This is the peer reviewed author accepted manuscript (post print) version of a published work that appeared in final form in:

Ferreira, Máverick André Dionísio, Ferreira Mello, Rafael, Kovanovic, Vitomir, Nascimento, André, Lins, Rafael, Gasevic, Dragan 2022 'NASC: network analytics to uncover socio-cognitive discourse of student roles', ACM International Conference Proceeding Series, pp. 415-425

This un-copyedited output may not exactly replicate the final published authoritative version for which the publisher owns copyright. It is not the copy of record. This output may be used for non-commercial purposes.

The final definitive published version (version of record) is available at:
https://doi.org/10.1145/3506860.3506978

Persistent link to the UniSA Research Outputs Repository:
https://researchoutputs.unisa.edu.au/11541.2/28696

General Rights

Copyright and moral rights for the publications made accessible in the UniSA Research Outputs Repository are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the persistent link identifying the publication in the UniSA Research Outputs Repository

If you believe that this document breaches copyright, please contact us and provide details. We will remove access to the work immediately and investigate your claim.

# NASC: Network analytics to uncover socio-cognitive discourse of student roles

Máverick Ferreira
Federal University of Pernambuco
Recife, PE, Brazil
amaverick70@gmail.com

Rafael Ferreira Mello
Cesar School
Recife, PE, Brazil
Federal Rural University of Pernambuco
Recife, PE, Brazil
rflm@cesar.school

Vitomir Kovanović
University of South Australia
Adelaide, Australia
Vitomir.Kovanovic@unisa.edu.au

André Nascimento
Federal Rural University of Pernambuco
Recife, PE, Brazil
andre.camara@ufrpe.br

Rafael Dueire Lins
Federal Rural University of Pernambuco
Recife, PE, Brazil
rdl.ufpe@gmail.com

Dragan Gašević
Monash University
Melbourne, VC, Australia
University of Edinburgh
Edinburgh, U.K.
King Abdulaziz University
Jeddah, Saudi Arabia
Dragan.Gasevic@monash.edu

## ABSTRACT

Roles that learners assume during online discussions are an important aspect of educational experience. The roles can be assigned to learners and/or can spontaneously emerge through student-student interaction. While existing research proposed several approaches for analytics of emerging roles, there is limited research in analytic methods that can i) automatically detect emerging roles that can be interpreted in terms of higher-order constructs of collaboration; ii) analyse the extent to which students complied to scripted roles and how emerging roles compare to scripted ones; and iii) track progression of roles in social knowledge progression over time. To address these gaps in the literature, this paper propose a network-analytic approach that combines techniques of cluster analysis and epistemic network analysis. The method was validated in an empirical study discovered emerging roles that were found meaningful in terms of social and cognitive dimensions of the well-known model of communities of inquiry. The study also revealed similarities and differences between emerging and script roles played by learners and identified different progression trajectories in social knowledge construction between emerging and scripted roles. The proposed analytic approach and the study results have implications that can inform teaching practice and development techniques for collaboration analytics.

## CCS CONCEPTS

• **Applied computing** → **Education**; **Distance learning**; • **Social and professional topics** → **Computing/technology**.

## KEYWORDS

Emerging Roles; Epistemic Network Analysis; CSCL; Clustering Analysis

**ACM Reference Format:**
Máverick Ferreira, Rafael Ferreira Mello, Vitomir Kovanović, André Nascimento, Rafael Dueire Lins, and Dragan Gašević. 2022. NASC: Network analytics to uncover socio-cognitive discourse of student roles. In *LAK22: 12th International Learning Analytics and Knowledge Conference (LAK22), March 21–25, 2022, Online, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3506860.3506978

## 1 INTRODUCTION

Asynchronous online discussions have been widely used to in computer-supported collaborative learning (CSCL) settings [9]. They promote social interactions and knowledge construction, which are essential aspects of online learning environments. In this line of research, the literature emphasizes a need to provide automatic methods to understand the effectiveness of pedagogies that seek to promote social knowledge construction through online discussions [6, 43]. This understanding can help instructors inform future decisions about pedagogical approaches that aim to support large numbers of students in online discussions [19].

Roles that learners assume during discussions are an important aspect of educational experience and have attracted much attention in the literature [6, 11, 18, 31, 54]. Roles in educational settings can be defined as groups of students with similar behaviour exhibited during their participation in a group activity [4]. In the case of online discussions, there are two categories of student roles [54]: (i) *emerging*, when the student assumes a role intrinsically during an online discussion, and (ii) *scripted*, when the moderator pre-defines

the role for each student. Such roles could influence student participation and consequently impact the social knowledge construction, group collaboration and cohesion during an online discussion [48].

Analytics of roles students take in online discussions can offer useful insights that can inform teaching practice [6, 18, 31]. However, to effectively provide analytic insights for instructors, the interpretation of students' roles should be connected to relevant learning designs and/or educational theory [18]. Also, the literature on student roles suggests that social and cognitive aspects of discussion should be considered [6, 18, 54]. In this context, the Community of Inquiry (CoI) model is a well-established framework that outlines how asynchronous online communication shapes student learning and their cognitive development [14]. The CoI model proposes three different constructs (known as presences) to explain educational experience through social interactions [12]: social presence, teaching presence, and cognitive presence. These presences explain the process of student inquiry through social interactions in an online community of learners [12], recognise the role of social context to create a productive community of learners [17, 42], emphasise the importance of teaching for the development of an effective community of learners [16, 19], and could provide information on how students participation is associated with the development of online communities [25, 28].

There is a large body of research that examined associations of the CoI model [12] and script theory of guidance [11] with student engagement, knowledge construction, and learning outcomes [6, 15, 21, 43]. Previous research has explored how scripted roles could mould social and cognitive presences in a CoI [19, 43]. Nevertheless, there is limited research that has investigated the impact of *emerging roles* on the development of social and cognitive presences in a CoI. A further investigation on this topic could (i) reveal new knowledge that advances the understanding how presences develop in a CoI and (ii) provide a analytic approach to inform teaching decisions.

This paper proposes a network analytic approach called Network Analytics for Socio-Cognitive (NASC) for analysis of online discussions based on social and cognitive presences of the CoI model. The NASC approach is based on epistemic network analysis [52] and allows for automatic detection of emerging roles, comparison of emerging roles with scripted roles, and tracking development of social knowledge construction over time. The approach was validated in a study that used scripted roles for guiding students to develop social and cognitive presence in asynchronous online discussions.

## 2 THEORETICAL BACKGROUND

### 2.1 The Community of Inquiry Model

The Community of Inquiry (CoI) model aims to describe the processes of social knowledge construction in asynchronous online discussions [15]. The CoI model defines three interdependent dimensions (called presences) that together shape students' online learning experience.

The central construct of the model is *cognitive presence*, that describes students' process of critical thinking and inquiry-based learning. It is formally defined as "the extent to which the participants in any particular configuration of a community of inquiry are able to construct meaning through sustained communication" [13, p. 89]. Operationalised within the practical inquiry model [12], cognitive presence consists of four interconnected phases of inquiry-based learning cycle: (I) **Triggering event** phase starts the learning cycle and is initiated by a problem or dilemma that initiates practical inquiry cycle; (ii) **Exploration** phase includes information sharing and investigation and debating of different ideas and potential solutions; (iii) **Integration** phase, where students connect and synthesise relevant ideas and information to construct new knowledge; and finally, (iv) **Resolution** phase, which includes the application of new knowledge to the originating problem, typically triggering a new learning cycle in the process.

Another critical construct within the CoI model is *social presence*, that captures social climate and interpersonal relationships within the course [45]. It is formally defined as "the ability of participants in a community of inquiry to project themselves socially and emotionally, as 'real' people (i.e., their full personality), through the medium of communication used" [13, p. 94]. Social presence consists of three sub-dimensions that capture different aspects of social interactions. The *affective* dimension capture the expression of students' real emotions and feelings map within the learning community. The *interactive* dimensions captures the open and interactive nature of student communication. Finally, *group cohesion* captures the sense of union and group commitment among students in the learning community. A meta-analysis reported in [42] showed a strong positive correlation of social presence with student satisfaction and perceived learning, while social presence has also also found to be positively associated with academic performance with role scripting as a moderator of the association [22].

Finally, *teaching presence* relates to the instructors' role before and during the course, and is formally defined as "the design, facilitation, and direction of cognitive and social processes for the purpose of realizing personally meaningful and educationally worthwhile learning outcomes" [1, p. 5]. It consists of the three components, namely, *design and facilitation* of online learning experience, discourse *facilitation* and *direct instruction*, when necessary. Teaching presence is crucial for the development of both cognitive and social presence. Similarly, social presence mediates the relationship between teaching and the cognitive presences [17, 27], providing necessary environment for providing foundation for the effective inquiry-based learning experience. The use of script roles is also used as a form of teaching presence in communities of inquiry[19].

### 2.2 Emerging and Scripted Roles in Online Discussions

Asynchronous discussion is a key element to support collaboration in online environments. However, a spontaneous interaction of students in online discussions does not necessarily lead to a high level of social knowledge construction, social presence, and cognitive development [11, 12]. Instead effective approaches to guiding learners in productive online discussions are necessary. In this context, the script theory of guidance [11] gained significant attention in the literature. From the perspective of the CSCL research, the script theory of guidance posits that productive student–student discussions can be supported and high-level of social knowledge construction achieved through scripting and assigning roles to students [11, 54]. According to Strijbos and Weinberger [54, p. 491] roles are "more

or less stated functions or responsibilities that guide individual behavior and regulate group interaction," while *scripted roles* are intentionally "designed to improve both learning processes and outcomes" [p. 492]. Role assignment is a type of collaboration script central to the script theory, and has been extensively studied in the CSCL literature as an effective approach to increasing the level of knowledge construction, cognitive processing, and argumentation [5, 43, 48]. Role assignment and scripting are also used in research on CoI to as a way to guide students to achieve high levels of social and cognitive presence [19].

In contrast to scripted roles, emerging roles arise spontaneously from interpersonal exchanges without prior guidance, capture aspects that are not predefined at the beginning of the group work [54]. Emerging roles are also important to promote learner agency given that scripted roles can be seen to promote compliance [57] and even be demotivating [40], in spite of being beneficial for learning [40]. Developments in learning analytics enabled for automated detection of emerging roles from the analysis of digital traces recorded in online discussion forums [32, 49]. Existing approaches are based on natural language processing, social network analysis, clustering algorithms and epistemic network analysis [6, 18, 31, 43, 47]. For instance, Dowell and Poquet [6] proposed an approach that combines automated content analysis and social network analysis to identify emerging roles in online discussion. Gašević et al. [18] suggested an approach that combines techniques of social and epistemic network analysis with techniques for automated content and cluster analysis to study emerging roles through the lenses of social and cognitive dimensions of online collaboration.

## 2.3 Research Questions

The automatic detection of emerging roles can provide insights into the extent to which roles that the instructors envisioned in their learning designs indeed hold [6, 18]. As such, analysis of emerging roles can inform decisions about future changes in learning design. However, at this stage there has not been sufficient research that looked specifically at the emergence of roles in online discussions based on high-order constructs such as cognitive and social presence. Analysis of emerging roles have so far been done based on lower-level text analysis in combination with cluster analysis [6], or measures of social networks analysis [18, 31] and manually coded messages for higher-order constructs [4, 54] but without focusing on analytic methods that aim to identify patters of participation through automated analysis of higher-order constructs of online group activity to identify emerging roles. In this context, we posit that social and cognitive presences of CoI, as higher order constructs of online participation, could provide a solid socio-cognitive conceptual foundation for analysis of emerging roles in online discussions. As such, the first research question addressed in the current study was:

> **RESEARCH QUESTION 1:** *To what extent can emerging roles students occupy in an asynchronous online discussion be identified through automated analysis of indicators of social and cognitive presences?*

If emerging roles can be automatically detected, it would be useful to determine how they relate to scripted roles to inform instructional decisions and potentially provide input for personalised feedback analytics-based tools such as OnTask [37]. While attempts for identification of emerging roles [6, 18, 31, 43, 47] and analysis of the extent to which students comply to the functions of scripted roles [55, 56] are noted, there has been limited literature that proposed data analytic methods to compare emerging and scripted roles especially on the level of higher-order constructs such as relationships between cognitive and social presence. Hence, the second research question was formulated as:

> **RESEARCH QUESTION 2:** *To what extent are scripted roles, which aim at promoting productive online discussions, associated with emerging roles that are identified through automated analysis of indicators of social and cognitive presence?*

Finally, it is equally important to track the progression of learners over time and how they perform certain roles and how these roles progress in their social knowledge construction. Again, this can inform decisions about teaching interventions and even be used to construct analytics-based personalised feedback at scale with tools to help students improve their social knowledge construction. There has been a shortage of work that looked at how emerging and scripted roles progress over time and data analytic methods that can provide analysis of progression on the level of higher-order constructs such as relations between social and cognitive presence. Thus, the third research question was:

> **RESEARCH QUESTION 3:** *To what extent can we track changes in the associations between social and cognitive presences, for each role, over time?*

# 3 METHOD

To answer the proposed research questions, we proposed a learning analytic approach called Network Analytics of Socio-Cognitive discourse (NASC). Figure 1 presents the overall steps that are included in NASC. The process is initiated (Step 1) with the collection of online discussion transcripts from a learning environment. Then, the data is annotated according to the categories of social and cognitive presence (Step 2). Although we used a dataset that was manually coded, the literature presents several approaches for automatic coding of online discussion messages according to the categories of social and cognitive presences [2, 3, 7, 34]. Step 3 entails the use of a clustering algorithm to identify the emerging roles and produces the assignment of students to the relevant emerging roles (Step 4). Student assignments to emerging and scripted roles used to construct epistemic networks and preform epistemic network analysis to compare emerging and scripted roles and to track changes in social knowledge construction for each role over time (Step 5). Finally, the NASC produces results that include information about emerging roles detected automatically, differences between emerging and scripted roles, and the trajectory graphs showing how the roles changed over time.

## 3.1 Discussion Transcript and CoI annotation

The dataset used in the study comprised six offerings (Winter 2008, Fall 2008, Summer 2009, Fall 2009, Winter 2010, Winter 2011) of a fully online master-level research-intensive course in software engineering, with 82 students that posted 1,747 messages. As part of

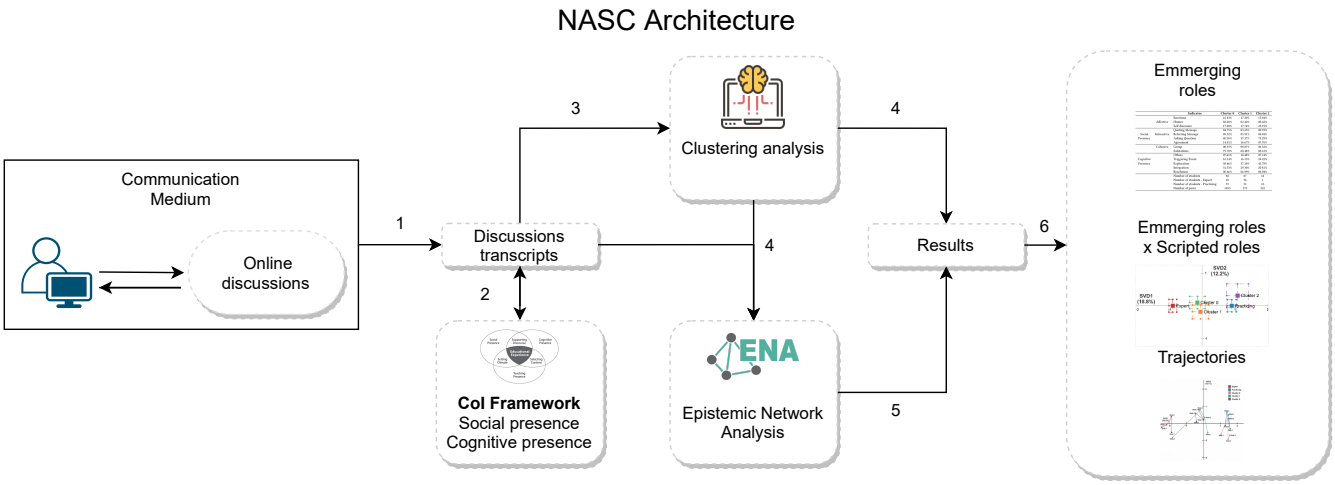



**Figure 1: The description of the proposed analytic approach NASC – network analytic of socio-cognitive discourse**

the course assessment (15% of the final mark), the students were instructed to record a video presentation about one research paper related to one of the course topics, in which the other students would engage in online discussions around their presentation. Within each discussion, the students could assume two scripted roles defined by the instructors: (i) **experts**, the students that posted the video presentation that triggered the discussion; they also had the responsibility to answer questions about the video and moderate the discussion; (ii) **practising researchers**, students who watched the video in order to interact with comments and questions for the experts [1]. It is important to mention that the students were assigned a specific role for each discussion. It means that the same student could assume the expert or practising roles in different discussions. Not all students played both roles. The discussions happened over the four weeks of the course (Weeks 2-5) and the students could assume the expert role in only one discussions while in all other discussions participated as practising researchers.

As already mentioned, the data analysed in this study was manually annotated. Two experts in CSCL and software engineering codded the dataset for both cognitive and social presence using the coding scheme proposed by Garrison et al. [14], Rourke et al. [44]. All online discussion messages were coded according to the indicators and phases of the social and cognitive presences, respectively. The coding unit of analysis was entire message. Each message could have more than one social presence indicator [41], but just one cognitive presence phase [14]. The coders achieved an excellent level of agreement for cognitive presences, reaching a percentage of agreement = 98.1% and Cohen's k = 0.974. The differences (32 messages) were resolved through discussion among coders. Table 1 shows the number of messages in each phase of cognitive presence in the dataset.

For social presence, the final number of indicator codes was 3,770 instead of 1,747 (the total number of messages) due to multiple codes assigned to one message. The coders reached the agreement of 84% of agreement for each indicator of social presence. Table 2 presents

[1]For further details on the course design [19].

**Table 1: Distribution of cognitive presence.**

| Phase | Messages | % |
|---|---|---|
| Triggering event | 308 | 17.63% |
| Exploration | 684 | 39.15% |
| Integration | 508 | 29.08% |
| Resolution | 107 | 6.13% |
| *Other* | 140 | 8.01% |
| *Total* | 1,747 | 100.00% |

the distribution of the indicators of social presence after the coding process. We omitted some of the social presence indicators (i.e., Continuing a thread, Complementing, and Vocatives) in our analysis as they had a disproportionately large number of messages with such codes. It follows the guideline used in previous work [26] and avoids the problem of a dominant code in epistemic network analysis [10].

## 3.2 Cluster Analysis

The current study adopted a cluster analysis to identify students' emerging roles during online discussions (RQ1). Clustering algorithms are unsupervised machine learning techniques applied to identify relevant subgroups with similar patterns within an extensive data collection [58]. Cluster analysis has been widely used to analyse online discussions under different perspectives [9, 25], including analysis of student roles [6, 24, 30]. Among the possible clustering algorithms, several previous studies in educational settings have adopted the k-means algorithm, with the use of the euclidean distance, to uncover patterns in online discussions [6, 9, 30]. K-means usually reaches a good performance with a small processing time [58]. However, k-means requires the number of clusters (k) as an input parameter. The literature recommends the use of the silhouette coefficient to determine an optimal number of clusters [20, 53]. The silhouette coefficient is generally evaluated with a

**Table 2: Social Presence Indicators**

| Category | Indicator | Count | Percent Agreement |
|---|---|---|---|
| Affective | Expression of emotions | 288 (16.5%) | 84.4 |
| | Use of humor | 44 (2.52%) | 93.1 |
| | Self-disclosure | 322 (18.4%) | 84.1 |
| Interactive | Continuing a Thread | 1664 (95.2%) | 98.9 |
| | Quoting from others messages | 65 (3.72%) | 95.4 |
| | Referring explicitly to otherâ€™s messages | 91 (5.21%) | 92.7 |
| | Asking questions | 800 (45.21%) | 89.4 |
| | Complementing, expressing appreciation | 1391 (79.6%) | 90.7 |
| | Expressing agreement | 243 (13.9%) | 96.6 |
| Cohesive | Vocatives | 1433 (82%) | 91.8 |
| | Addresses or refers to the group using inclusive pronouns | 144 (8.24%) | 88.8 |
| | Phatics, salutions | 1281 (73.3%) | 96.1 |

number from 2 to n (depends on the dataset), and the higher value indicates the ideal number of clusters for the k-means algorithm.

As presented in Section 3.1, our data was collected from 82 students, who could assume the expert and practicing researcher roles in different discussion threads. It is important to note that not every student assumed both roles during the course (i.e., some students simply decided not to perform their scripted role). To answer the first research questions, the unit of analysis of the clustering algorithm was a pair of students in a specific scripted role. Therefore, we did not evaluate only 82 instances (i.e., number of students), but a total of 163 instances, which is equivalent to all pairs of (student,role). We adopted social presence indicators and the phases of cognitive presence as features for the clustering algorithm. The final feature vector of each instance contained the number of indicators of each message posted in the online discussion by the pairs of (student,role).

## 3.3 Epistemic Network Analysis

To answer the second and the third research questions we adopted epistemic network analysis (ENA) [52]. ENA is a technique based on network analysis employed to explore rich relationships between a set of concepts. Initially applied to examine the relationships between concepts in discourse transcripts [52], ENA has been applied to study different educational phenomena such as the assessment of students, time management, learning strategies, the analysis of verbal data in CSCL, the evaluation of the development of different students groups in an online discussion, and measurement of the performance of an educational intervention [8, 35, 36, 43]. ENA can model complex domains as networks especially adopted for problems with a small set of concepts characterized by highly dynamic and dense interactions [50]. It provides a mechanism to compare the differences between different groups of analysis units – such as between the emerging and scripted roles in the present study.

ENA models connections among different *codes* (e.g., indicators for social presence) for each *unit of analysis* (e.g., student) based on co-occurrences of codes in data subsets called *stanzas* or *conversations* (e.g., co-occurrence of codes for indicators of social presence in every post of a online discussion). Based on code co-occurrences of all units of analysis, ENA builds a high-dimensional network representation called *analytic space* . Then, the units of analysis are plotted in a *projection space*, which is derived from the analytic space through singular value decomposition (SVD), a dimension reduction technique similar to principal component analysis. In the end, the output of ENA is a series of two-dimensional graph models which capture the relationships between the different codes and which are based on two SVD dimensions which typically explain the high proportion of variance in data[51].

Typical diagrams used in ENA include: (i) projection graphs, which plot centroids[2] of each unit of analysis as dots in the diagram and with squares representing the mean values of centroids of different groups; ii) the network graphs, which represent connections among the codes of a unit of analysis or a group of units of analysis; (iii) the subtraction network graph, which displays the differences in the connections of the codes related to two groups. Furthermore, it is also possible to use ENA to build a trajectory graph, which explains how networks of different groups changed over time. That is, trajectory graphs plot connections between centriods of the same unit of analysis.

To answer research question 2, each discussion message was coded with several binary codes capturing the presence/absence of the social presence indicators and the cognitive presence phases. Both units of analysis and stanzas were students (i.e., all student messages) within different roles (emerging and scripted). The use of students as units of analysis here enabled us to see, for each student, their connections between the phases of cognitive presence and the indicators of social presence. Specifically, we answered research question 2 through the projection and subtraction graphs. Moreover, the differences between the roles on both svd1 and svd2 values were then compared by using a series of the Mann-Whitney tests [46] with $\alpha$=0.05.

[2] A centroid is the mean value of edge weights for a given epistemic network [51]

As research question 3 aimed to analyse the temporal development in the relationship between cognitive and social presences for emerging and scripted roles, we used a different configuration in the application of ENA. We adopted week and day as the unit of analysis and stanza, respectively. This analysis produced a different network graph (i.e., the values of the svd dimensions differed from those produced by ENA applied to research questions 2). The goal is to use the trajectory graph to assess the change of each scripted and emerging role over the weeks of the course discussion.

# 4 RESULTS

## 4.1 Research Question 1

The analysis of different values for k (2..30) identified 3 as the optimal number of clusters with the value of 0.3644 for the silhouette coefficient. Then, we employed the k-means algorithm to identify the three groups (cluster 0, cluster 1, and cluster 2), which are considered the *emerging roles* in this analysis. Table 3 shows the percentage of indicators for each cluster identified and the scripted roles (expert and practicing research) and the number of students and messages for each role.

Clusters 0 and 1 presented similar values for cognitive presence and affective category of social presence. However, cluster 0 showed a strong focus on interactivity in the discussions (i.e., interactive category of social presence) and creating an environment where the students are deeply involved in a conversation (Cohesive category). Moreover, it is possible to see that students in cluster 1 were less active during the discussion (with an average of 5.55 posts by students). In short, students in both clusters (0 and 1) can be described as cognitively engaged, but cluster 0 also had relevant social interactions. Therefore, we named clusters 0 and 1 as the ***socio-cognitively engaged*** and ***cognitively engaged*** emerging roles, respectively.

Cluster 2 had a similar trend with that of the practicing-researcher scripted role according to the indicators of social presence. This trend could be explained by the small number of students who performed the role of experts in this cluster (only one). However, this cluster also contained most of its messages in the triggering event (which could also be associated with asking questions, an indicator of social presence) and exploration phases, both of which represent low levels of cognitive presence. As the main focus of students in cluster 2 was the social presence, we categorise this cluster as the **socially-focused** emerging role.

## 4.2 Research Question 2

Figure 2 presents the projection of the average networks of each emerging and scripted roles. The rectangles represent group-average networks (95% CI are also outlined) for students in the expert (red), practicing researcher (blue), socio-cognitively engaged (green), cognitively engaged (orange), and socially-focused (purple) roles. The graph shows that svd1 and svd2 dimensions accounted for 18.8 and 12.2 percent of variance in data, respectively. The differences between the students in the expert and practicing researcher roles can be observed visually, with the key difference along the svd1 (U=283.50, p=0.00, r=0.91). In terms of emerging roles, the difference between socio-cognitively engaged students and cognitively engaged students was along the svd2(U=2206.00, p=0.04, r=0.20), while

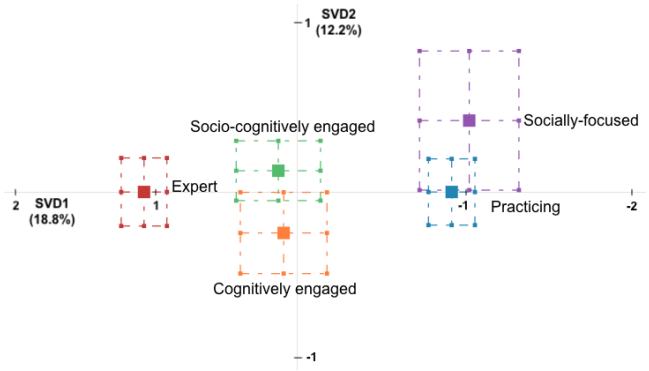



**Figure 2: ENA projections of the means values of group networks of the students who assumed different scripted and emerging roles. Each colored square (95% CI are outlined around the group means) represent a mean network of one role (emerging or scripted).**

the students when played the socially-focused role differentiated from the cognitively engaged role along both svd1(U=190.00, p=0.00, r=0.59) and svd2(U=672.00, p=0.01, r=0.43), and from the socio-cognitively engaged role on the svd1 (U=254.00, p=0.00, r=0.56). Finally, Figure 2 also shows that socially-focused role overlapped with the practicing researcher scripted role, and the socio-cognitively engaged and cognitively engaged roles were relatively close to the expert role.

Figure 3 shows the subtraction networks comparing the scripted and emerging roles. It reveals that the socio-cognitively engaged emerging role had a similar level of participation compared to that of the expert script role; however, the students in the socio-cognitively engaged emerging role made more connections with the indicators of social presence (Asking_Question, Salutation, Referring_Message, Group) than the expert emerging role did (Figure 3a). Following a similar trend, the cognitively engaged emerging role also possessed connections with the aforementioned indicators of social presence (Figure 3c), but the expert scripted role had stronger connections of the exploration and integration phases with the indicators of social presence plotted on the left side of the svd1-axis. Finally, Figure 3f confirms the similarity between socially-focused and the practicing researcher role, while Figure 3e indicates a weak connection of socially-focused role and the cognitive presence phases when compared to the expert role.

## 4.3 Research Question 3

The ENA trajectory graph was employed to answer research question 3, providing details on how networks of the students in different roles changed in terms of connections between social and cognitive presences over the four weeks of the discussion. Figure 4a presents the final network built using the entire dataset. Again, the visualisation was done using svd1 and svd2, which accounted for 63.8 and 10.1 percent of variability, respectively. The network in Figure 4a shows that svd1 and svd2 tended to explain the indicators of social presence and the cognitive presence phases, respectively. In terms

**Table 3: Percentage of indicators for emerging (Clusters 0–2) and scripted (Expert and Practicing Researcher) roles**

| | Emerging roles | | | Scripted roles | |
|---|---|---|---|---|---|
| | Cluster 0: Socio-cognitively engaged | Cluster 1: Cognitively engaged | Cluster 2: Socially-focused | Expert | Practicing |
| **Social presence** | | | | | |
| *Affective* | | | | | |
| Emotions | 16.43% | 17.20% | 15.84% | 22.12% | 11.15% |
| Humor | 02.28% | 02.42% | 00.42% | 02.94% | 02.12% |
| Self discosure | 17.00% | 17.74% | 23.91% | 17.06% | 19.73% |
| *Interactive* | | | | | |
| Continuing Thread | 95.73% | 90.32% | 99.38% | 90.24% | 100.00% |
| Quoting Message | 04.75% | 03.23% | 00.93% | 06.47% | 01.11% |
| Referring Message | 05.32% | 05.91% | 04.04% | 05.65% | 04.79% |
| Asking Question | 40.36% | 37.37% | 73.29% | 14.47% | 75.47% |
| Complimenting | 81.67% | 64.25% | 90.68% | 70.94% | 87.85% |
| Agreement | 14.81% | 16.67% | 07.76% | 17.18% | 10.81% |
| *Cohesive* | | | | | |
| Vocatives | 84.14% | 66.94% | 92.55% | 77.41% | 86.40% |
| Group | 08.55% | 08.87% | 06.52% | 09.41% | 07.13% |
| Salutations | 75.78% | 60.48% | 80.12% | 72.00% | 74.58% |
| **Cognitive presence** | | | | | |
| Triggering Event | 16.14% | 16.13% | 24.22% | 17.65% | 17.61% |
| Exploration | 38.46% | 37.10% | 43.79% | 39.76% | 38.57% |
| Integration | 31.53% | 29.30% | 20.81% | 29.18% | 28.99% |
| Resolution | 06.46% | 06.99% | 04.04% | 06.12% | 06.13% |
| Others | 07.41% | 10.48% | 07.14% | 07.29% | 08.70% |
| No. of students | 82 | 67 | 14 | 82 | 81 |
| No. of students - Expert | 45 | 36 | 1 | 82 | 0 |
| No. of students - Practicing | 37 | 31 | 13 | 0 | 81 |
| No. of posts | 1053 | 372 | 322 | 850 | 897 |
| No. of posts per student | 12.84 | 5.55 | 23 | 10.36 | 11.07 |

of social presence, the right-hand side of svd1 contains indicators related to basic interactions (asking questions and referring messages) and affective indicators (emotion and humor), while the left-hand side encompasses more indicators related to group cohesion (group, agreement, self_disclosure). Another interesting observation is that the integration and exploration phases of cognitive presence were plotted in distinct parts of the graph, leaving the resolution and triggering event categories in the middle.

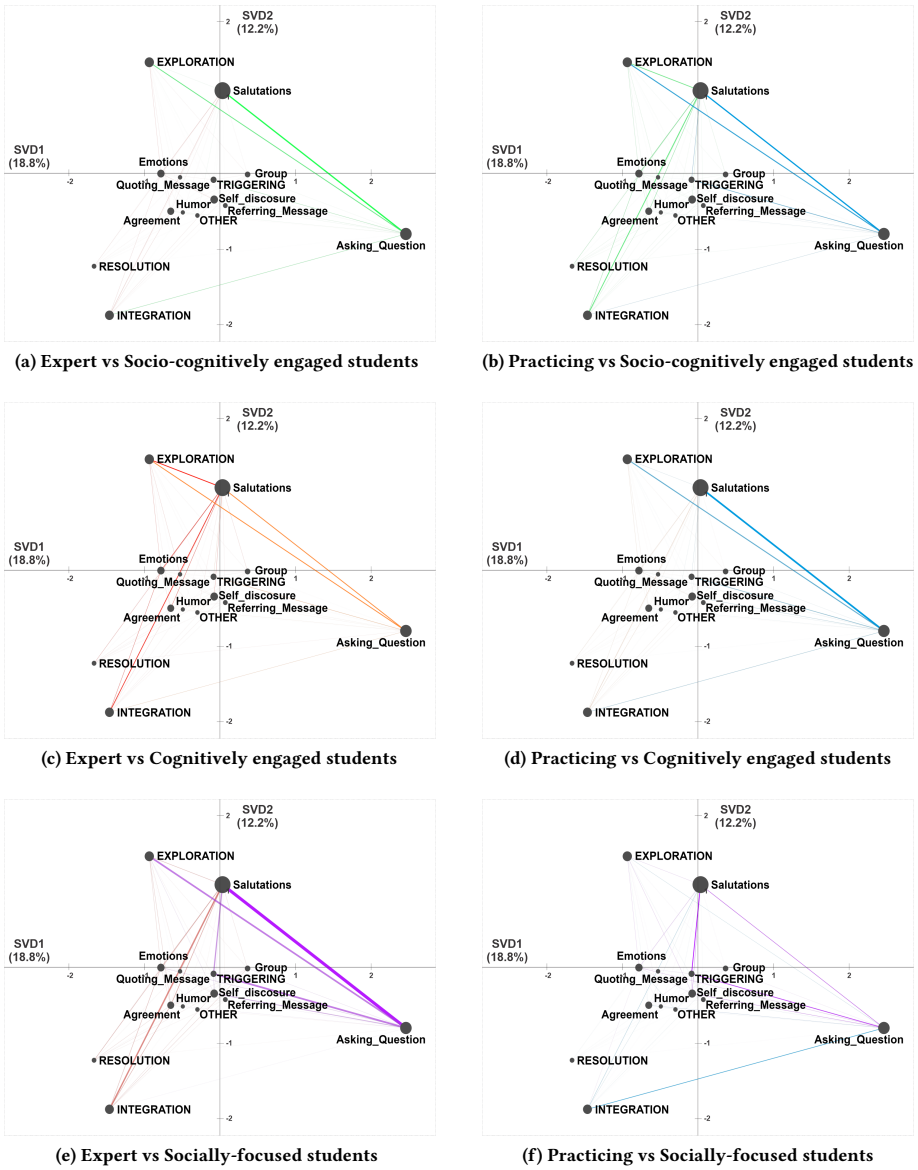
The trajectories of the role over the four weeks (Figure 4b) were plotted in the same high-dimensional space as it shown in Figure 4a. It means that we could analyse the trajectory of each group in relation to the position of social and cognitive presence in Figure 4a. As mentioned before, the trajectory graph shows the location of the central activity performed by students from each role in each week of the discussion connected by a line that represents how the groups evolved from one week to the other. This analysis revealed a similar trend between students in the socially-focused emerging role and practicing researcher scripted role, where they were more related to indicators such as asking questions, referring to previous messages, and triggering events. The socio-cognitively engaged and cognitively engaged emerging roles had some similarities, as they were plotted on the central part of the graph. The socio-cognitively engaged emerging role was closer to the loop between the triggering event-resolution phases of cognitive presence, while the cognitively engaged emerging role initiated week 1 closer to the exploration phase of social presence. Finally, the trajectory graph revealed that the students in the expert scripted role favoured the exploration and group cohesion indicators.

# 5 DISCUSSION

## 5.1 Interpretation of the Results

The results related to research question 1 showed that the indicators of social presence and phases of cognitive presence could be suitable to identify emerging roles (defined as clusters) as confirmed with the silhouette value of 0.3644 [53]. The results presented in Table 3 detail the main features of different emerging roles, which were categorised as socio-cognitively engaged, cognitively engaged and socially-focused roles. It is note that the emerging roles demonstrated some similarities and differences with scripted roles.

The emerging roles identified with the use of the NASC approach had some parallels with the existing research on scripted and emerging roles. First, the literature shows that when an instructor assign the students to scripted roles, the general trend is that students participate according to their scripted roles, especially when the




(a) Expert vs Socio-cognitively engaged students

(b) Practicing vs Socio-cognitively engaged students

(c) Expert vs Cognitively engaged students

(d) Practicing vs Cognitively engaged students

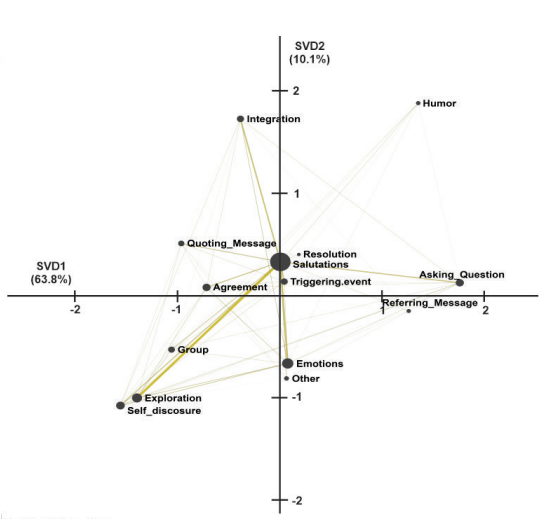
(e) Expert vs Socially-focused students

(f) Practicing vs Socially-focused students

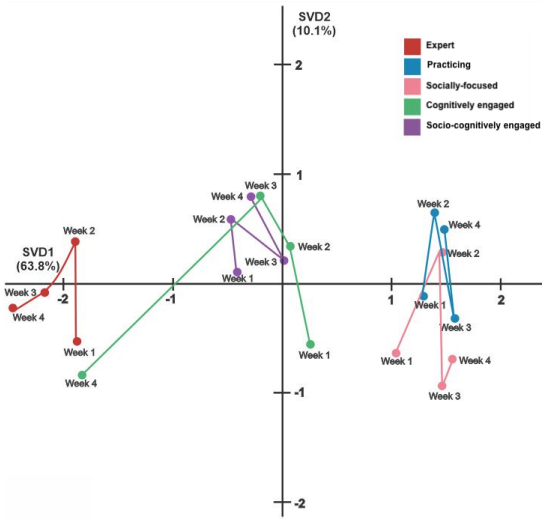
**Figure 3: ENA subtraction of student networks related to the scripted (Expert and Practicing) and emerging roles (clusters 0, 1 and 2)**

participation in group work counts towards final marks [33, 54]. However, when the same student needs to act differently in distinct online discussions, a previous role the student was asked to play could influence the attitudes of the students in the future role assignments [39]. Our results confirm this trend as two out of the three emerging roles (except the socially-focused emerging role) contained students from both scripted roles. Second, it is important to highlight that the scripted roles investigated in this study were related to simple instructions (experts – initiating and maintain a discussion and practicing researchers – asking questions and post comments). However, the CSCL community usually suggests the adoption of more cognitively-related (e.g., argumentative roles) scripted roles [11, 54]. Therefore, the emerging roles identified in this study show a higher similarity with previous CSCL literature papers and provide additional information on the students' behaviour during the discussion.

Research question 2 aimed to unpack the association between social and cognitive presences for scripted and emerging roles in asynchronous online discussions. The results confirmed that each emerging role tended to be related to one of the scripted roles. While socio-cognitively engaged and cognitively engaged emerging roles were closer to the expert scripted role, students in the




(a) Group-average networks graph of the relationship between social and cognitive presences for the development of all students over the four weeks of discussions.




(b) Mean trajectory graph for the students in the different groups including expert (red), practicing (blue), cluster 0 (salmon), cluster 1 (green), and cluster 2 (purple).

**Figure 4: Students' trajectory analysis over four weeks of discussions.**

socially-focused emerging role had a participation level similar to that of the practicing researcher scripted role. Despite the similarities of socio-cognitively engaged and cognitively engaged emerging roles with and the expert scripted role, the results showed that the expert role leaned towards higher cognitive presence levels. This result is sensible as the expert role was designed to be in charge of discussions [19]. Moreover, the participation of students in socio-cognitively engaged and cognitively engaged emerging roles was slightly different. On the one hand, socio-cognitively engaged role predominantly showed social presence in terms of interactive indicators; this can explain the strong relation of the socio-cognitively engaged role with the exploration and integration phases, which are reported in the literature to be more related to the the interactive category of social presence [43, 44]. In contrast, students in the cognitively engaged emerging role focused on affective messages and asked questions representative of the triggering event phase [43, 44].

As the socially-focused emerging role encompassed a subgroup of the students who assumed the roles of practicing researchers, its relationship with the expert role is similar to the relationship between the practicing and expert scripted roles as reported in the literature [43]. In short, the socially-focused emerging role mostly remained on low levels of cognitive presence and was related to the salutation and asking question indicators of social presence. The direct comparison between the socially-focused emerging role and the practicing researcher script role showed that students in the socially-focused emerging role emerging role did not engage deeply in the discussion as their messages were more related to the triggering event phase and the salutation and asking questions indicators of social presence. Thus, students in the socially-focused role could potentially be in the lower performance group in the discussion [6].

The last research question aimed to analyse changes in the participation of each emerging role over time through the use of trajectory analysis provided by ENA. This analysis reemphasised the need for analysing the emerging roles even if the course design previously introduced scripted roles. It is clear that the students acting as experts focused on answering questions (i.e., exploration phase) and encourage group cohesion (i.e., group and self-disclosure). However, they did not entirely act as expected as the goal of this role was the initiation and conclusion of the discussion that could be related to the triggering and resolution phases, respectively [19, 43]. On the other hand, both socio-cognitively engaged and cognitively engaged roles were closely related to this goal.

## 5.2 Implications

The findings of the study showed that the use of the NASC approach is promising for the analysis of emerging roles in online discussions. Not only could the approach automatically identify theoretically meaningful emerging roles, but it could compare them to the script roles and track progression over time in terms of well-established higher-order constructs of social and cognitive presence. This can offer useful insights to the teachers to inform their decisions and improve overall teaching presence in a community of inquiry [12]. NASC can provide teachers with insights in the efficacy of their pedagogical intentions (i.e., script roles), so that they can take relevant action such as changes to the scripts of roles. It can also help them track student progress in terms of social and cognitive presences and create a foundation for enhancement of educational experience by giving timely feedback and recognising students that are not cognitively engaged in the discussion [43]. Moreover, NASC also holds a potential to be used in formative assessment of collaboration skills [6] by emphasising the developmental nature of assessment [32] through the use of trajectory graphs of ENA .

While the NASC approach showed premising results in the analysis of student roles even bigger potential lies in its integration with the existing body of research in learning and collaboration analytics [49]. The approach can effectively be used together with existing approaches for automatic classification of online discussion messages for cognitive and social presence [2, 29, 34]. This in turn eases potential adoption of the proposed analytic approach without a need for the users to manually code messages, as it was done in the current study. It would however be a promising research direction to investigate how the NASC approach can be complemented with existing work in collaboration analytics that makes use of social network analysis to further probe the social dimension of collaboration [18, 38] and automatic methods for automatic analysis of other constructs of the cognitive dimension of collaboration [23].

## 5.3 Limitations and future work

While this proposed approach shows a promise in addressing significant issues in the research on the association of emerging and scripted roles with social and cognitive presence in developing communities of inquiry, there are several limitations of the present study that should be acknowledged. First, this study was based on the data from six offerings of the same course at a single institution, which can negatively affect the broader implications of the presented analytic approach and the generalisability of the results obtained. Second, the findings of the present study findings may be somewhat limited, given the specifics of the adopted course design and the scripted roles. Therefore, it is important to apply the same analytic approach to other datasets collected in different course settings and with online discussions in a different language other than English to address those problems. Finally, as common for cluster analysis, it involves making many methodological decisions, such as deciding on the number of clusters and the features used, it might be that the findings in the present study would be different if different algorithms and parameters were used. Thus, it would be prudent to evaluate different clustering algorithms and experiment with other features in cluster analysis as suggested in the related literature [6, 39].

## ACKNOWLEDGMENTS

This research was partially supported by the fellowship 310060/2018-4 of the National Council for Scientific and Technological Development – CNPq.

## REFERENCES

[1] Terry Anderson, Liam Rourke, D. Randy Garrison, and Walter Archer. 2001. Assessing Teaching Presence in a Computer Conferencing Context. *Journal of Asynchronous Learning Networks* 5 (2001), 1–17.

[2] Arthur Barbosa, Máverick Ferreira, Rafael Ferreira Mello, Rafael Dueire Lins, and Dragan Gasevic. 2021. The impact of automatic text translation on classification of online discussions for social and cognitive presences. In *LAK21: 11th International Learning Analytics and Knowledge Conference*. 77–87.

[3] Gian Barbosa, Raissa Camelo, Anderson Pinheiro Cavalcanti, Péricles Miranda, Rafael Ferreira Mello, Vitomir Kovanović, and Dragan Gašević. 2020. Towards automatic cross-language classification of cognitive presence in online discussions. In *Proceedings of the tenth international conference on learning analytics & knowledge*. 605–614.

[4] Maarten De Laat and Vic Lally. 2004. It's not so easy: Researching the complexity of emergent participant roles and awareness in asynchronous networked learning discussions. *Journal of Computer Assisted Learning* 20, 3 (2004), 165–171.

[5] Bram De Wever, Hilde Van Keer, Tammy Schellens, and Martin Valcke. 2007. Applying multilevel modelling to content analysis data: Methodological issues in the study of role assignment in asynchronous discussion groups. *Learning and instruction* 17, 4 (2007), 436–447.

[6] Nia MM Dowell and Oleksandra Poquet. 2021. SCIP: Combining Group Communication and Interpersonal Positioning to Identify Emergent Roles in Scaled Digital Environments. *Computers in Human Behavior* (2021), 106709.

[7] Máverick Ferreira, Vitor Rolim, Rafael Ferreira Mello, Rafael Dueire Lins, Guanliang Chen, and Dragan Gašević. 2020. Towards automatic content analysis of social presence in transcripts of online discussions. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. 141–150.

[8] Rafael Ferreira, Vitomir Kovanović, Dragan Gašević, and Vitor Rolim. 2018. Towards combined network and text analytics of student discourse in online discussions. In *International conference on artificial intelligence in education*. Springer, 111–126.

[9] Rafael Ferreira-Mello, Máverick André, Anderson Pinheiro, Evandro Costa, and Cristobal Romero. 2019. Text mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 9, 6 (2019), e1332.

[10] Rafael Ferreira Mello and Dragan Gašević. 2019. What is the Effect of a Dominant Code in an Epistemic Network Analysis?. In *Advances in Quantitative Ethnography*, Brendan Eagan, Morten Misfeldt, and Amanda Siebert-Evenstone (Eds.). Springer International Publishing, Cham, 66–76.

[11] Frank Fischer, Ingo Kollar, Karsten Stegmann, and Christof Wecker. 2013. Toward a script theory of guidance in computer-supported collaborative learning. *Educational psychologist* 48, 1 (2013), 56–66.

[12] D Randy Garrison. 2015. *Thinking collaboratively: Learning in a community of inquiry*. Routledge.

[13] D Randy Garrison, Terry Anderson, and Walter Archer. 2000. Critical inquiry in a text-based environment: Computer conferencing in higher education. *The internet and higher education* 2, 2-3 (2000), 87–105. https://doi.org/10.1016/S1096-7516(00)00016-6

[14] D. Randy Garrison, Terry Anderson, and Walter Archer. 2001. Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education* 15, 1 (2001), 7–23. https://doi.org/10.1080/08923640109527071

[15] D. Randy Garrison, Terry Anderson, and Walter Archer. 2010. The first decade of the community of inquiry framework: A retrospective. *The Internet and Higher Education* 13, 1-2 (2010), 5–9.

[16] D Randy Garrison and Martha Cleveland-Innes. 2005. Facilitating cognitive presence in online learning: Interaction is not enough. *The American journal of distance education* 19, 3 (2005), 133–148.

[17] D Randy Garrison, Martha Cleveland-Innes, and Tak Shing Fung. 2010. Exploring causal relationships among teaching, cognitive and social presence: Student perceptions of the community of inquiry framework. *The internet and higher education* 13, 1-2 (2010), 31–36.

[18] Dragan Gašević, Srećko Joksimović, Brendan R Eagan, and David Williamson Shaffer. 2019. SENS: Network analytics to combine social and cognitive perspectives of collaborative learning. *Computers in Human Behavior* 92 (2019), 562–577.

[19] Dragan Gašević, Olusola Adesope, Srećko Joksimović, and Vitomir Kovanović. 2015. Externally-facilitated regulation scaffolding and role assignment to develop cognitive presence in asynchronous online discussions. *The Internet and Higher Education* 24 (2015), 53–65. https://doi.org/10.1016/j.iheduc.2014.09.006

[20] Greg Hamerly and Charles Elkan. 2004. Learning the k in k-means. *Advances in neural information processing systems* 16 (2004), 281–288.

[21] Friedrich Hesse, Esther Care, Juergen Buder, Kai Sassenberg, and Patrick Griffin. 2015. Assessment and Teaching of 21st Century Skills. *The Internet and Higher Education* (2015), 37–56.

[22] Srećko Joksimović, Dragan Gašević, Vitomir Kovanović, Bernhard E Riecke, and Marek Hatala. 2015. Social presence in online discussions as a process predictor of academic performance. *Journal of Computer Assisted Learning* 31, 6 (2015), 638–654.

[23] Srećko Joksimović, Jelena Jovanović, Vitomir Kovanović, Dragan Gašević, Nikola Milikić, Amal Zouaq, and Jan Paul Van Staalduinen. 2020. Comprehensive analysis of discussion forum participation: from speech acts to discussion dynamics and course outcomes. *IEEE Transactions on Learning Technologies* 13, 1 (2020), 38–51.

[24] Min Kyu Kim and Tuba Ketenci. 2019. Learner participation profiles in an asynchronous online collaboration context. *The Internet and Higher Education* 41 (2019), 62–76.

[25] Vitomir Kovanović, Dragan Gašević, Srećko Joksimović, Marek Hatala, and Olusola Adesope. 2015. Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions. *The Internet and Higher Education* 27 (2015), 74–89.

[26] Vitomir Kovanović, Srecko Joksimović, Dragan Gašević, and Marek Hatala. 2014. What is the source of social capital? The association between social network position and social presence in communities of inquiry. In *Workshop on Graph-Based Educational Data Mining*. 1.

[27] Vitomir Kovanović, Srećko Joksimović, Oleksandra Poquet, Thieme Hennis, Iva Čukić, Pieter De Vries, Marek Hatala, Shane Dawson, George Siemens, and

Dragan Gašević. 2018. Exploring communities of inquiry in massive open online courses. *Computers & Education* 119 (2018), 44–58.

[28] Vitomir Kovanović, Srećko Joksimović, Oleksandra Poquet, Thieme Hennis, Pieter de Vries, Marek Hatala, Shane Dawson, George Siemens, and Dragan Gašević. 2019. Examining communities of inquiry in Massive Open Online Courses: The role of study strategies. *The Internet and Higher Education* 40 (2019), 20–43.

[29] Vitomir Kovanović, Srećko Joksimović, Zak Waters, Dragan Gašević, Kirsty Kitto, Marek Hatala, and George Siemens. 2016. Towards automated content analysis of discussion transcripts: A cognitive presence case. In *Proceedings of the sixth international conference on learning analytics & knowledge*. ACM, 15–24.

[30] Nale Lehmann-Willenbrock, Stephenson J Beck, and Simone Kauffeld. 2016. Emergent team roles in organizational meetings: Identifying communication patterns via cluster analysis. *Communication Studies* 67, 1 (2016), 37–57.

[31] José-Antonio Marcos-García, Alejandra Martínez-Monés, and Yannis Dimitriadis. 2015. DESPRO: A method based on roles to provide collaboration analysis support adapted to the participants in CSCL situations. *Computers & Education* 82 (2015), 335–353.

[32] Roberto Martinez-Maldonado, Dragan Gaševic, Vanessa Echeverria, Gloria Fernandez Nieto, Zachari Swiecki, and Simon Buckingham Shum. 2021. What Do You Mean by Collaboration Analytics? A Conceptual Model. *Journal of Learning Analytics* 8, 1 (2021), 126–153.

[33] Rosa M Mayordomo and Javier Onrubia. 2015. Work coordination and collaborative knowledge construction in a small group collaborative virtual task. *The Internet and Higher Education* 25 (2015), 96–104.

[34] Valter Neto, Vitor Rolim, Anderson Pinheiro Cavalcanti, Rafael Dueire Lins, Dragan Gasevic, and Rafael Ferreiramello. 2021. Automatic Content Analysis of Online Discussions for Cognitive Presence: A Study of the Generalizability across Educational Contexts. *IEEE Transactions on Learning Technologies* (2021).

[35] Nataša Pantić, Sarah Galey, Lani Florian, Srećko Joksimović, Gil Viry, Dragan Gašević, Helén Knutes Nyqvist, and Krystallia Kyritsi. 2021. Making sense of teacher agency for change with social and epistemic network analysis. *Journal of Educational Change* (2021), 1–33.

[36] Luc Paquette, Theodore Grant, Yingbin Zhang, Gautam Biswas, and Ryan Baker. 2021. Using Epistemic Networks to Analyze Self-regulated Learning in an Open-Ended Problem-Solving Environment. In *Proceedings of the 2nd International Conference on Quantitative Ethnography*, Andrew R. Ruis and Seung B. Lee (Eds.). Springer International Publishing, Cham, 185–201.

[37] Abelardo Pardo, Kathryn Bartimote, Simon Buckingham Shum, Shane Dawson, Jing Gao, Dragan Gašević, Steve Leichtweis, Danny Liu, Roberto Martínez-Maldonado, Negin Mirriahi, et al. 2018. OnTask: Delivering data-informed, personalized learning support actions. *Journal of Learning Analytics* 5, 3 (2018), 235–249.

[38] Oleksandra Poquet and Jelena Jovanovic. 2020. Intergroup and interpersonal forum positioning in shared-thread and post-reply networks. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. 187–196.

[39] Francesca Pozzi. 2011. The impact of scripted roles on online collaborative learning processes. *International Journal of Computer-Supported Collaborative Learning* 6, 3 (2011), 471–484.

[40] Anika Radkowitsch, Freydis Vogel, and Frank Fischer. 2020. Good for learning, bad for motivation? A meta-analysis on the effects of computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning* 15, 1 (2020), 5–47.

[41] Jennifer Richardson, Karen Swan, Patrick Lowenthal, and Phil Ice. 2016. Social presence in online learning: Past, present, and future. In *Global Learn*. Association for the Advancement of Computing in Education (AACE), 477–483.

[42] Jennifer C Richardson, Yukiko Maeda, Jing Lv, and Secil Caskurlu. 2017. Social presence in relation to students' satisfaction and learning in the online environment: A meta-analysis. *Computers in Human Behavior* 71 (2017), 402–417.

[43] Vitor Rolim, Rafael Ferreira, Rafael Dueire Lins, and Dragan Găsević. 2019. A network-based analytic approach to uncovering the relationship between social and cognitive presences in communities of inquiry. *The Internet and Higher Education* 42 (2019), 53–65.

[44] Liam Rourke, Terry Anderson, D. Randy Garrison, and Walter Archer. 1999. Assessing Social Presence In Asynchronous Text-based Computer Conferencing. *The Journal of Distance Education* 14, 2 (1999), 50–71. https://www.learntechlib.org/p/92000/

[45] Liam Rourke, Terry Anderson, D. Randy Garrison, and Walter Archer. 2007. Assessing Social Presence In Asynchronous Text-based Computer Conferencing. *The Journal of Distance Education / Revue de l'Éducation à Distance* 14, 2 (2007), 50–71. http://eric.ed.gov/?id=EJ616753

[46] Graeme D Ruxton. 2006. The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test. *Behavioral Ecology* 17, 4 (2006), 688–690.

[47] Mohammed Saqr and Olga Viberg. 2020. Using Diffusion Network Analytics to Examine and Support Knowledge Construction in CSCL Settings. In *In Proceedings of the 15th European Conference on Technology Enhanced Learning*. Springer, Cham, Switzerland, 158–172.

[48] Tammy Schellens, Hilde Van Keer, Bram De Wever, and Martin Valcke. 2007. Scripting by assigning roles: Does it improve knowledge construction in asynchronous discussion groups? *International Journal of Computer-Supported Collaborative Learning* 2, 2 (2007), 225–246.

[49] Bertrand Schneider, Nia Dowell, and Kate Thompson. 2021. Collaboration Analytics—Current State and Potential Futures. *Journal of Learning Analytics* 8, 1 (2021), 1–12.

[50] David Williamson Shaffer. 2004. Epistemic Frames and Islands of Expertise: Learning from Infusion Experiences. In *Proceedings of the 6th International Conference on Learning Sciences (ICLS '04)*. International Society of the Learning Sciences, Santa Monica, California, 473–480.

[51] David Williamson Shaffer, Wesley Collier, and Andrew R Ruis. 2016. A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics* 3, 3 (2016), 9–45.

[52] David Williamson Shaffer, David Hatfield, Gina Navoa Svarovsky, Padraig Nash, Aran Nulty, Elizabeth Bagley, Ken Frank, André A Rupp, and Robert Mislevy. 2009. Epistemic network analysis: A prototype for 21st-century assessment of learning. *International Journal of Learning and Media* 1, 2 (2009).

[53] Artur Starczewski and Adam Krzyżak. 2015. Performance evaluation of the silhouette index. In *International Conference on Artificial Intelligence and Soft Computing*. Springer, 49–58.

[54] Jan-Willem Strijbos and Armin Weinberger. 2010. Emerging and scripted roles in computer-supported collaborative learning. *Computers in Human Behavior* 26, 4 (2010), 491–494.

[55] Alyssa Friend Wise and Ming Ming Chiu. 2011. Analyzing temporal patterns of knowledge construction in a role-based online discussion. *International Journal of Computer-Supported Collaborative Learning* 6, 3 (2011), 445–470.

[56] Alyssa Friend Wise, Marzieh Saghafian, and Poornima Padmanabhan. 2012. Towards more precise design guidance: Specifying and testing the functions of assigned student roles in online discussions. *Educational Technology Research and Development* 60, 1 (2012), 55–82.

[57] Alyssa Friend Wise and Baruch B Schwarz. 2017. Visions of CSCL: Eight provocations for the future of the field. *International Journal of Computer-Supported Collaborative Learning* 12, 4 (2017), 423–467.

[58] Rui Xu and Don Wunsch. 2008. *Clustering*. Vol. 10. John Wiley & Sons.