*, *20*(2), 177–203.

Reich, J. (2014, December 8). MOOC Completion and Retention in the Context of Student Intent. *EDUCAUSE Review*. Retrieved from http://er.educause.edu/articles/2014/12/mooc-completion-and-retention-in-the-context-of-student-intent

Reich, J. (2015). Rebooting MOOC research. *Science, 347*(6217), 34–35. http://dx.doi.org/10.1126/science.1261627

Ren, Y., Harper, F. M., Drenner, S., Terveen, L. G., Kiesler, S. B., Riedl, J., & Kraut, R. E. (2012). Building Member Attachment in Online Communities: Applying

Theories of Group Identity and Interpersonal Bonds. *Mis Quarterly*, *36*(3), 841–864.

Ren, Y., Kraut, R., & Kiesler, S. (2007). Applying common identity and bond theory to design of online communities. *Organization Studies*, *28*(3), 377–408. https://doi.org/10.1177/0170840607076007

Rivard, R. (2013, July 18). Udacity Project on "Pause." *Inside Higer Ed*. Retrieved from https://www.insidehighered.com/news/2013/07/18/citing-disappointing-student-outcomes-san-jose-state-pauses-work-udacity

Robertson, S., Bonal, X., & Dale, R. (2002). GATS and the Education Service Industry. The Polities of Scale and Reterritoriazation. *Comparative Education Review*, *46*(4), 472–496. http://dx.doi.org/10.1086/343122

Robertson, S., Dale, R., Moutsios, S., Nielsen, G., Shore, C., & Wright, S. (2012). *Globalisation and regionalisation in higher education: Toward a new conceptual framework* (Working Papers on University Reform). Copenhagen, Denmark: Department of Education, Aarhus University.

Robertson, S., & Keeling, R. (2008). Stirring the lions: strategy and tactics in global higher education. *Globalisation, Societies and Education*, *6*(3), 221–240. https://dx.doi.org/10.1080/14767720802343316

Robins, G. (2015). *Doing social network research: Network-based research design for social scientists*. Sage Publications Ltd.

Robins, G., Elliott, P., & Pattison, P. (2001). Network models for social selection processes. *Social Networks*, *23*(1), 1–30. https://dx.doi.org/10.1016/S0378-8733(01)00029-6

Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social Networks*, *29*(2), 173–191. http://dx.doi.org/10.1016/j.socnet.2006.08.002

Robins, G., Snijders, T., Wang, P., Handcock, M., & Pattison, P. (2007). Recent developments in exponential random graph (p*) models for social networks. *Social Networks*, *29*(2), 192–215. http://dx.doi.org/10.1016/j.socnet.2006.08.003

Rosé, C. P., Carlson, R., Yang, D., Wen, M., Resnick, L., Goldman, P., & Sherer, J. (2014). Social factors that contribute to attrition in MOOCs. In *Proceedings of the First ACM Conference on Learning @ Scale Conference (L@S'14),* 197–198. https://dx.doi.org/10.1145/2556325.2567879

Rosé, C. P., Ferschke, O., Tomar, G., Yang, D., Howley, I., Aleven, V., … Gašević, D. (2015). Challenges and opportunities of dual-layer MOOCs: Reflections from an edX deployment study. In *Proceedings of the 11th International Conference on Computer Supported Collaborative Learning (CSCL 2015),* (Vol. 2, pp. 848–851). Gothenburg, Sweden: International Society of the Learning Sciences.

Ross, J., Sinclair, C., Knox, J., Bayne, S., & Macleod, H. (2014). Teacher experiences and academic identity: The missing components of MOOC Pedagogy. *MERLOT*

Journal of Online Learning and Teaching, *10*(1), 57–69. Retrieved from http://jolt.merlot.org/vol10no1/ross_0314.pdf

Rossi, L. A., & Gnawali, O. (2014). Language independent analysis and classification of discussion threads in Coursera MOOC forums. In *Information Reuse and Integration (IRI), 2014 IEEE 15th International Conference on* (pp. 654–661). http://dx.doi.org/10.1109/IRI.2014.7051952

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (2001). Methodological issues in the content analysis of computer conference transcripts. *International Journal of Artificial Intelligence in Education (IJAIED)*, *12*, 8–22.

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (2007). Assessing social presence in asynchronous text-based computer conferencing. *International Journal of E-Learning & Distance Education*, *14*(2), 50–71. Retrieved from http://www.ijede.ca/index.php/jde/article/view/153

Rourke, L., & Kanuka, H. (2009). Learning in Communities of Inquiry: A Review of the Literature. *Journal of Distance Education*, *23*(1), 19–48. Retrieved from http://ijede.ca/index.php/jde/article/view/474

Rovai, A. (2002). Building sense of community at a distance, *The International Review of Research in Open and Distributed Learning, 3*(1), 1–16. http://dx.doi.org/10.19173/irrodl.v3i1.79

Rovai, A., & Barnum, K. T. (2007). On-line course effectiveness: An analysis of student interactions and perceptions of learning. *International Journal of E-Learning & Distance Education*, *18*(1), 57–73. Retrieved from http://www.ijede.ca/index.php/jde/article/view/121

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American Psychologist*, *55*(1), 68–78. http://dx.doi.org/10.1037/0003-066X.55.1.68

Perez-Sanagustin, M., Maldonaldo, J., & Morales, N. (2016). *Status report on the adoption of MOOCs in higher education in Latin America and Europe* (Report as a part of the MOOC-Maker Project "Building Management Capacity for MOOCs in Higher Education"). Retrieved from http://www.mooc-maker.org/wp-content/files/D1.1-InformeMOOCLatam-vFINALDEFINITIVO_English.pdf

Sassenberg, K. (2002). Common-bond and common-identity groups on the Internet: Attachment and normative behavior in on-topic and off-topic chats. *Group Dynamics: Theory, Research, and Practice*, *6*(1), 27.

Scardamalia, M., & Bereiter, C. (1996). Computer support for knowledge building communities. In T. Koschmann (Ed.), *CSCL: Theory and Practice of an Emerging Paradigm* (249–269). Malwah, New Jersey: Lawrence Erlbaum Associates Inc. Publishers.

Schmid, L., Manturuk, K., Simpkins, I., Goldwasser, M., & Whitfield, K. E. (2015). Fulfilling the promise: do MOOCs reach the educationally underserved?

Educational Media International, *52*(2), 116–128. https://dx.doi.org/10.1080/09523987.2015.1053288

Schneider, E., & Kizilcec, R. (2014). Why did you enroll in this course?: Developing a standardized survey question for reasons to enroll. In *Proceedings of the first ACM Conference on Learning @ Scale* (L@S'14), 147–148. https://dx.doi.org/10.1145/2556325.2567852

Schrire, S. (2006). Knowledge building in asynchronous discussion groups: Going beyond quantitative analysis. *Computers & Education*, *46*(1), 49–70. https://dx.doi.org/10.1016/j.compedu.2005.04.006

Scott, J. (2013). *Social network analysis* (3rd Edition). SAGE Publications Ltd.

Selwyn, N., & Bulfin, S. (2014). *The discursive construction of MOOCs as educational opportunity and educational threat*. MOOC Research Initiative (MRI) - Final Report. Retrieved from http://goo.gl/ok43eT

Selwyn, N., Bulfin, S., & Pangrazio, L. (2015). Massive open online change? Exploring the discursive construction of the "MOOC" in newspapers. *Higher Education Quarterly*, *69*(2), 175–192. http://dx.doi.org/10.1111/hequ.12061

Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher*, *27*(2), 4–13. https://dx.doi.org/10.3102/0013189X027002004

Shackelford, J. L., & Maxwell, M. (2012). Sense of community in graduate online education: Contribution of learner to learner interaction. *The International Review of Research in Open and Distributed Learning*, *13*(4), 228–249. http://dx.doi.org/10.19173/irrodl.v13i4.1339

Shah, D. (2015, December 21). By the numbers: MOOCs in 2015. [Web log post]. Retrieved from https://www.class-central.com/report/moocs-2015-stats/

Shah, D. (2016a, October 16). XuetangX: A look at China's first and biggest MOOC platform — Class Central. [Web log post]. Retrieved from https://www.class-central.com/report/xuetangx/

Shah, D. (2016b, December 22). 6 biggest MOOC trends of 2016. [Web log post]. Retrieved from https://www.class-central.com/report/biggest-mooc-trends-2016/

Sharma, K., Jermann, P., & Dillenbourg, P. (2015). Identifying styles and path towards success in MOOCs. In O. Santos, J. Boticario, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, J.M. Luna, C. Mihaescu, P. Moreno, A.Hershkovitz, S.Ventura, and M. Desmarais (Eds.) Proceedings of the 8th International Conference on Educational Data Mining (EDM'15), Madrid, Spain (pp. 408–411). Retrieved from http://eric.ed.gov/?id=ED560766

Shea, P., Sau Li, C., & Pickett, A. (2006). A study of teaching presence and student sense of learning community in fully online and web-enhanced college courses. *The Internet and Higher Education*, *9*(3), 175–190. https://dx.doi.org/10.1016/j.iheduc.2006.06.005

Shillair, R., & Wash, R. (2015). Are you listening?: Social roles and perceived value of statements in online learning communities. In *Proceedings of the Second (2015) ACM Conference on Learning @ Scale* (pp. 377–380). https://dx.doi.org/10.1145/2724660.2728697

Shirky, C. (2012, November 12). Napster, Udacity, and the Academy. [Web log post]. Retrieved from http://www.shirky.com/weblog/2012/11/napster-udacity-and-the-academy/

Short, J., Williams, E., & Christie, B. (1976). The social psychology of telecommunications. Retrieved from http://www.citeulike.org/group/1662/article/327279

Siemens, G. (2005). Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning*, *2*(1), 3–10. Retrieved from http://er.dut.ac.za/handle/123456789/69

Siemens, G. (2008, October 30). Connectivism Course (CCK08). [Web log post]. Retrieved from http://www.elearnspace.org/blog/2008/10/30/connectivism-course-cck08/

Siemens, G. (2013a). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*, *57*(10), 1380–1400. https://dx.doi.org/10.1177/0002764213498851

Siemens, G. (2013b). Massive open online courses: Innovation in education? In R. McGreal, W. Kinuthia, S. Marshall (Eds.), *Open Educational Resources: Innovation, Research and Practice* (pp.5–15). Athabasca: Commonwealth of Learning, Athabasca University.

Siemens, G. (2014) (Ed.) Special Issue: Research into Massive Open Online Courses. *The International Review of Open and Distributed Learning, 15*(5). Retrieved from http://www.irrodl.org/index.php/irrodl/issue/view/64

Siemens, G., & Baker, R. (2012). Learning analytics and educational data mining: towards communication and collaboration. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 252–254). http://dx.doi.org/10.1145/2330601.2330661

Siemens, G., Gašević, D. & Dawson, S. (2015). Preparing for the Digital University: a review of the history and current state of distance, blended, and online learning. Arlington: Link Research Lab. Retrieved October 24, 2016 from http://linkresearchlab.org/PreparingDigitalUniversity.pdf

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, *46*(5), 30.

Simmel, G. (2009/1858–1918). *Sociology: Inquiries into the construction of social forms (2 Vols.)* (*Anthony J. Blasi, Anton K. Jacobs, Mthew Kanjirathinkal Trans)*. Leiden/Boston: Brill.

Sinha, T. (2014). Together we stand, together we fall, together we win: Dynamic team formation in massive open online courses. In *Applications of Digital Information and Web Technologies (ICADIWT), 2014 Fifth International Conference on the* (pp. 107–112). https://dx.doi.org/10.1109/ICADIWT.2014.6814694

Skrypnyk, O., Joksimović, S., Kovanović, V., Gašević, D., & Dawson, S. (2014). Roles of course facilitators, learners, and technology in the flow of information of a cMOOC. *The International Review of Research in Open and Distributed Learning*, *16*(3). http://dx.doi.org/10.19173/irrodl.v16i3.2170

Smith, T. W., Marsden, P. V., & Hout, M. (2015). *General Social Surveys, 1972–2014: Cumulative codebook.* Chicago: NORC.

Snijders, T. A., Pattison, P. E., Robins, G. L., & Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology*, *36*(1), 99–153.

Spencer, M. H. (1974). *Contemporary economics*. New York, NY: Worth

Stewart, B. (2013). Massiveness + Openness = New Literacies of Participation? *MERLOT Journal of Online Learning and Teaching*, *9*(2). Retrieved from http://jolt.merlot.org/vol9no2/stewart_bonnie_0613.htm

Stich, A. E., & Reeves, T. D. (2017). Massive open online courses and underserved students in the United States. *The Internet and Higher Education*, *32*, 58–71. http://dx.doi.org/10.1016/j.iheduc.2016.09.001

Strijbos, J.-W., Martens, R. L., Prins, F. J., & Jochems, W. M. G. (2006). Content analysis: What are they talking about? *Computers & Education*, *46*(1), 29–48. https://dx.doi.org/10.1016/j.compedu.2005.04.002

Stump, G. S., DeBoer, J., Whittinghill, J., & Breslow, L. (2013). Development of a framework to classify MOOC discussion forum posts: Methodology and challenges. [TLL Working Paper]. Retrieved from https://tll.mit.edu/sites/default/files/library/Coding_a_MOOC_Discussion_Forum.pdf

Suthers, D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning*, *1*(3), 315–337. https://dx.doi.org/10.1007/s11412-006-9660-y

Suthers, D. (2015). From contingencies to network-level phenomena: Multilevel analysis of activity and actors in heterogeneous networked learning environments. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge (LAK'15)*, 368–377. https://dx.doi.org/10.1145/2723576.2723626

Suthers, D., Hoppe, H. U., de Laat, M., & Shum, S. B. (2012). Connecting levels and methods of analysis in networked learning communities. In *Proceedings of the*

2nd International Conference on Learning Analytics and Knowledge (pp. 11–13). http://dx.doi.org/10.1145/2330601.2330608

Suthers, D., & Verbert, K. (2013). Learning analytics as a middle space. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge (LAK'13),* 1–4. http://dx.doi.org/10.1145/2460296.2460298

Sutton, L. A. (2001). The principle of vicarious interaction in computer-mediated communications. *International Journal of Educational Telecommunications*, *7*(3), 223–242.

Swan, K. (2002). Building learning communities in online courses: The importance of interaction. *Education, Communication & Information*, *2*(1), 23–49. http://dx.doi.org/10.1080/1463631022000005016

Swan, K. (2003a). Developing social presence in online course discussions. In S. Naidu (Ed.), *Learning and teaching with technology: Principles and tractices*, 147–164. UK: Routledge.

Swan, K. (2003b). Learning effectiveness: what the research tells us. In J. Bourne & J. C. Moore (Eds.), *Elements of quality online wducation, practice and direction*. Needham, MA: Sloan Center for Online Education, 13-45.

Swan, K. (2004). Relationships between interactions and learning in online environments. *The Sloan Consortium*. Retrieved from http://bestpracticemodels.wiki.staffs.ac.uk/@api/deki/files/99/=interactions.pdf

Tajfel, H. (1978). *Differentiation between social groups*. London: Academic Press.

Takahashi, N. (2000). The emergence of generalized exchange. *American Journal of Sociology*, *105*(4), 1105–1134. http://dx.doi.org/10.1086/210400

Tashakkori, A., & Teddie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches (Applied Social Research Methods Series,* v.46). SAGE Publications.

Taylor, J. (2001). Fifth generation distance education. *e-Journal of Instructional Science & Technology, 4*(1), 1-14. Retrieved from http://eprints.usq.edu.au/136/

Telecompaper. (2016, March 2). Telefonica's MiriadaX e-learning platform hits 2 mln users. *Telecompaper*. Retrieved from https://www.telecompaper.com/news/telefonicas-miriadax-e-learning-platform-hits-2-mln-users--1131236

Thomas, M. (2002). Learning within incoherent structures: the space of online discussion forums. *Journal of Computer Assisted Learning*, *18*, 351–366. https://dx.doi.org/10.1046/j.0266-4909.2002.03800.x

Thomas, S. (2000). Ties that bind: A social network approach to understanding student integration and persistence. *Journal of Higher Education*, 591–615.

Thompson, M. (2007). From Distance Education to E-Learning. In R. Andrews & C. Haythornthwaite (Eds.), *The SAGE handbook of e-learning research* (pp. 159–179). UK: SAGE Publications Ltd.

Tomei, L. A. (2006). The impact of online teaching on faculty load: Computing the ideal class size for online courses. *Journal of Technology and Teacher Education*, *14*(3), 531.

Tu, C.-H. (2002). The Measurement of Social Presence in an Online Learning Environment. *International Journal on E-Learning*, *1*(2), 34–45.

Tu, C.-H., & McIsaac, M. (2002). The relationship of social presence and interaction in online classes. *The American Journal of Distance Education*, *16*(3), 131–150. http://dx.doi.org/10.1207/S15389286AJDE1603_2

Tucker, R. (2016). FutureLearn and Deakin University the First to Offer Range of Degrees Delivered Entirely on a MOOC Platform. [Media Release]. Retrieved from http://www.deakin.edu.au/about-deakin/media-releases/articles/futurelearn-and-deakin-university-the-first-to-offer-range-of-degrees-delivered-entirely-on-a-mooc-platform

Tucker, C., Pursel, B., & Divinsky, A. (2014). *Mining student-generated textual data in MOOCs and quantifying their effects on student performance and learning outcomes.* Presented at 2014 ASEE Annual Conference, Indianapolis, Indiana, Indianapolis, Indiana.

Turner, J. C. (1985). Social categorization and the self-concept: A social cognitive theory of group behavior. *Advances in Group Processes*, *2*, 77–122.

Turner, T. C., Smith, M. A., Fisher, D., & Welser, H. T. (2005). Picturing Usenet: Mapping computer-mediated collective action. *Journal of Computer-Mediated Communication*, *10*(4). http://dx.doi.org/10.1111/j.1083-6101.2005.tb00270.x

van Mierlo, T. (2014). The 1% rule in four digital health social networks: an observational study. *Journal of Medical Internet Research*, 16(2), e33. http://dx.doi.org/10.2196/jmir.2966

Vaquero, L. M., & Cebrian, M. (2013). The rich club phenomenon in the classroom. *Scientific Reports*, *3*, 1174. https://dx.doi.org/10.1038/srep01174

Veletsianos, G. (Ed.) (2013). Learner experiences with MOOCs and open online learning. *Hybrid Pedagogy*. Retrieved from https://hybrid-pedagogy.github.io/LearnerExperiencesInMOOCs/LearnerExperiencesInMOOCs.pdf

Veletsianos, G., & Shepherdson, P. (2016). A systematic analysis and synthesis of the empirical MOOC literature published in 2013–2015. *The International Review of Research in Open and Distributed Learning*, *17*(2). http://dx.doi.org/10.19173/irrodl.v17i2.2448

Vrasidas, C., & McIsaac, M. S. (1999). Factors influencing interaction in an online course. *American Journal of Distance Education*, *13*(3), 22–36. http://dx.doi.org/10.1080/08923649909527033

Vu, D., Pattison, P., & Robins, G. (2015). Relational event models for social learning in MOOCs. *Social Networks*, *43*, 121–135. https://dx.doi.org/10.1016/j.socnet.2015.05.001

Vygotsky, L. S. (1980/1978). *Mind in society: The development of higher psychological processes*. USA: Harvard University Press.

Walji, S., Deacon, A., Small, J., & Czerniewicz, L. (2016). Learning through engagement: MOOCs as an emergent form of provision. *Distance Education*, *37*(2), 208–223. http://dx.doi.org/10.1080/01587919.2016.1184400

Walsh, T. (2011). *Unlocking the gates: How and why leading universities are opening up access to their courses*. Princeton University Press.

Walther, J. (1996). Computer-mediated communication impersonal, interpersonal, and hyperpersonal interaction. *Communication Research*, *23*(1), 3–43. https://dx.doi.org/10.1177/009365096023001001

Walther, J., Anderson, J. F., & Park, D. W. (1994). Interpersonal effects in computer-mediated interaction: A meta-analysis of social and antisocial communication. *Communication Research*, *21*(4), 460–487. https://dx.doi.org/10.1177/009365094021004002

Wang, X., Wen, M., & Rosé, C. P. (2016). Towards Triggering Higher-order Thinking Behaviors in MOOCs. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (*LAK'16),* 398–407. https://dx.doi.org/10.1145/2883851.2883964

Wang, X., Yang, D., Wen, M., Koedinger, K., & Rosé, C. P. (2015). Investigating how student's cognitive behavior in MOOC discussion forums affect learning gains. In O. Santos, J. Boticario, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, J.M. Luna, C. Mihaescu, P. Moreno, A.Hershkovitz, S.Ventura, and M. Desmarais (Eds.) *Proceedings of the 8th International Conference on Educational Data Mining (EDM'15*), Madrid, Spain (pp.226–223). Retrieved from http://www.educationaldatamining.org/EDM2015/proceedings/full226-233.pdf

Wasko, M. M., & Faraj, S. (2005). Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly*, 35–57.

Wasko, M. M., Faraj, S., & Teigland, R. (2004). Collective action and knowledge contribution in electronic networks of practice. *Journal of the Association for Information Systems*, *5*(11), 2.

Wasko, M. M., Teigland, R., & Faraj, S. (2009). The provision of online public goods: Examining social structure in an electronic network of practice. *Decision Support Systems*, *47*(3), 254–265. http://dx.doi.org/10.1016/j.dss.2009.02.012

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.

Wasserman, S., & Pattison, P. (1996). Logit models and logistic regressions for social networks: I. An introduction to Markov graphs andp. *Psychometrika*, *61*(3), 401–425. http://dx.doi.org/10.1007/BF02294547

Watson, S. L., Watson, W. R., Richardson, J., & Loizzo, J. (2016). Instructor's Use of Social Presence, Teaching Presence, and Attitudinal Dissonance: A Case Study of an Attitudinal Change MOOC. *The International Review of Research in Open and Distributed Learning*, *17*(3). http://dx.doi.org/10.19173/irrodl.v17i3.2379

Watters, A. (2015a, August 23). The MOOC Revolution That Wasn't. *The Kernel. The Daily Dot*. Retrieved from http://kernelmag.dailydot.com/issue-sections/headline-story/14046/mooc-revolution-uber-for-education/

Watters, A. (2015b, December 1). Top Ed-Tech Trends of 2015. [Web log post]. Retrieved from http://hackeducation.com/2015/12/01/trends

Webb, N. M. (1982). Student interaction and learning in small groups. *Review of Educational Research*, *52*(3), 421–445. https://dx.doi.org/10.3102/00346543052003421

Wellman, B. (1979). The community question: The intimate networks of East Yorkers. *American Journal of Sociology*, 1201–1231. http://dx.doi.org/10.1086/226906

Wellman, B. (1997). An electronic group is virtually a social network. In S. Kiesler (Ed.), *Culture of the Internet* (pp. 179–205). New York: Lawrence Erlbaum Associates, Inc.

Wellman, B. (2001). Computer networks as social networks. *Science*, *293*, 2031–4. https://dx.doi.org/10.1126/science.1065547

Wellman, B., & Gulia, M. (1999). Net surfers don't ride alone: Virtual communities as communities. In B. Wellman (Ed.), *Networks in the Global Village* (pp.331–366). Boulder, Co: Westview.

Wellman, B., Quan-Haase, A., Boase, J., Chen, W., Hampton, K., Díaz, I., & Miyata, K. (2003). The social affordances of the Internet for networked individualism. *Journal of Computer-Mediated Communication*, *8*(3). http://dx.doi.org/10.1111/j.1083-6101.2003.tb00216.x

Weltzer-Ward, L. (2011). Content analysis coding schemes for online asynchronous discussion. *Campus-Wide Information Systems*, *28*(1), 56–74. https://dx.doi.org/10.1108/10650741111097296

Wen, M., Yang, D., & Rosé, C. (2014). Sentiment Analysis in MOOC Discussion Forums: What does it tell us? In In J. Stamper, J., Z. Pardos, M. Mavrikis, B. McLaren (Eds.) *Proceedings of the 7th International Conference on Educational Data Mining (EDM'14)*, 130–137. Retrieved from http://educationaldatamining.org/EDM2014/uploads/procs2014/longpapers/130_EDM-2014-Full.pdf

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. New York: Cambridge university press.

Wenger, E. (2000). Communities of practice and social learning systems. *Organization*, *7*(2), 225–246. https://dx.doi.org/10.1177/135050840072002

Wenger, E., McDermott, R. A., & Snyder, W. (2002). *Cultivating communities of practice: A guide to managing knowledge*. USA: Harvard Business Press.

Wenger-Traynor, E., & Wenger-Traynor, B. (2011, December 28). *Communities versus networks*. [Web log post]. Retrieved from http://wenger-trayner.com/resources/communities-versus-networks/

White, D. S., & Le Cornu, A. (2011). Visitors and Residents: A new typology for online engagement. *First Monday*, *16*(9). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/viewArticle/3171

White, S., Davis, H., Dickens, K. P., Leon Urrutia, M., & Sanchez Vera, M. M. (2014). MOOCs: What motivates the producers and participants? *Communications in Computer and Information Science*, 1–16. Retrieved from http://eprints.soton.ac.uk/370440/

Whitehill, J., Williams, J. J., Lopez, G., Coleman, C. A., & Reich, J. (2015). *Beyond prediction: First steps toward automatic intervention in MOOC student stopout* (SSRN Scholarly Paper No. ID 2611750). Rochester, NY: Social Science Research Network. http://dx.doi.org/10.2139/ssrn.2611750

Wiley, D. (2002). A proposed measure of discussion activity in threaded discussion spaces (v0. 9). *Unpublished Working Draft Received from the Author (David. Wiley@ Byu. Edu)*. Retrieved from http://davidwiley.org/docs/discussion09.pdf

Wiley, D. (2009, November 16). Defining "Open." [Web log post]. Retrieved from http://opencontent.org/blog/archives/1123

Wise, A., Cui, Y., Jin, W., & Vytasek, J. (2017). Mining for gold: Identifying content-related MOOC discussion threads across domains through linguistic modeling. *The Internet and Higher Education*, *32*, 11–28. http://dx.doi.org/10.1016/j.iheduc.2016.08.001

Wise, A., Cui, Y., & Vytasek, J. (2016). Bringing order to chaos in MOOC discussion forums with content-related thread identification. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK'16),* 188–197. http://dx.doi.org/10.1145/2883851.2883916

Wise, A., & Shaffer, D. (2015). Why theory matters more than ever in the age of big data. *Journal of Learning Analytics*, *2*(2), 5–13. https://dx.doi.org/10.18608/jla.2015.22.2

Wise, A., Speer, J., Marbouti, F., & Hsiao, Y.-T. (2012). Broadening the notion of participation in online discussions: examining patterns in learners' online listening behaviors. *Instructional Science*, *41*(2), 323–343. https://dx.doi.org/10.1007/s11251-012-9230-9

Witthaus, G., Inamorato dos Santos. A., Childs, M., Tannhäuser, A., Conole, G., Nkuyubwatsi, B., Punie, Y. (2016). *Validation of non-formal MOOC-based learning: An analysis of assessment and recognition practices in Europe (OpenCred)* (JRC Science for Policy Report;27660). http://dx.doi.org/10.2791/809371

Woo, Y., & Reeves, T. C. (2007). Meaningful interaction in web-based learning: A social constructivist interpretation. *The Internet and Higher Education*, *10*(1), 15–25. https://dx.doi.org/10.1016/j.iheduc.2006.10.005

Woo, Y., & Reeves, T. C. (2008). Interaction in Asynchronous Web-Based Learning Environments. *Journal of Asynchronous Learning Networks*, *12*, 179–194.

Wu, C.-G., Gerlach, J. H., & Young, C. E. (2007). An empirical analysis of open source software developers' motivations and continuance intentions. *Information & Management*, *44*(3), 253–262. https://dx.doi.org/10.1016/j.im.2006.12.006

Wu, P. F., & Korfiatis, N. (2013). You scratch someone's back and we'll scratch yours: Collective reciprocity in social Q&A communities. *Journal of the American Society for Information Science and Technology*, *64*(10), 2069–2077. http://dx.doi.org/ 10.1002/asi.22913

Wubbels, T., den Brok, P., van Tartwijk, J., & Levy, J. (2012). *Interpersonal relationships in education: An overview of contemporary research* (Vol. 3). Springer Science & Business Media.

Yang, D., Sinha, T., Adamson, D., & Rosé, C. P. (2013). Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In *Proceedings of the 2013 NIPS Data-driven education workshop* (Vol. 11, p. 14). Retrieved from http://lytics.stanford.edu/datadriveneducation/papers/yangetal.pdf

Yang, D., Wen, M., Kumar, A., Xing, E. P., & Rosé, C. (2014). Towards an integration of text and graph clustering methods as a lens for studying social interaction in MOOCs. *The International Review of Research in Open and Distributed Learning*, *15*(5). http://dx.doi.org/10.19173/irrodl.v15i5.1853

Yang, D., Wen, M., Howley, I., Kraut, R., Rosé, C. (2015). Exploring the effect of confusion in discussion forums of Massive Open Online Courses. In *Proceedings of the Second (2015) ACM Conference on Learning @ Scale (L@S'15),* 121–130. http://dx.doi.org/10.1145/2724660.2724677

Yang, J., Morris, M. R., Teevan, J., Adamic, L. A., & Ackerman, M. S. (2011). Culture matters: A survey study of social Q&A behavior. *Icwsm*, *11*, 409–416.

Ye, C., Kinnebrew, J. S., Biswas, G., Evans, B. J., Fisher, D. H., Narasimham, G., & Brady, K. A. (2015). Behavior prediction in MOOCs using higher granularity temporal information. In *Proceedings of the Second (2015) ACM Conference on Learning @ Scale (L@S'15),* 335–338. https://dx.doi.org/10.1145/2724660.2728687

Zhang, J., Skryabin, M., & Song, X. (2016). Understanding the dynamics of MOOC discussion forums with simulation investigation for empirical network analysis

(SIENA). *Distance Education*, *37*(3), 270–286. http://dx.doi.org/10.1080/01587919.2016.1226230

Zhang, W. (2007). Trust in Electronic Networks of Practice: An Integrative Model. In C. Steinfield, B. T. Pentland, M. Ackerman, & N. Contractor (Eds.), *Communities and Technologies 2007* (pp. 351–370). Springer London. https://dx.doi.org/10.1007/978-1-84628-905-7_18

Zheng, S., Rosson, M. B., Shih, P. C., & Carroll, J. M. (2015). Understanding student motivation, behaviors and perceptions in MOOCs. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing (CSCW'15),* 1882–1895. http://dx.doi.org/10.1145/2675133.2675217

Zhenghao, C., Alcorn, B., Christensen, G., Eriksson, N., Koller, D., & Emanuel, E. J. (2015, September 22). Who's benefitting from MOOCs, and why. *Harvard Business Review*. Retrieved from https://hbr.org/2015/09/whos-benefiting-from-moocs-and-why

Zhu, M., Bergner, Y., Zhang, Y., Baker, R., Wang, Y., & Paquette, L. (2016). Longitudinal engagement, performance, and social connectivity: a MOOC case study using exponential random graph models. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK'16)*, 223–230. https://dx.doi.org/10.1145/2883851.2883934

Appendix A Ethics Clearance

1) First application to the Online Human Research Ethics Committee

From: no_reply@unisa.edu.au <no_reply@unisa.edu.au>
Sent: Friday, November 28, 2014 1:29 PM
To: Skrypnyk, Oleksandra - skroy001; humanethics@unisa.edu.au
Subject: Human ethics: Application exempt

Dear Applicant

Re: Ethics protocol "Understanding Scaled Online Learning" (Application ID: 0000033878)

Thank you for submitting your application through the Online Human Research Ethics Application and Review System. Based on the answers you provided, the risk associated with your research project was assessed and your research project does not require further ethics review. Your application is thus considered Exempt and you can commence your research.

Please retain a copy of this system generated email as evidence of this assessment for auditing purposes.

Kind regards

Executive Officer
UniSA's Human Research Ethics Committee
CRICOS provider number 00121B

This is an automated email and cannot be replied to.
Please refer to our FAQ page: http://w3.unisa.edu.au/res/ethics/human/ethicsfaq.asp for any system-related queries. Otherwise, please direct your queries to humanethics@unisa.edu.au.

2) Second application to the Online Human Research Ethics Committee (upon the decision to add more empirical research to the thesis)

From: no_reply@unisa.edu.au <no_reply@unisa.edu.au>
Sent: Wednesday, November 30, 2016 5:05 PM
To: humanethics@unisa.edu.au; Skrypnyk, Oleksandra - skroy001; Ruth Geer
Subject: Human Ethics: Application approved

Dear Applicant

Re: Ethics protocol "Understanding Scaled Online Learning - Addition" (Application ID: 0000036182)

Thank you for submitting your ethics protocol for consideration. Your protocol has been considered by the E2 Committee Review Group.

I am pleased to advise that your protocol has been granted ethics approval and meets the requirements of the National Statement on Ethical Conduct in Human Research.

Please note that the E2 Committee Review Group's decision will be reported to the next meeting of the Human Research Ethics Committee for endorsement.

Please regard this email as formal notification of approval.

Ethics approval is always made on the basis of a number of conditions detailed at http://www.unisa.edu.au/res/forms/docs/humanresearchethics_conditions.doc; it is important that you are familiar with, and abide by, these conditions. It is also essential that you conduct all research according to UniSA guidelines, which can be found at http://www.unisa.edu.au/res/ethics/default.asp

Please note, if your project is a clinical trial you are required to register it in a publicly accessible trials registry prior to enrolment of the first participant (e.g. Australian New Zealand Clinical Trials Registry http://www.anzctr.org.au/) as a condition of ethics approval. Best wishes for your research.

Executive Officer
UniSA's Human Research Ethics Committee
CRICOS provider number 00121B

This is an automated email and cannot be replied to. Please direct your query to humanethics@unisa.edu.au.

Appendix B edX terms of Service and Privacy Policy

1. edX Terms of Service

The terms of service were retrieved from https://www.edx.org/edx-terms-service. On October 22, 2014, edX adopted amended Terms of Service, providing as follows

Welcome to edX. Please read these Terms of Service ("TOS") and the edX Privacy Policy and Honor Code prior to registering for edX.org or using any portion of the edX website (the "Site," which consists of all content and pages located within the edX.org web domain), including accessing any course material, chat rooms, or other electronic services. These TOS, the Privacy Policy and the Honor Code are agreements (the "Agreements") between you and edX. By using the Site, you accept and agree to be legally bound by the Agreements, whether or not you are a registered user. If you do not understand or do not wish to be bound by the terms of the Agreements, you should not use the Site.

EdX reserves the right to modify these TOS at any time without advance notice. Any changes to these TOS will be effective immediately upon posting on this page, with an updated effective date. By accessing the Site after any changes have been made, you signify your agreement on a prospective basis to the modified TOS and all of the changes. Be sure to return to this page periodically to ensure familiarity with the most current version of these TOS.

DESCRIPTION OF EDX

EdX offers online courses that include opportunities for professor-to-student and student-to-student interactivity, individual assessment of a student's work and, for students who demonstrate their mastery of subjects, a certificate of achievement or other acknowledgment.

RULES FOR ONLINE CONDUCT

You agree that you are responsible for your own use of the Site and for your User Postings. "User Postings" include all content submitted, posted, published or distributed on the Site by you or other users of the Site, including but not limited to all forum posts, wiki edits, notes, questions, comments, videos and file uploads. You agree that you will use the Site in compliance with these TOS, the Honor Code and all applicable local, state, national and international laws, rules and regulations, including copyright laws, any laws regarding the transmission of technical data exported from your country of residence, and all United States export control laws.

As a condition of your use of the edX services, you will not use the Site in any manner intended to damage, disable, overburden or impair any edX server or the network(s) connected to any edX server or to interfere with any other party's use and enjoyment of the Site. You may not attempt to gain unauthorized access to the Site, other accounts, computer systems or networks connected to any edX server through hacking, password mining or any other means. You may not obtain or attempt to obtain any materials or information stored on the Site, its servers or associated computers through any means not intentionally made available through the Site.

THE FOLLOWING LIST OF ITEMS IS STRICTLY PROHIBITED ON THE SITE:

- Content that defames, harasses or threatens others;
- Content that discusses illegal activities with the intent to commit them;
- Content that infringes another's intellectual property, including, but not limited to, copyrights or trademarks;
- Profane, pornographic, obscene, indecent or unlawful content;
- Advertising or any form of commercial solicitation;
- Content related to partisan political activities;
- Viruses, trojan horses, worms, time bombs, corrupted files, malware, spyware or any other similar software that may damage the operation of another's computer or property; and
- Content that contains intentionally inaccurate information or that is posted with the intent of misleading others.

Furthermore, you agree not to scrape, or otherwise download in bulk, any Site content, including but not limited to a list or directory of users on the system, on-line textbooks, User Postings or user information. You agree not to misrepresent or attempt to misrepresent your identity while using the Site (although you are welcome and encouraged to use an anonymous username in the forums and to act in a manner that keeps your identity concealed).

USER ACCOUNTS AND AUTHORITY

In order to participate fully in Site activities, you must provide your name, an email address and a user password in order to create a user account ("User Account"). You agree that you will never divulge or share access or access information for your User Account with any third party for any reason. In setting up your User Account, you may be prompted to enter additional optional information (e.g., your address). You represent that all information provided by you is accurate and current. You agree to maintain and update your information to keep it accurate and current.

We care about the confidentiality and security of your personal information. Please see our Privacy Policy for more information about what information about you edX collects and how edX uses that information.

YOUR RIGHT TO USE CONTENT ON THE SITE

Unless indicated as being in the public domain, the content on the Site is protected by United States and foreign copyright laws. Unless otherwise expressly stated on the Site, the texts, exams, video, images and other instructional materials provided with the courses offered on this Site are for your personal use in connection with those courses only. We aim to make much of the edX course content available under more open license terms that will help create a vibrant ecosystem of contributors and further edX's goal of making education accessible and affordable to the world.

Certain reference documents, digital textbooks, articles and other information on the Site are used with the permission of third parties, and use of that information is subject to certain rules and conditions, which will be posted along with the information. By using this Site you agree to abide by all such rules and conditions.

You agree to retain all copyright and other notices on any content you obtain from the Site. All rights in the Site and its content, if not expressly granted, are reserved.

USER POSTINGS

User Postings Representations and Warranties. By submitting or distributing your User Postings, you affirm, represent and warrant (1) that you have the necessary rights, licenses, consents and/or permissions to reproduce and publish the User Postings and to authorize edX and its users to reproduce, modify, publish and otherwise use and distribute your User Postings in a manner consistent with the licenses granted by you below, and (2) that neither your submission of your User Postings nor the exercise of the licenses granted below will infringe or violate the rights of any third party. You, and not edX, are solely responsible for your User Postings and the consequences of posting or publishing them.

License Grant to edX. By submitting or distributing your User Postings, you hereby grant to edX a worldwide, non-exclusive, transferable, assignable, sub licensable, fully paid-up, royalty-free, perpetual, irrevocable right and license to host, transfer, display, perform, reproduce, modify, distribute, re-distribute, relicense and otherwise use, make available and exploit your User Postings, in whole or in part, in any form and in any media formats and through any media channels (now known or hereafter developed).

License Grant to edX Users. By submitting or distributing your User Postings, you hereby grant to each user of the Site a non-exclusive license to access and use your User Postings in connection with their use of the Site for their own personal purposes.

CERTIFICATES, ETC.

Generally. EdX and the colleges, universities and other institutions providing courses on the Site (collectively, the "Members") may offer a certificate of achievement or other acknowledgment (a "Certificate") for students who, in their judgment, have satisfactorily demonstrated mastery of the course material. Certificates will be issued by edX under the name of the underlying Member from where the course originated, i.e. HarvardX, MITx, etc. The decision whether a Certificate will be awarded to a given student will be solely within the discretion of the awarding Member. EdX and/or the Members providing courses on the Site may choose not to offer a Certificate for some courses.

Registration and Payment. Subject to the foregoing, you may be required to pay a fee to receive an ID Verified Certificate of Achievement (a “Verified Certificate”) with respect to certain courses on the Site, in which case, the following processes and requirements shall apply:

- As part of your registration, you will be prompted to submit a photo of yourself, along with one other form of photo identification (“photo ID”). Although these items are collected by edX in accordance with these TOS and the edXPrivacy Policy, you should be aware that the actual authentication of your identify is performed by an edX third party service provider. Please refer to the Privacy Policy for additional details on how we handle those items once received by us.
- In addition, the processing of your payment information is done by an edX third party vendor, and you will be re-routed to a secure website to complete the payment transaction. The privacy statement posted at the payment landing page operated by such third party vendor will govern the submission and processing of your payment details. Please familiarize yourself with those terms prior to initiating a transaction.
- Once your payment transaction is completed, you will be re-routed back to an edX confirmation page and will receive a confirmation email with your name, order number and the payment amount. Please retain this email for your records as this information will be required if you seek a refund from edX, as described below.

Course Withdrawal and Refunds. For a period of fourteen (14) calendar days from the course start date (“Drop Period”), you may withdraw or unregister from the course by clicking on the unregister link on the student dashboard and request a refund by emailing billing@edx.org. Please include your full name and order number in your email request. Refunds will be credited

to the credit card used when you registered and may take up to two billing cycles to process. EdX has no obligation to issue a refund after the Drop Period, but if you believe a refund is warranted, please email us atbilling@edx.org.

No Other Enrollment. When you take a course through edX, you will not be an applicant for admission to, or enrolled in, any degree program of the Member as a result of registering for or completing a course provided by such Member through edX. You will not be entitled to use any of the resources of the Member beyond the online courses provided on the Site, nor will you be eligible to receive student privileges or benefits provided to students enrolled in degree programs of the Member.

AUTHENTICATION

In order that edX can authenticate your identity as required by these TOS, you will be prompted to take a webcam photo of yourself, as well as a photo of an acceptable form of photo ID (described below). Once submitted to edX, this information will be subject to the edX Privacy Policy and will be used only for the purpose of verifying your identity. Acceptable forms of photo ID's are:

- Government or State-issued driver's license
- Passport
- National ID card
- State or Province ID card (including cards issued by motor vehicle agencies)

In order to be accepted by edX, your photo ID must:

- Contain your full name exactly (excluding hyphens, accents and spaces)
- Contain a relatively current photograph of yourself
- Be an original document; photocopied documents cannot be accepted
- Be current and valid; expired documents cannot be accepted

Unfortunately, if you do not posses a photo ID meeting the criteria described above, edX is unable to provide you with an ID Verified Certificate at this time.

TRADEMARKS

Use of edX, MIT, Harvard University and Member Names, Trademarks and Service Marks. The "edX," "MIT," and "Harvard University" names, logos and seals are trademarks ("Trademarks") of the respective entities. Likewise, the names, logos, and seals of the other Members are Trademarks owned by the respective Member. You may not use any of these Trademarks, or any variations thereof, without the owner's prior written consent. You may not use any of these Trademarks, or any variations thereof, for promotional purposes, or in any way that deliberately or inadvertently claims, suggests or, in these institutions' sole judgment, gives the appearance or impression of a relationship with or endorsement by these institutions.

All Trademarks not owned by edX or the Members that appear on the Site or on or through the services made available on or through the Site, if any, are the property of their respective owners.

Nothing contained on the Site should be construed as granting, by implication, estoppel or otherwise, any license or right to use any Trademark displayed on the Site without the written permission of the owner of the applicable Trademark.

DIGITAL MILLENNIUM COPYRIGHT ACT

Copyright owners who believe their material has been infringed on the Site should contact edX's designated copyright agent at dmca-agent@edx.org or at 141 Portland St. 9th Floor, Cambridge, MA 02139, Attention: edX DMCA Agent/General Counsel.

Notification must include:

- Identification of the copyrighted work, or, in the case of multiple works at the same location, a representative list of such works at that site.
- Identification of the material that is claimed to be infringing or to be the subject of infringing activity. You must include sufficient information for us to locate the material (e.g., URL, IP address, computer name).
- Information for us to be able to contact the complaining party (e.g., email address, phone number).
- A statement that the complaining party believes that the use of the material has not been authorized by the copyright owner or an authorized agent.
- A statement that the information in the notification is accurate and that the complaining party is authorized to act on behalf of the copyright owner.

DISCLAIMERS OF WARRANTY / LIMITATIONS OF LIABILITIES

THE SITE AND ANY INFORMATION, CONTENT OR SERVICES MADE AVAILABLE ON OR THROUGH THE SITE ARE PROVIDED "AS IS" AND "AS AVAILABLE" WITHOUT WARRANTY OF ANY KIND (EXPRESS, IMPLIED OR OTHERWISE), INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NON-INFRINGEMENT, EXCEPT INSOFAR AS ANY SUCH IMPLIED WARRANTIES MAY NOT BE DISCLAIMED UNDER APPLICABLE LAW.

EDX AND THE EDX PARTICIPANTS (AS HERINAFTER DEFINED) DO NOT WARRANT THAT THE SITE WILL OPERATE IN AN UNINTERRUPTED OR ERROR-FREE MANNER, THAT THE SITE IS FREE OF VIRUSES OR OTHER HARMFUL COMPONENTS, OR THAT THE COURSES OR CONTENT PROVIDED WILL MEET YOUR NEEDS OR EXPECTATIONS. EDX AND THE EDX PARTICIPANTS ALSO MAKE NO WARRANTY ABOUT THE ACCURACY, COMPLETENESS, TIMELINESS, OR QUALITY OF THE SITE OR ANY COURSES OR CONTENT, OR THAT ANY PARTICULAR COURSES OR CONTENT WILL CONTINUE TO BE MADE AVAILABLE. “EDX PARTICIPANTS” MEANS MIT, HARVARD, THE OTHER MEMBERS, THE ENTITIES PROVIDING INFORMATION, CONTENT OR SERVICES FOR THE SITE, THE COURSE INSTRUCTORS AND THEIR STAFFS.

USE OF THE SITE, AND THE CONTENT AND SERVICES OBTAINED FROM OR THROUGH THE SITE, IS AT YOUR OWN RISK. YOUR ACCESS TO OR DOWNLOAD OF INFORMATION, MATERIALS OR DATA THROUGH THE SITE OR ANY REFERENCE SITES IS AT YOUR OWN DISCRETION AND RISK, AND YOU WILL BE SOLELY RESPONSIBLE FOR ANY DAMAGE TO YOUR PROPERTY (INCLUDING YOUR COMPUTER SYSTEM) OR LOSS OF DATA THAT RESULTS FROM THE DOWNLOAD OR USE OF SUCH MATERIAL OR DATA, UNLESS OTHERWISE EXPRESSLY PROVIDED FOR IN THE EDXPRIVACY POLICY.

User Postings Disclaimer. You understand that when using the Site you will be exposed to User Postings from a variety of sources and that neither edX nor the edX Participants are responsible for the accuracy, usefulness, reliability or intellectual property rights of or relating to such User Postings. You further understand and acknowledge that you may be exposed to User Postings that are inaccurate, offensive, defamatory, indecent or objectionable and you agree to waive, and hereby do waive, any legal or equitable rights or remedies you have or may

have against edX or any of the edX Participants with respect thereto. Neither edX nor any of the edX Participants endorse any User Postings or any opinion, recommendation or advice expressed therein. Neither edX nor any of the edX Participants have any obligation to monitor any User Postings or any other user communications through the Site.

However, edX reserves the right to review User Postings and to exercise its sole discretion to edit or remove, in whole or in part, any User Posting at any time and for any reason, or to allow the edX Participants to do so. Without limiting the foregoing, upon receiving notice from a user or a content owner that a User Posting allegedly does not conform to these TOS, edX may investigate the allegation and determine in its sole discretion whether to remove the User Posting, which it reserves the right to do at any time and without notice.

Links to Other Sites. The Site may include hyperlinks to sites maintained or controlled by others. EdX and the edX Participants are not responsible for and do not routinely screen, approve, review or endorse the contents of or use of any of the products or services that may be offered at these sites. If you decide to access linked third-party websites, you do so at your own risk.

TO THE FULLEST EXTENT PERMITTED BY APPLICABLE LAW, YOU AGREE THAT NEITHER EDX NOR ANY OF THE EDX PARTICIPANTS WILL BE LIABLE TO YOU FOR ANY LOSS OR DAMAGES, EITHER ACTUAL OR CONSEQUENTIAL, ARISING OUT OF OR RELATING TO THESE TERMS OF SERVICE, OR YOUR (OR ANY THIRD PARTY'S) USE OF OR INABILITY TO USE THE SITE, OR YOUR PLACEMENT OF CONTENT ON THE SITE, OR YOUR RELIANCE UPON INFORMATION OBTAINED FROM OR THROUGH THE SITE, WHETHER YOUR CLAIM IS BASED IN CONTRACT, TORT, STATUTORY OR OTHER LAW.

IN PARTICULAR, TO THE FULLEST EXTENT PERMITTED BY APPLICABLE LAW, NEITHER EDX NOR ANY OF THE EDX PARTICIPANTS WILL HAVE ANY LIABILITY FOR ANY CONSEQUENTIAL, INDIRECT, PUNITIVE, SPECIAL, EXEMPLARY OR INCIDENTAL DAMAGES, WHETHER FORESEEABLE OR UNFORESEEABLE AND WHETHER OR NOT EDX OR ANY OF THE EDX PARTICIPANTS HAS BEEN NEGLIGENT OR OTHERWISE AT FAULT (INCLUDING, BUT NOT LIMITED TO, CLAIMS FOR DEFAMATION, ERRORS, LOSS OF PROFITS, LOSS OF DATA OR INTERRUPTION IN AVAILABILITY OF DATA).

CERTAIN STATE LAWS DO NOT ALLOW LIMITATIONS ON IMPLIED WARRANTIES OR THE EXCLUSION OR LIMITATION OF CERTAIN DAMAGES. IF THESE LAWS APPLY TO YOU, SOME OR ALL OF THE ABOVE DISCLAIMERS, EXCLUSIONS, OR LIMITATIONS MAY NOT APPLY TO YOU, AND YOU MIGHT HAVE ADDITIONAL RIGHTS.

INDEMNIFICATION

You agree to defend, hold harmless and indemnify edX and the edX Participants, and their respective subsidiaries, affiliates, officers, faculty, students, fellows, governing board members, agents and employees from and against any third-party claims, actions or demands arising out of, resulting from or in any way related to your use of the Site, including any liability or expense arising from any and all claims, losses, damages (actual and consequential), suits, judgments, litigation costs and attorneys' fees, of every kind and nature. In such a case, edX or one of the edX Participants will provide you with written notice of such claim, suit or action.

MISCELLANEOUS

Termination Rights; Discontinuation of Courses and Content. You agree that edX, in its sole discretion, may terminate your use of the Site or your participation in it, for any reason or no reason, upon notice to you. It is edX's policy to terminate in appropriate circumstances users of the Site who are repeat copyright infringers. EdX and the edX Participants reserve the right at any time in their sole discretion to cancel, delay, reschedule or alter the format of any course offered through edX, or to cease providing any part or all of the Site content or related services, and you agree that neither edX nor any of the edX Participants will have any liability to you for such an action. If you no longer desire to participate in the Site, you may terminate your participation at any time. The rights granted to you hereunder will terminate upon any termination of your right to use the Site, but the other provisions of these TOS will survive any such termination.

Entire Agreement. These TOS, the Honor Code, and the Privacy Policy together constitute the entire agreement between you and edX with respect to your use of the Site, superseding any prior agreements between you and edX regarding your use of the Site.

Waiver and Severability of TOS. The failure of edX to exercise or enforce any right or provision of these TOS shall not constitute a waiver of such right or provision. If any provision of these TOS is found by a court of competent jurisdiction to be invalid, the parties nevertheless agree that the court should endeavor to give effect to the parties' intentions as reflected in the provision and the other provisions of these TOS shall remain in full force and effect.

Choice of Law/Forum Selection. You agree that these TOS and any claim or dispute arising out of or relating to these TOS or any content or service obtained from or through the Site will be governed by the laws of the Commonwealth of Massachusetts, excluding its conflicts of law provisions. You agree that all such claims and disputes will be heard and resolved exclusively in the federal or state courts located in and serving Cambridge, Massachusetts, U.S.A. You consent to the personal jurisdiction of those courts over you for this purpose, and you waive and agree not to assert any objection to such proceedings in those courts (including any defense or objection of lack of proper jurisdiction or venue or inconvenience of forum).

Effective Date: October 22, 2014.

2. Honor Code Pledge

COLLABORATION POLICY

By enrolling in a course on edX, you are joining a special worldwide community of learners. The aspiration of edX is to provide anyone in the world who has the motivation and ability to engage coursework from the Massachusetts Institute of Technology, Harvard University and the other Members providing courses on the Site the opportunity to attain the best educational experience that internet technology enables. You are part of the community who will help edX achieve this goal.

EdX depends upon your motivation to learn the material and to do so with honesty. In order to participate in edX, you must agree to the Honor Code below and any additional terms specific to a class. This Honor Code, and any additional terms, will be posted on each class website.

EDX HONOR CODE PLEDGE

By enrolling in an edX course, I agree that I will:

- Complete all mid-terms and final exams with my own work and only my own work. I will not submit the work of any other person.
- Maintain only one user account and not let anyone else use my username and/or password.
- Not engage in any activity that would dishonestly improve my results, or improve or hurt the results of others.
- Not post answers to problems that are being used to assess student performance.

Unless otherwise indicated by the instructor of an edX course, learners on edX are encouraged to:

- Collaborate with others on the lecture videos, exercises, homework and labs.
- Discuss with others general concepts and materials in each course.
- Present ideas and written work to fellow edX learners or others for comment or criticism.

CHANGING THE HONOR CODE

Please note that we review and may make changes to this Honor Code from time to time. Any changes to this Honor Code will be effective immediately upon posting on this page, with an updated effective date. By accessing the Site after any changes have been made, you signify your agreement on a prospective basis to the modified Honor Code and any changes contained therein. Be sure to return to this page periodically to ensure familiarity with the most current version of this Honor Code.

Effective Date: October 31, 2014

3. edX Privacy Policy
The policy was retrieved from https://www.edx.org/edx-privacy-policy. On October 22, 2014, edX adopted an amended Privacy Policy, providing as follows:

At edX, we care about the confidentiality and security of your personal information. We will use commercially reasonable efforts to keep your Personal Information secure in compliance with this Privacy Policy. "Personal Information" is defined below. As used in this Privacy Policy, "we," "us," and "our" refer to edX.

This Privacy Policy only applies to information that we collect through the edX Website (the "Site," which consists of all content and pages located within the edX.org web domain) and does not apply to information that we may collect from you in other ways (for example, this policy does not apply to information that you may provide to us over the phone, by fax, or through conventional mail). In addition, please note that your education records are protected by the Family Educational Rights and Privacy Act ("FERPA") to the extent FERPA applies.

USERNAMES AND POSTINGS

Comments or other information posted by you to our forums, wikis or other areas of the Site designed for public communications or communications among registered class members may be viewed and downloaded by others who visit the Site. For this reason, we encourage you to use discretion when deciding whether to post any information that can be used to identify you to those forums (or other public or classwide areas).

CONSENT TO PROCESSING OUTSIDE HOME COUNTRY

By using our Site, or providing information to edX, you understand and unambiguously consent to the collection, use, disclosure and retention of information in the United States and other countries and territories, for the purposes set forth in this Privacy Policy and the edX Terms of Service. You hereby consent to the collection, use, disclosure, and retention by edX of your Personal Information as described under this Privacy Policy, including but not limited to the transfer of your personal data between edX and the third parties, affiliates, and subsidiaries described in this Privacy Policy. For further certainty, any consent relating to a right to transfer information referred to in this paragraph shall be deemed to include your consent to the transfer of the applicable Personal Information to a jurisdiction which may provide a different level of privacy protection than that available in your own country.

If you do not agree with these terms, then please do not access, browse, or register for the Site. If you choose not to provide us with certain information required to provide you with various services offered on our Site, you may not be able to establish a user account and we may not be able to provide you with those services.

PERSONAL INFORMATION

As used in this Privacy Policy, "Personal Information" means any information about yourself that you may provide to us when using the Site, such as when you sign up for a user account or enter into a transaction through the Site, which may include (but is not limited to) your name, contact information, gender, date of birth, occupation, and if you register for an ID Verified Certificate of Achievement, a driver’s license or other government issued identification.

We endeavor to limit the collection of Personal Information to what we need with respect to specific Site activities in order to satisfy business, legal, or regulatory requirements. For example, if you wish to purchase an ID Verified Certificate of Achievement, we will collect certain Personal Information necessary to authenticate your identity, as described in greater

detail in the edX Terms of Service. If you do not wish to submit the required authentication information, then you will not be able to purchase a certificate via the Site.

EdX does not currently collect financial information from you; however, if you choose to purchase an ID Verified Certificate of Achievement, you will be directed to our third party payment processor; and the submission of your payment information will be subject to the terms found at the payment landing page accessible via edX Terms of Service.

INFORMATION WE COLLECT AND HOW WE USE IT

We collect information, including Personal Information (described above), when you sign up for a user account, participate in online courses, register for a paid certificate, send us email messages and/or participate in our public forums. We also collect certain usage information about student performance and patterns of learning. In addition, we track information indicating, among other things, which pages of our Site were visited, the order in which they were visited, when they were visited and which hyperlinks and other user interface controls were used.

We may log the IP address, operating system, and browser software used by each user of the Site, and we may be able to determine from an IP address a user's Internet Service Provider and the geographic location of his or her point of connectivity. Various web analytics tools are used to collect this information. Some of the information is collected through cookies (small text files placed on your computer that store information about you, which can be accessed by the Site). You should be able to control how and whether cookies will be accepted by your web browser. Most browsers offer instructions on how to reset the browser to reject cookies in the "Help" section of the toolbar. If you reject our cookies, many functions and conveniences of this Site may not work properly.

USE AND DISCLOSURE TO THIRD PARTIES

Among other things, we and the educational institutions and other leading global members that provide courses through edX (the "Members") may use the information about you collected through the Site (including your Personal Information) in connection with the following. In cases where we share or disclose your Personal Information: (1) the third party recipients are required to handle the Personal Information in a confidential manner and to maintain adequate security to protect the information from loss, misuse, unauthorized access or disclosure, alteration, and destruction; and (2) we will only disclose and share the Personal Information that is required by the third party to fulfill the purpose stated at the time of collection.

- To enable the Members to provide, administer, and improve the courses.
- To help us and the Members improve edX offerings, both individually (e.g., by course staff when working with a student) and in aggregate, and to individualize the experience and to evaluate the access and use of the Site and the impact of edX on the worldwide educational community.
- For purposes of scientific research, particularly, for example, in the areas of cognitive science and education.
- For the purpose for which you specifically provided the information, for example, to respond to a specific inquiry or provide you with access to the specific course content and/or services you select.
- To track both individual and aggregate attendance, progress and completion of an online course, and to analyze statistics on student performance and how students learn.
- To monitor and detect violations of the Honor Code, the Terms of Service, as well as other misuses and potential misuses of the Site.
- To publish information, but not Personal Information, gathered about edX access, use, impact, and student performance.

- To send you updates about online courses offered by edX or other events, to send you communications about products or services of edX, edX affiliates, or selected business partners that may be of interest to you, or to send you email messages about Site maintenance or updates.
- To archive this information and/or use it for future communications with you.
- To maintain and improve the functioning and security of the Site and our software, systems, and network.
- For purposes described elsewhere in this Privacy Policy (including, e.g., Sharing with Third Parties).
- As otherwise described to you at the point of collection or pursuant to your consent.
- To authenticate your identity when you register for a paid certificate.
- To process refunds, as applicable.

We will share information we collect with the Members, and we and the Members may share this information (including Personal Information) with third parties as follows:

- With service providers or contractors that perform certain functions on our or the Members' behalf, including processing information that you provide to us on the Site, processing purchases via third party providers, and other transactions through the Site, operating the Site or portions of it, providing or administering courses, or in connection with other aspects of edX or Member services.
- With other visitors to the Site, to the extent that you submit comments, course work, or other information or content (collectively, "Postings") to a portion of the Site designed for public communications; and with other members of an edX class of which you are a member, to the extent you submit Postings to a portion of the Site designed for viewing by those class members. We may provide your Postings to students who later enroll in the same classes as you, within the context of the forums, the courseware, or otherwise. If we do re-post your Postings originally made to non-public portions of the Site, we will post them without your real name and e-mail (except with your express permission), but we may use your username without your consent.
- For purposes of scientific research, particularly, for example, in the areas of cognitive science and education. However, we will only share Personal Information about you for this purpose to the extent doing so complies with applicable law and is limited to the Personal Information required to fulfill the purposes stated at the time of collection.
- To provide opportunities for you to communicate with other users who may have similar interests or educational goals. For instance, we may recommend specific study partners or connect potential student mentees and mentors. In such cases, we may use all information collected about you to determine who might be interested in communicating with you, but we will only provide other users your username, and not disclose your real name or e-mail address to your contact.
- To respond to subpoenas, court orders, or other legal process; to investigate, or prevent, or take action regarding illegal activities, suspected fraud, or security or technical issues, or to enforce our Terms of Service, our Honor Code, or this Privacy Policy; as otherwise may be required by applicable law; or to protect our rights, property, or safety or those of others.
- With affiliates of edX or the Members, or with successors in the event of a merger, acquisition or reorganization, for their use consistent with this Privacy Policy.
- As otherwise described to you at the point of collection or pursuant to your consent.
- For integration with third party services. For example, videos and other content may be hosted on YouTube and other websites not controlled by edX.

In addition, we may share aggregated information that does not personally identify you with the public and with third parties, including but not limited to researchers and business partners.

USE FOR PERSONALIZATION AND PEDAGOGICAL IMPROVEMENTS

Our goal is to provide current and future visitors with the best possible educational experience. To further this goal, we sometimes present different users with different versions of course materials and software. We do this to personalize the experience to the individual learner (assess the learner's level of ability and learning style, and present materials best suited to the learner), to evaluate the effectiveness of our course materials, to improve our understanding of the learning process and to otherwise improve the effectiveness of our offerings. We may publish or otherwise publicize results from this process, but, unless otherwise permitted under this Privacy Policy, those publications or public disclosures will not include Personal Information.

LINKS TO OTHER WEBSITES

The Site contains links to websites published by third parties, including other content providers as well as certain service providers, such as those who process payments on behalf of edX. These other websites are not under our control, and you acknowledge and agree that we are not responsible for the collection and use of your information by such websites, unless we have otherwise indicated in this Privacy Policy. We encourage you to be aware when you are be redirected to a third party website and review the privacy policies of each website you visit and use.

SECURITY

EdX has a program designed to protect Personal Information in its possession or control. This is done through a variety of privacy and security policies, processes, and procedures, including administrative, physical, and technical safeguards that reasonably and appropriately protect the confidentiality, integrity, and availability of the Personal Information that it creates, receives, maintains, or transmits. Nonetheless, no method of transmission over the Internet, or method of electronic storage, is 100% secure; and therefore, edX cannot guarantee its absolute security.

CHANGING OUR PRIVACY POLICY

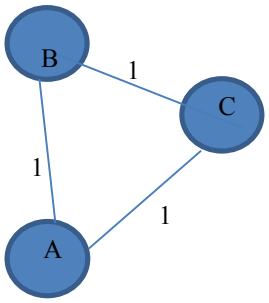
Please note that we review and may make changes to this Privacy Policy from time to time. When changes are made, the Privacy Policy link will include a notation "Revised (date)." indicating that you should review the new terms, which will be effective immediately upon posting on this page, with an updated effective date. The notation will be included in the Privacy Policy link for at least seven (7) days after the update has been made. By accessing the Site after any changes have been made, you signify your agreement on a prospective basis to the modified Privacy Policy and any changes contained therein. In case you miss the notification referenced above, be sure to return to this page periodically to ensure familiarity with the most current version of this Privacy Policy.

PRIVACY CONCERNS

If you have privacy concerns, have disclosed data you would prefer to keep private, or would like to access the information we maintain about you, please contact us at privacy@edx.org.

Effective Date: October 22, 2014

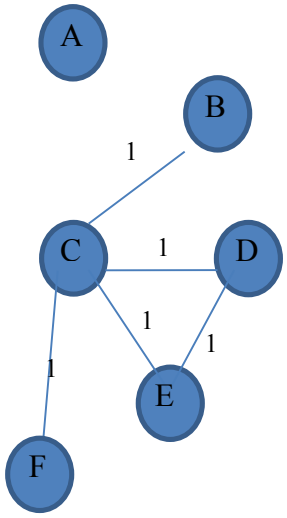
Appendix C Examples of Network Construction

How it appeared on the forum	How it was reprented in network construction	Graph representation
1) Non-Manipulated Thread: Typical discussion thread		
1st order comment: *Topic: Can not enter values into dotted box:* **Person A:** Hi, I need your help please. I am working on Assessment trying to assembling the right formula, but the dotted boxes are not functioning. Is there any other way to do this exercise? Please advise. Thanks (username)	A, 1st order thread id-1	
2nd order comment. Person B: I believe you have to match the parts of the function separated by the commas to the references in the strip below. You can then drag them to the correct sequence.	B, 1order thread id-1	
3rd order comment. Person B: Ok reverse that – you have to drag the cell references.	B, 1st order thread id-1	
3rd order comment. Person B: They don't make these things very intuitive.	B, 1st order thread id-1	
2nd order comment. Person A: Hi (username B), thanks for your reply and the info but not being able to complete this part on the screen, so I have saved it in Dropbox	A, first order thread	
2nd order comment. Person C: You just need to drag your answers to the right spot	C, first order thread	

2) Manipulated Thread: Example of an Exception
(example below continues to include 124 more posters)

1st order comment:
Topic: [Official]Week 2: HW Assignment 2.5

A; 1st order thread id



Person A: Here you can post your questions about homework assignment HW 2.5. We really appreciate students helping other students out, but please try not to reveal the answer in doing so.

2nd order comment. Person B: I'm stuck on this problem, I get a very high amperage value of the order of hundreds of amperes, and maybe it's right. Then for the voltage I have any idea to calculate the total intensity, because it should be zero, but I have problems to calculate the electron density because we have two different and which is not used.

B, 2nd order thread id -1

> ***3rd order comment. Person C*:** Hello, pay a bit more attention to the problem description for the 2.5.b

C, 2nd order thread id-1

2nd order comment. Person D: I'm stuck here too. In point (a): should I introduce the sign of the answer? In point (b): should I estimate n like Integrate[dn/dx * dx] in 2um = delta n ?

D, 2nd order thread id-2

> ***3rd order comment. Person C***: no sign included in point a. in point b there is no need to estimate the n, the n is given already in the problem description. You just need to think, according to problem description which of the n values provided (lower density or higher density of electrons) you should choose.

C, 2nd order thread id-2

> ***3rd order comment. Person E***: Thanks. I had the same questions.

E, 2nd order thread id-2

2nd order comment. Person F: to solve part b), are we to set Je(diff) = Je(drift) and solve for E (electric field)? the only question i now have is for what value of 'n' to use. the two values are only different by a factor of ten.

F, 2nd order thread id-3

> ***3rd order comment. Person F:*** to answer my own question: yes, and there only 2 choices for 'n'.

F, 2nd order thread id-3

> ***3rd order comment. Person C***: You are right, but in a more precisely way Je(diff) + Je(drift) = 0 So Je(diff) = - Je(drift) [But introduce them without sign] Lets take this direction for dn/dx "more electrons -> less electrons" Then you need to apply an electric field that moves the electrons in the other way, from less to more
> .

C, 2nd order thread id-3

3) Manipulated Thread: Typical Introductory Thread
(example below continues to include over 700 posters)

1st order comment:
Topic: The one and only place for Excelfies

Person A: Hi students, Me, myself and Excel! Looking forward to seeing all of your spread-sheeting tools! Please scroll down and post your Excelfie as a response, do not click the "New Post" button above :) Attachment: Picture of the person posting

2nd order comment. Person B: Coffee and Excel. This is how we drink our Coffee in Kuwait. Arabic style :D Attachment: Picture of the person with the laptop and a coffee

> ***3rd order comment.*** **Person C**: With sugar or sketo? Love the cup. Looking forward to the work here, hope you enjoy it as well.
> ***3rd order comment.*** **Person D**: Sheikh l'shabeb !

2nd order comment. **Person E**: Hi There, A lot of excel sheet under my brain ☺ Attachment: A screenshot of a spreadsheet.

2nd order comment. Person F: Hi everyone! This is me and my daily work full of figures, pivots and numerous calculations. I never feel bored with this stuff. Attachment: Picture of the poster with the laptop.

> ***3rd order comment.*** **Person G**: good work [username F]

A; 1st order thread id

B; 2nd order thread id-1

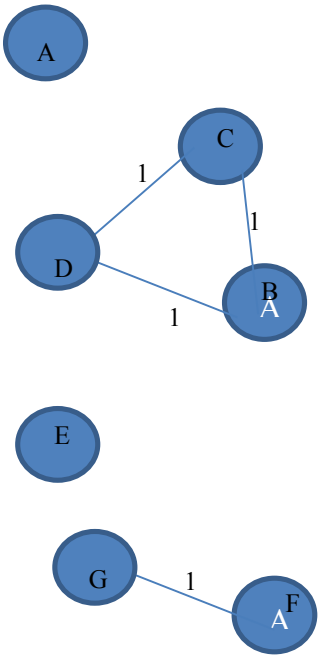
C, 2nd order thread id-1

D, 2nd order thread id-1

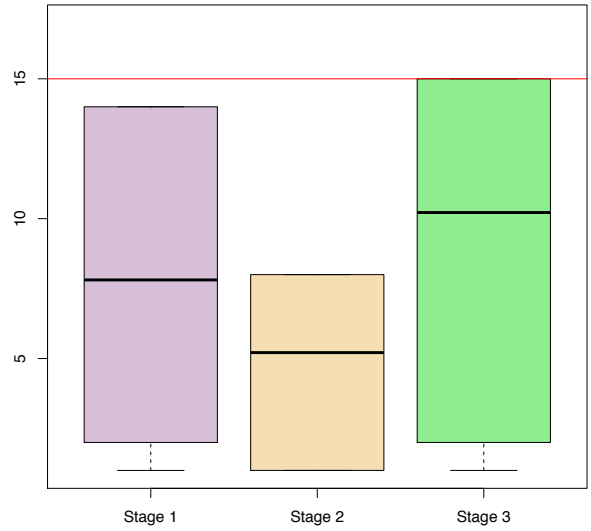
E, 2nd order thread id-2

F, 2nd order thread id-3

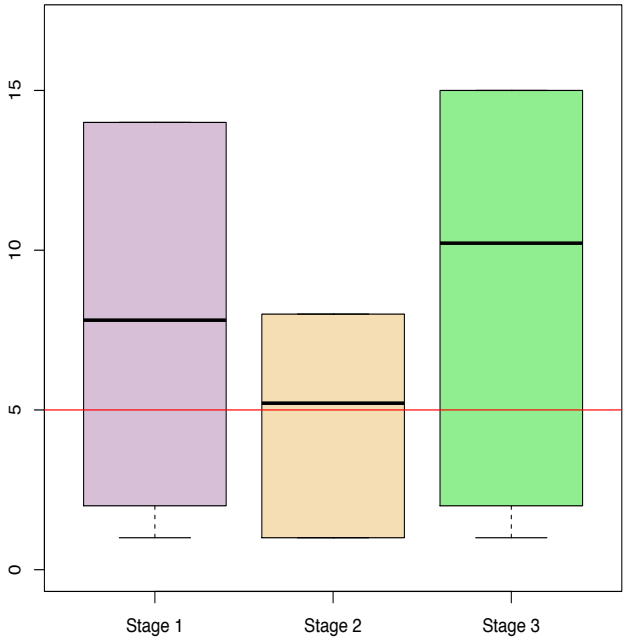
G, 2nd order thread id-3



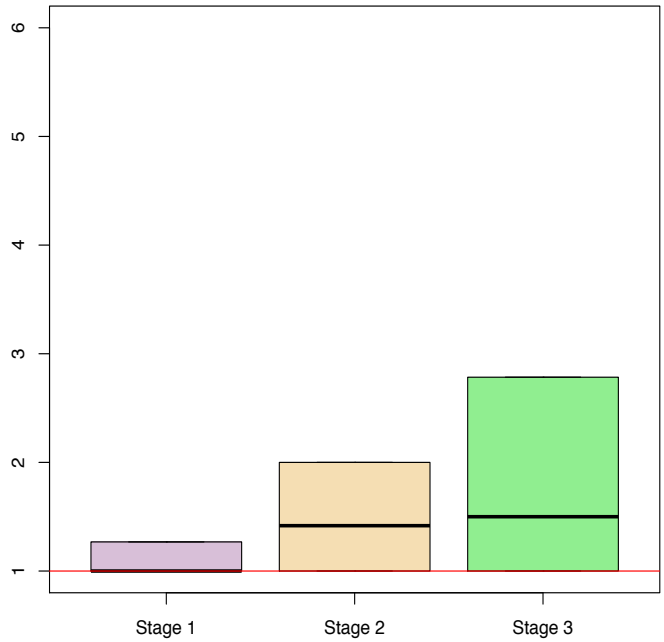
Appendix D Accumulation of edge and degree over time



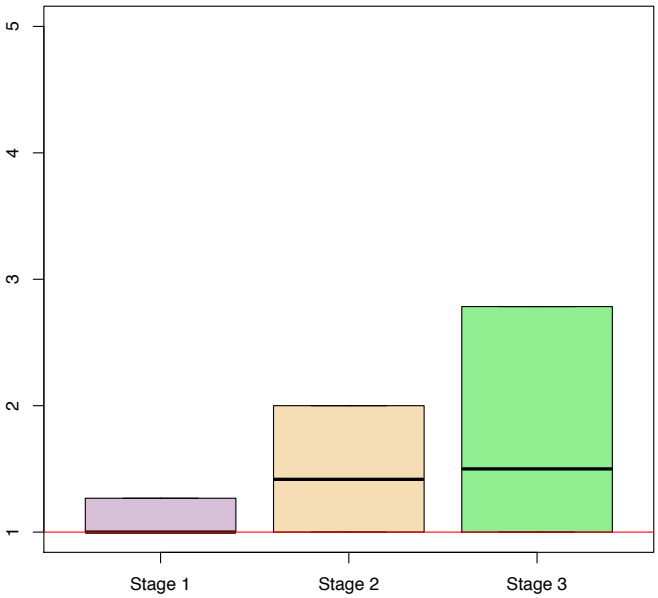
a) Degree Distribution Over Time in Water Forum Core



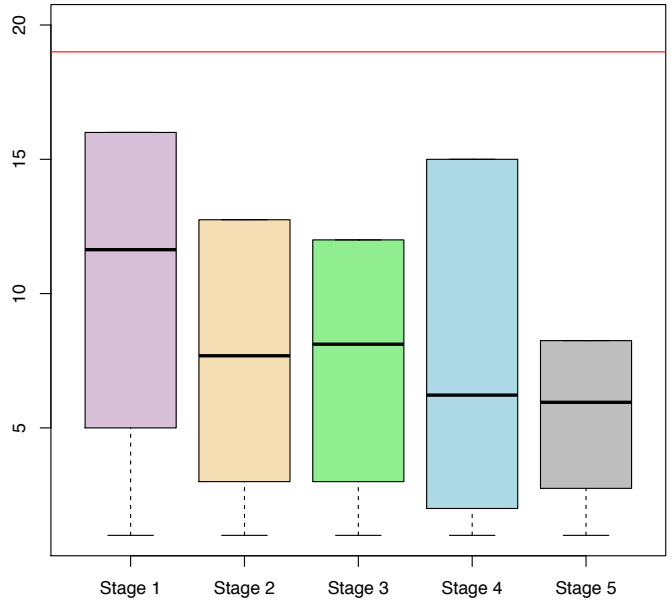
a) Degree Distribution Over Time in Water All-Posters Network



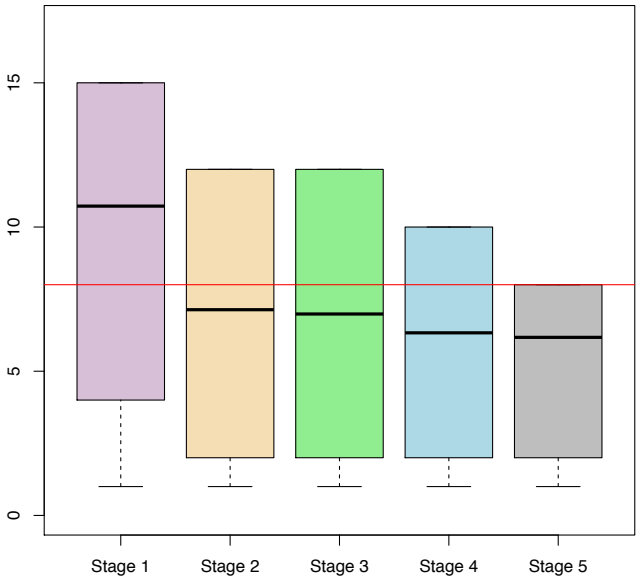
c) Edge Weight Distribution Over Time in Water Forum Core



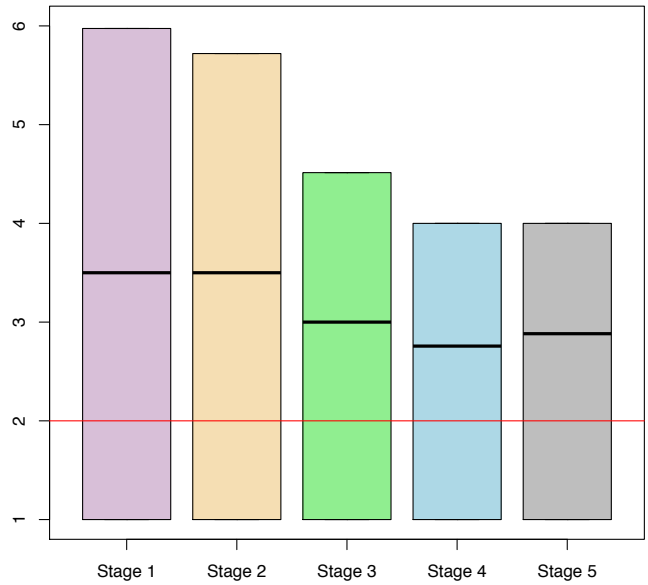
d) Edge Weight Distribution Over Time in Water All-Posters Network



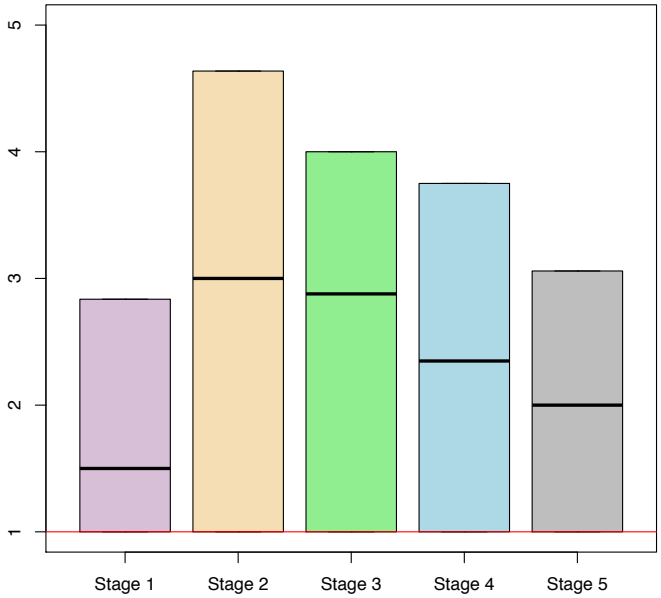
e) Degree Distribution Over Time in FP Forum Core



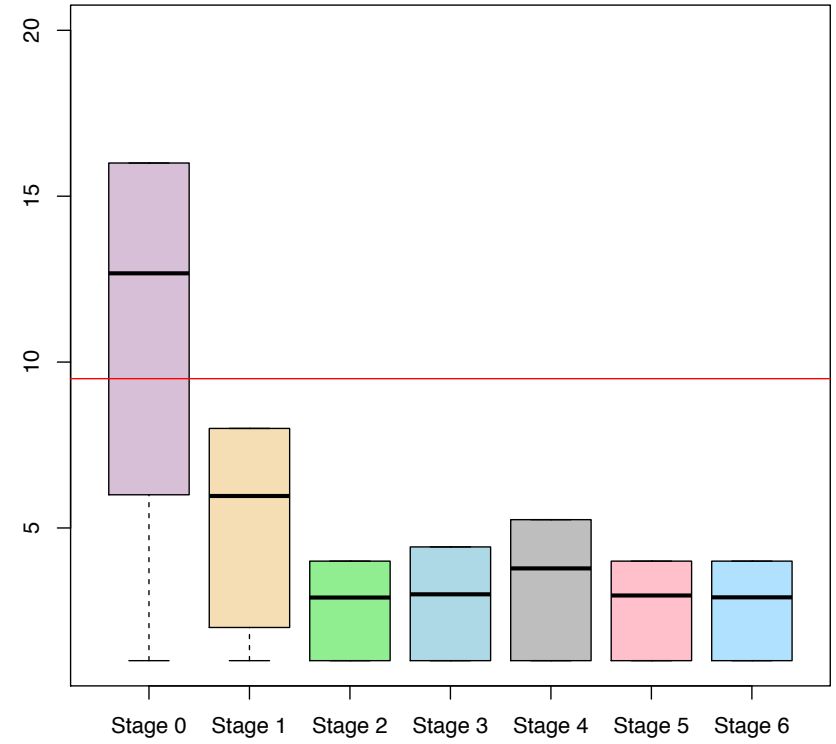
f) Degree Distribution Over Time in FP All Posters Network



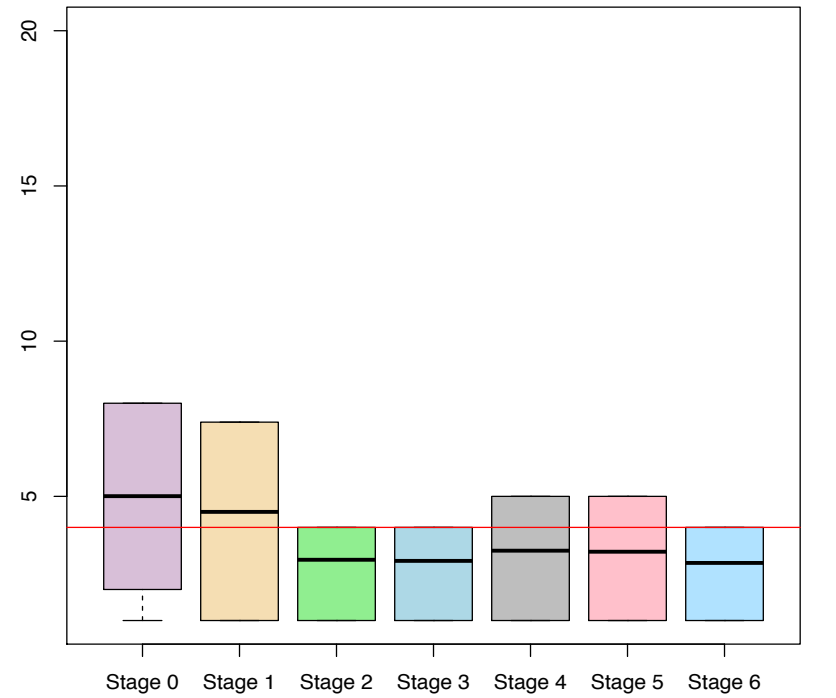
g) Edge Weight Distribution Over Time in FP Forum Core



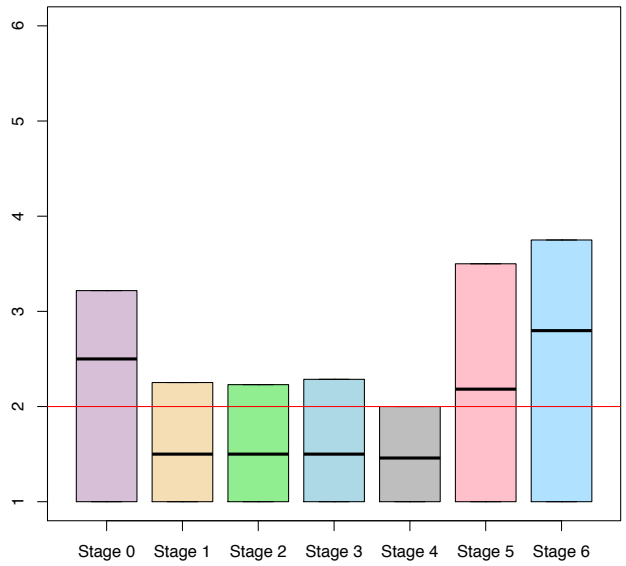
h) Edge Weight Distribution Over Time in FP All Posters Network



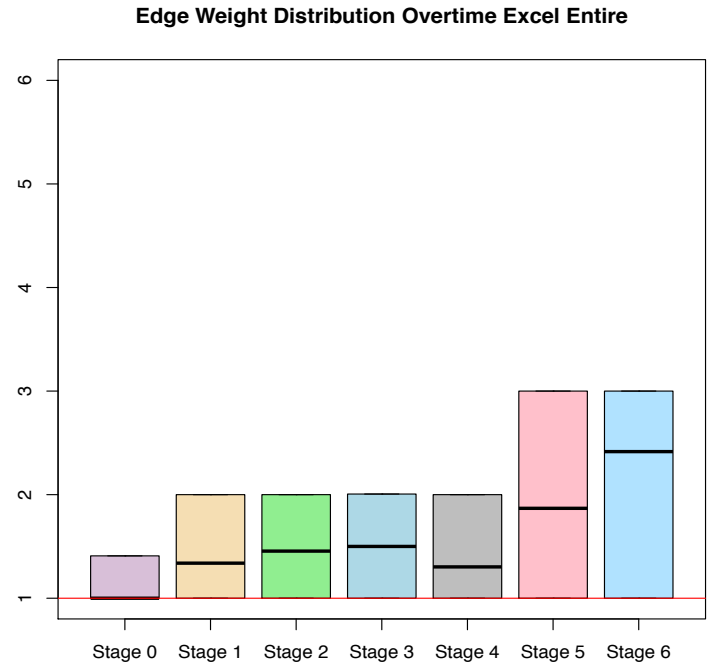
l) Degree Distribution Over Time in Excel Forum Core



j) Degree Distribution Over Time in Excel All Posters Network



k) Edge Weight Distribution Over Time in Excel Forum Core



l) Edge Weight Distribution Over Time in Excel Forum Core

Appendix E Coding Framework

CURRICULUM-RELATED CONVERSATIONS			
Label	**Keywords**	**Description**	**Example**
Content Task (CT)	Explanation OR not clear OR clarification OR confused OR frustrated AND Task OR homework OR exam OR assignment OR quiz	A learner is asking for help or clarification because the learner is not clear about something in the exam/ homework/ in the assignment/ in the lab/ in the quiz. The learner is discussing a calculation in the task or a curriculum concept from the course task (explicitly refers to the course task). The learner is referring to the mistake in the exam, homework or quiz. The learner may be referring to his emotional state in relation to the task, but the conversation purpose is to ask for help with task solution.	1) Does any one got answers to efficiency problems of exam 1 problems (6,7) i am getting some errors in % value 2) Without any Up converter: only photons 2 eV & above are absorbed which is applicable for photons around 620nm and below. All calculations as usual & let's assume current density Jsc1 With Up converter 1: 2 photons (one from B, another form C) add up @ 65% success to make 1 photon with > 2eV and you get your Jsc2' but the total Jsc2 is the sum of Jsc1 & Jsc2' 3) Are the answers for the questions e,f and g wrong ? cause i am getting the answers in 10^-6 range after applying the formula discussed in week 5. Please illuminate someone. Thank u
Content Non-task (CN)	Explanation OR not clear OR clarification, OR confused OR frustrated **AND** The name of the concept, issue or practical invention related to the subject matter **BUT NOT** Task OR homework OR exam OR	A learner is asking for help or clarification because the learner is not clear about something in the lecture. The learner is discussing a calculation in the lecture The learner is discussing a subject matter concept but not from the course task (does not mention a course task) The learner is referring to the mistake in the lecture The learner is discussing his hobby or something he did that is related to the subject, e.g. personal solar system, cooling	1) The **PbS** Band Gap of 0.37 eV is even smaller, but we know how **poisonous** it can be to animals, including Humans (damaging the nervous system and disturbing the brain)…so it is not a good option for our future…..or for the past….. 2) also in (6) since V=0 there is only a current divider and there is no negative term in the expression. I dont get what am I missing! THX 3) **Lecture 2.6 - WHy shouldn't n & p regions not be electrically neutral ?** In the above lecture (at time 3.42 secs)

	assignment OR quiz	water to get rid of the smell, tricks for excel spreadsheet operation, books about programming. The learners are discussing subject related topics that are outside of the course's content, for example, politics, engineering, which batteries to buy, etc.	the prof. says that n & p regions are not electrically neutral. Why? Since doping atoms are electrically neutral (all atoms are), and they become positively charged, after adding to the base material, by giving away an electron, shouldn't the n region as a whole be electrically neutral ? Same for p region. Please explain
SELF-RELATED CONVERSATIONS			
Label	**Keywords**	**Description**	**Example**
Social Task (ST)	Frustrated OR Confused OR Unfair OR wrong OR difficult OR easy OR challenging Focus on emotional support rather than content-related solution	The post reflects the learner's reflection on his ability to do the task. The conversation **does not focus on the request for task solution**, but on the emotion experienced while dealing with the task, as well as self-reflection on learners ability to cope. The learner is complaining about something. The learner is frustrated and the rest are offering support. The learner is insinuating that wants to quit or does not know how to continue. The learner is venting about the difficulty of the task/course/subject matter, and the rest are responding In the conversation the happiness or excitement about the task or question or subject matter is expressed	1) A: So this first lecture came as kind of a shock to me. There was a lot of new expressions and exotic material names dropped, that I have no previous knowledge about. B: Just as Lasciel mentioned, the III-V devices are indeed photovoltaic devices made from the 3rd and 5th group of the periodic table. But don't worry, this group of solar cells (and other materials) will be discussed extensively in the upcoming weeks. 2) I don't have a PV system and I don't know anyone with one!

SETTING-RELATED CONVERSATIONS			
Technical and administrative issues (INF)	**Administrative question** I joined the course late, can I still do it; can't find syllabus. **IT(system)-related problem:** progress bar, where is the video, my scores do not show well; how do I download videos, how to turn on the subtitles, can we have pdfs of formulas, youtube does not work in my country; submitting assignment (typing formula, rounding, etc.)	Student is experiencing what he believes to be a computer problem that does not allow to complete homework or watch the lecture. Student wants to perform a technical operation, i.e. download, annotate, submit, and something does not work. Student is confused as to where to find course materials. Student is confused about course grading or progress bar. Student is not sure whether can still join the course and is asking whether course schedule and requirements would still allow. Student is confused about course deadlines.	1) I've done a couple of the practice questions throughout the chapter 1 and only received points for 1 question. The rest are still 0/0. I had 2 questions that I knew I got wrong (for the cost/kWh chart question that got me confused) but I had the other previous questions correct. Why did I not receive credit for them? 2) These subtopics cover a lot more of the subject,but I am unable to download. Kindly let me know how to download slides under Lecture. In the absence of that, kindly add that powerful feature, which would help. Good day! S.E.Kannan 3) I recently discovered this class and therefore joined late. Therefore, I missed the deadline for the 1st and 2nd assignment. If I take the Exam 1 and submit all assignments and exams, thereafter, would I still be able to get a certificate for this class?
Social Non-task (SN)	Rules, norms, power negotiation, challenge the authority, setting common ground (including introductions); present or discuss expectations, offers perspective on values of learning and teaching	Students introduce themselves or where they are from. Students challenge the status quo of the course, e.g. number of test attempts, the amount of feedback from staff, and discusses why it is a problem. Student challenges the status quo of the course, and discusses what it should be but does not	A: Hi there. Most probably the designers of edX have thought of all the possibilities mentioned above. But remember that the courses are trying to meet higher levels of education, wich mean higher commitment, atention to details and self-discipline. Please, please, read the guidelines for this course (and many others) in this link: [Collaboration Guidelines) B: I have done the MITx 6.002x course last year - the

		offer a constructive solution. Student tells another student not to do something, e.g. not to reveal the answer, not to hack the system, etc.	original pilot course. On that course assessed questions were allowed three attempts to get the answer and non-assessed questions allowed (seemingly) infinite attempts. Problems did arise on that course where varying the degree of precision in an answer course cause one answer to be accepted but not another. I therefore see legitimate concerns being expressed by Darkbound. C: Oh, and as a cautionary warning to anyone thinking of setting up duplicate accounts - they were spotted on the MITx 6.002x course and the offenders didn't get to complete the course. The honour code needs to be honoured!
OTHER CATEGORIES			
Label	**Keywords**	**Description**	**Example**
Meta-cognitive (META)	A suggested strategy on how to learn best in the MOOC	Student offers a constructive solution, e.g. downloads the formulas, pdfs them and offers to others. Student shares her strategy of learning from the lectures. Student shares his or her resources about how to best find resources when he does not know the answer. Student shares his strategy for preparing to the exam and gives out study tips.	1) EDIT: You can now download lecture slides from the 'Courseware' tab by clicking 'Slides' button below the streaming video. They are *not* annotated. I kinda hoped staff would upload this... That didn't happen. Anyway...All PDFs in one .RAR file:LINK. Individual lectures (click 'Download' button to download) **2)** A: Your suggestion has been noted, and I'll pass it on to staff as I think it's a good one.They are REALLY busy lately though, so I really can't promise anything. Maybe if you took notes? I always learn better when extracting notes by myself. B: Each week, I create 20 pages from personal notations.... C: Taking notes really helps us in understanding the concepts. i think i had beat you in this one as it comes around 30 pages for me each week except for the first week. ;)
Mixed (MIX)	You can't determine which category to which the conversation falls under because its focus overlaps across categories. A student is both asking for help in the task and venting looking for emotional support.		

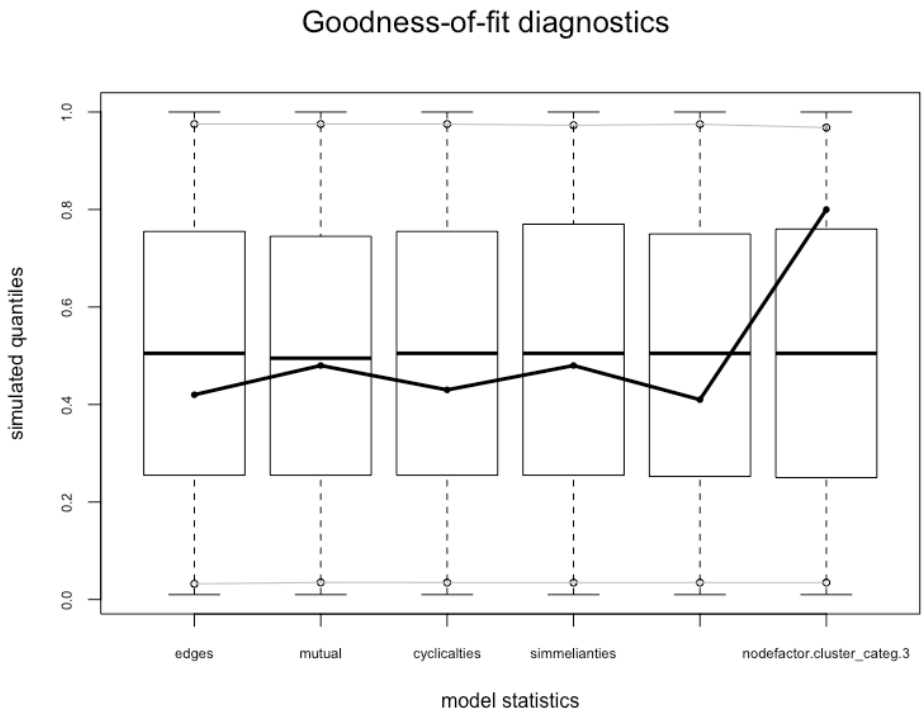
Appendix F K-Means Clustering of Network Actors

Size of clusters and measures for representative actors are summarised below.

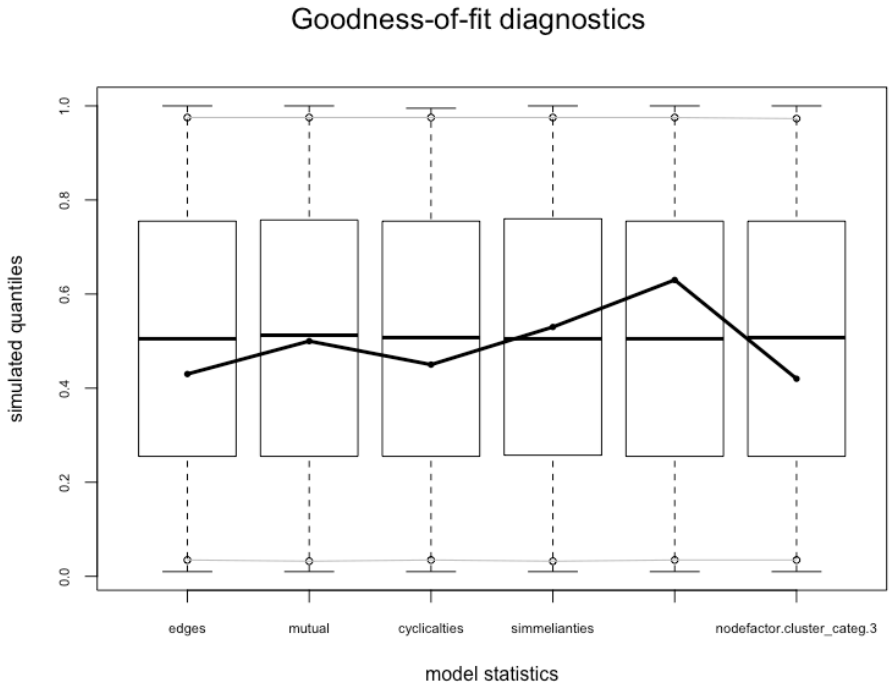
Course	Water	Solar	Excel	FP
Cluster 3 - High Posting Activity				
Posters, N	**6**	**7**	**3**	**4**
In-reach Centroid	6045	18797	14630	69788
Out-reach Centroid	7808	22576	12994	81533
Cluster 2 - Moderate Posting Activity				
Posters, N	**273**	**40**	**38**	**41**
In-reach Centroid	655	4484	2054	16114
Out-reach Centroid	710	6116	2864	21084
Cluster 1 - Low Posting Activity				
Posters, N	**2998**	**4082**	**1994**	**1317**
In-reach Centroid	32	170	65	743
Out-reach Centroid	35	159	58	619

Appendix G

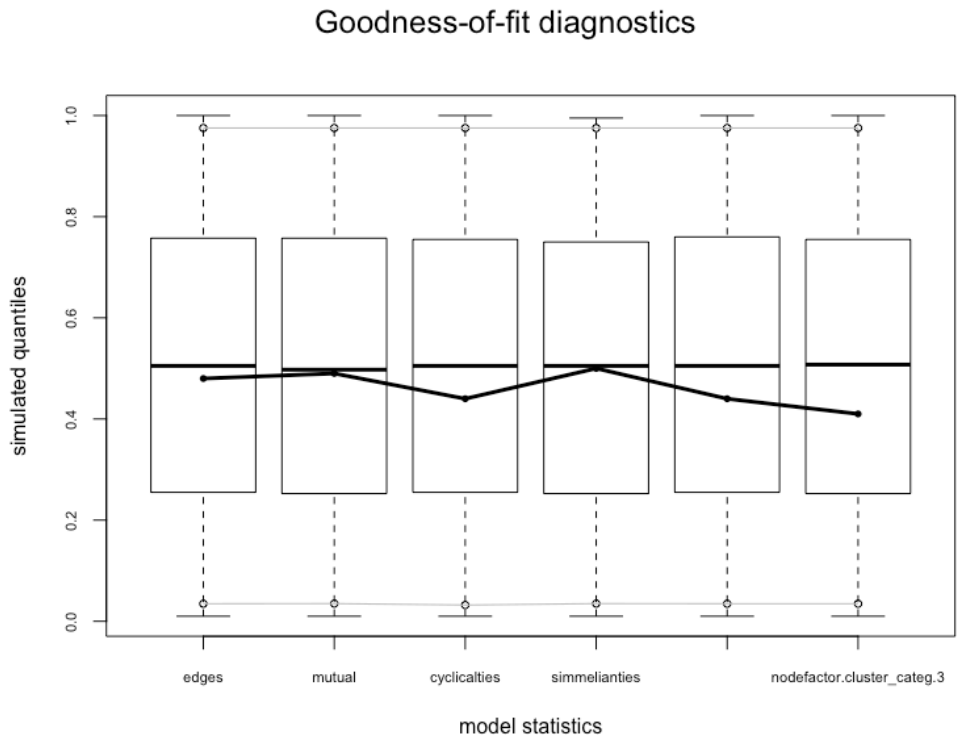
a) Goodness of fit for FP Model



b) Goodness of fit for Excel Model



c) Goodness of fit for Solar Mode



d) Goodness of fit for Water Model

