Computers & Education 189 (2022) 104581

Contents lists available at ScienceDirect

# Computers & Education

journal homepage: www.elsevier.com/locate/compedu

# How CSCL roles emerge, persist, transition, and evolve over time: A four-year longitudinal study

Mohammed Saqr [a,*], Sonsoles López-Pernas [a,b]

[a] University of Eastern Finland, School of Computing, Joensuu, Yliopistokatu 2, fi-80100, Joensuu, Finland
[b] Universidad Politécnica de Madrid, ETSI Sistemas Informáticos, Departamento de Sistemas Informáticos, C/ Alan Turing s/n, 28031, Madrid, Spain

ARTICLE INFO

*Keywords:*
Computer-supported collaborative learning
Learning analytics
Roles
Person-centered
Longitudinal analysis

ABSTRACT

A prevailing trend in CSCL literature has been the study of students' participatory roles. The majority of existing studies examine a single collaborative task or, at most, a complete course. This study aims to investigate the presence —or the lack thereof— of a more enduring disposition that drives student participation patterns across courses. Based on data from a 4-year program where 329 students used CSCL to collaborate in 10 successive courses (amounting up to 84,597 interactions), we identify the emerging roles using centrality measures and latent profile analysis (LPA) and trace the unfolding of roles over the entire duration of the program. Thereafter, we use Mixture Hidden Markov Models (MHMM) —methods that are particularly useful in detecting "latent traits" in longitudinal data— to identify how students' roles, transition, persist or evolve over time. Relevant covariates were also examined to explain students' membership of different trajectories. We identified three different roles (*leader*, *mediator*, *isolate*) at the course level. At the program level, we found three distinct trajectories: an *intense* trajectory with mostly leaders, a *fluctuating* trajectory with mostly mediators, and a *wallowing-in-the-mire* trajectory with mostly isolates. Our results show that roles re-emerge consistently regardless of the task or the course over extended periods of time and in a predictable manner. For instance, isolates "assumed" such a role in almost all of their courses over four years.

## 1. Introduction

Computer-supported collaborative learning (CSCL) has gained increasing adoption among practitioners, researchers and educators over the past four decades resulting in a stream of tools, research traditions and diverse applications (Dillenbourg & Fischer, 2007; Ludvigsen et al., 2021). Several definitions and conceptualizations exist: the most fundamental view of CSCL encompasses a tripartite structure where collaboration and social interaction, within a group of individuals, is mediated or facilitated by computational artifacts (Ludvigsen et al., 2021; Stahl et al., 2014). While such a conceptualization may look crudely reductionist, CSCL involves complex interactions and interdependencies among multifarious components, creating vast opportunities for learners and educators as well as

*Abbreviations:* CSCL, Computer-supported collaborative learning; MHMM, Mixture Hidden Markove Models; SNA, Social Network Analysis; MOOC, Massive Open Online Course; PBL, Problem-based Learning; LPA, Latent Profile Analysis.

* Corresponding author. University of Eastern Finland, School of Computing, Joensuu Campus, Yliopistokatu 2, P.O. Box 111, fi-80100. Joensuu, Finland.

*E-mail addresses:* mohammed.saqr@uef.fi (M. Saqr), sonsoles.lopez.pernas@upm.es (S. López-Pernas).

https://doi.org/10.1016/j.compedu.2022.104581
Received 4 March 2022; Received in revised form 13 May 2022; Accepted 30 June 2022
Available online 5 July 2022
0360-1315/

tangible challenges (Dillenbourg et al., 2009; Stahl et al., 2014). On the one hand, CSCL offers learners the opportunity to jointly work on a task, interact, exchange resources, co-construct knowledge, establish communities, as well as monitor and co-regulate their learning (Jeong & Hmelo-Silver, 2016). On the other hand, CSCL has obvious challenges regarding design, orchestration, roles, contextualization and management of the multiplicity of individuals, group processes and cultures (Dillenbourg et al., 2009; Järvelä & Hadwin, 2013; Kreijns et al., 2003). Driven by the potentials CSCL could offer, the application thereof has extended or integrated into several pedagogical approaches, e.g., game-based learning, team-based learning, and problem-based learning (PBL) (Car et al., 2019; Lu et al., 2010).

A growing body of knowledge has established that productive collaboration in CSCL —and collaboration at large— is hardly spontaneous and therefore, several methods have been developed to steer the process into an effective collaboration (De Wever & Strijbos, 2021; Järvelä & Hadwin, 2013; Strijbos et al., 2006). Such methods include —inter alia— scaffolding, teacher orchestration, facilitating, structuring, as well as supporting roles. Roles —the main focus of our article— are a set of responsibilities and functions that guide individuals' behavior toward the group as well as towards other collaborators (De Wever & Strijbos, 2021; Hare, 1994; Järvelä & Hadwin, 2013). Roles have several important functions and advantages for the collaborators and the success of the collaborative process. Roles facilitate group regulation, distribution of tasks, and coordination of efforts (De Wever et al., 2008; Strijbos & De Laat, 2010). Groups with defined roles have a rather smooth and effective collaborative process with fewer conflicts (De Wever et al., 2008). Roles can also enhance individual responsibility, positive interdependence, accountability, as well as group cohesion. All such factors are essential ingredients for a functioning collaborative group (De Wever et al., 2008; De Wever & Strijbos, 2021; Strijbos & De Laat, 2010).

The large volume of research at hand has extensively studied various types of roles at different levels. Nonetheless, previous literature reviews (e.g., De Wever & Strijbos, 2021; Strijbos & De Laat, 2010) as well as our updated review of the literature (see section 2.1.2) show that research has been hitherto confined to limited periods of time (a course or two). A gap —therefore— exists in our body of knowledge regarding role dynamics over longer periods of time (i.e., a full educational program). Little is known —if any— about students' pervasive pattern of collaboration, which define and shape their emergent roles at longitudinal levels, e.g., a program (Strijbos & De Laat, 2010). To that end, the present work aims to fill such a gap by studying the transitions and longitudinal unfolding of roles across a program, the heterogeneity, and subtypes of roles' trajectories, as well as the factors that predict or explain the unfolding of such trajectories. Our research questions are as follows:

- **RQ1**: What are the types of roles that students assume in CSCL across a full program and what are their defining criteria?
- **RQ2**: Are there distinct trajectories of students' roles across a full program? If so, what are their defining criteria?
- **RQ3**: How long do students assume a role, consistently stay in such a role and how likely are they to conclude the program in that role?
- **RQ4**: What are the variables that predict or explain students taking a certain trajectory?

## 2. Background

### 2.1. Roles in CSCL

The worth of roles as a subject of research and practice predates the birth of CSCL by several decades (Hare, 1994; Roethlisberger & Dickson, 1939) and has captured the attention of CSCL researchers since the early days (De Wever & Strijbos, 2021; Hare, 1994; Strijbos & De Laat, 2010) resulting in a wealth of research insights about types, methods of identification and possible structuring methods. In the next section, we offer an overview of roles in CSCL, methods for identification of roles and, in particular, Social Network Analysis (SNA). We then review previous research regarding roles using SNA methods as well as the dynamics of changes of roles.

#### 2.1.1. Types of roles

The literature identifies several types of roles across different dimensions. From the viewpoint of structuring, two types of roles are commonly described: a scripted role with predefined tasks and duties (Kollar et al., 2006), and an emergent role that spontaneously emerges during the collaborative process regardless of teacher involvement (Weinberger et al., 2010). Another dimension of roles exists on the product vs. process level. A product-oriented role is concerned with the delivery of the product or accomplishing the requirements of a task. For instance, a student with the role of *scribe* in PBL would record the main points of discussion and the learning agenda. A process-oriented role is more concerned with facilitating the collaborative process. For example, a student with a *leader* role would assist a productive discussion, invite participants to contribute and may manage conflicts (Strijbos & De Laat, 2010).

Roles may also exist at different granularity levels (*micro*, *meso* and *macro*). The micro-level ("role as a task") occurs when the role pertains to a single task or a piece of work within a limited time. A large corpus of CSCL literature has studied the "role as a task" due to the popularity of structuring the collaborative process with scripts in primary education (Strijbos & De Laat, 2010). The meso-level ("role as a pattern") manifests within a set of multiple collaborative tasks, whereas the macro-level ("roles as a stance") exists on a more general level, i.e., as an attitude or orientation towards collaborative learning. The *stance* offers a basis for explaining or understanding how the collaborators execute a task (micro-level) or group of tasks (meso-level). For instance, a student who does not value the collaborative process is likely to be less engaged and assume a role of a passive participant in a collaborative task (Strijbos & De Laat, 2010).

A stance may explain behavior in some situations; however, a stance does not necessarily lead to the display of a certain behavior

**Table 1**
Review of roles identified and methods used in the CSCL literature.

| Reference | Context | Methods | Centralities | Roles |
|---|---|---|---|---|
| Aviv et al. (2003) | 2 courses (35 students, 2 teachers) | Network equivalence analysis and visual inspection | | Guide<br>Bridging and triggering roles<br>Lurkers |
| Temdee et al. (2006) | 1 course (46 students) | Maximum Leadership index in the group | Degree<br>Closeness<br>Betweenness | Leader |
| Laghos and Zaphiris (2007) | 1 course (618 students) | Structural equivalence analysis and visual inspection | Centrality in-degree<br>Centrality out-degree<br>Neighbor in-degree<br>Neighbor out-degree | Role groups R1 – R4 |
| Stuetzer et al. (2011) | 120 discussion boards (834 participants) | Classification of roles according to predefined thresholds | Degree<br>Betweenness<br>Weight<br>Eigenvector | Sightseer<br>Cosmopolitan<br>Broker<br>Individualist<br>Alpha Dog |
| Jimoyiannis et al. (2013) | 1 course (44 students, 1 teacher) | Qualitative evaluation | Degree | Leaders<br>Connectors<br>Peripheral members<br>Lurkers |
| Stuetzer et al. (2013) | 834 participants in a shared collaborative learning platform (11 universities) | Classification of roles according to predefined thresholds | Degree<br>Betweenness<br>Weighted degree<br>Eigen | Broker |
| Chen and Chang (2014) | 1 course (33 students) | Classification of roles according to predefined thresholds | In-degree<br>Out-degree<br>"Social score"<br>"Interactive score" | Hub<br>Source<br>Sink<br>Island |
| Marcos-García et al. (2015) | 1 course (46 students, 1 teacher) | Classification of roles according to predefined thresholds | Out-degree<br>Out-closeness<br>Out-centralization<br>Density<br>In-degree<br>In-closeness<br>Centralization<br>Power<br>Betweenness | Students<br>Leader<br>Coordinator<br>Animator*<br>Active<br>Peripheral<br>Quiet<br>Missing<br>Teachers<br>Guide<br>Facilitator*<br>Observer |
| Medina et al. (2016) | 1 course (18 students) | Indirect blockmodeling | Automorphic equivalence<br>Structural equivalence | Leader<br>Coordinator<br>Active<br>Peripheral<br>Missing |
| Saqr et al. (2018) | 3 courses (82 students, 3 teachers) | Classification of roles according to predefined thresholds | Degree<br>In-degree<br>Out-degree<br>Betweenness<br>Closeness<br>Information | Leader<br>Coordinator<br>Active participatory<br>Active Non-participatory<br>Peripheral |
| Xie et al. (2018) | 1 course (57 students, 1 teacher) | Clustering of leadership changes using k-means | Katz centrality | Leader |
| Dowell et al. (2019) | 1 course (2429 students, 1 teacher) | Clustering of roles using k-means | Participation<br>Internal Cohesion<br>Responsivity<br>Social Impact<br>Newness<br>Communication Density | Chatterers<br>Driver<br>Follower<br>Lurker<br>Detached<br>Influential |
| Kim and Ketenci (2019) | 3 course Sect. (56 students) | Classification of roles according to predefined thresholds and confirmation with k-means clustering | In-degree<br>Out-degree<br>Betweenness | Full participants<br>Inbound participants<br>Peripheral participants |
| Ouyang and Chang (2019) | 1 course (120 students) | Classification of roles according to predefined thresholds (Social participatory role detection method) | Out-degree<br>Out-closeness<br>In-degree | Leader<br>Starter<br>Influencer |

(*continued on next page*)

**Table 1** (*continued*)

| Reference | Context | Methods | Centralities | Roles |
|---|---|---|---|---|
| | | | In-closeness<br>Betweenness | Mediator<br>Regular<br>Peripheral |
| Saqr and Viberg (2020) | 1 course (129 students, 5 teachers) | Clustering of roles using k-means | Degree<br>Betweenness<br>Closeness<br>Diffusion<br>Cross-clique connectivity | Leaders<br>Arbitrators<br>Satellites |
| Turkkila and Lommi (2020) | 1 course (11 students) | Functional role count | Katz-centrality | 1-sink, 1 source, 1-recip, 2-source, 2-sink, Relay, Relay & sink, Relay & source |
| Saqr and López-Pernas (2021b) | 69 courses (3277 students) | Clustering of roles using k-means | Weighted Degree<br>Betweenness<br>Closeness<br>Eigenvector<br>Neighborhood<br>Coreness Total<br>Cross Clique Connectivity<br>Diffusion Degree Total | Influencers<br>Mediators<br>Isolates |

consistently, i.e., assuming the same role in subsequent courses (Bair, 2017). Instead, a *disposition* is what drives an individual to exhibit similar patterns of behavior in most situations (Katz, 1993). A disposition can be conceived as an enduring behavioral attitude that results in frequently acting or exhibiting the same behavior (Ajzen, 1987; Claxton & Carr, 2004), e.g., having certain social dispositions would lead a student to assume a coordinator role in most —if not all— courses. The previous literature has extensively studied roles as a task, or as a stance (De Wever & Strijbos, 2021; Strijbos & De Laat, 2010) (see section 1.1.2). Yet, little is known about the more enduring participatory disposition which would manifest as a disposition to assume similar roles in a consistent manner, e.g., a leader's tendency to assume a leadership role in most collaborative tasks, a gap which is the main motivation for our study. We use computational methods to identify roles, use longitudinal modeling to trace the unfolding of roles and explain students' different patterns of assuming roles.

#### 2.1.2. *Identification of roles in CSCL*

Coding of students' interactions and identification of the roles through qualitative analysis was the common method during 1990–2010 (Strijbos & De Laat, 2010). The roles identified with qualitative analysis were diverse with relatively contextualized labels (e.g., emphatic, editor and project planner). Today, the increasing use of digital tools and the growing data sizes make such coding impractical. Researchers have sought computational methods that could facilitate the analysis of large amounts of data, offer objective measures, and help automate the process. Some examples of such methods are learning analytics and SNA (Saqr & López-Pernas, 2021b; Xie et al., 2018).

SNA is particularly useful for the analysis of relations, interactions, and communities, offering a practical and effective solution for automating the analysis of large volumes of data. In SNA, students are represented as "nodes", and interactions (e.g., posts, replies, comments, or messages) are represented as "edges". In other words, the edge is a representation of the interaction between the source and the target of the interaction. A *network* is constructed by aggregating all of the edges which, in turn, can be studied by means of SNA (Borgatti, 2005; Borgatti et al., 2009). SNA has a rich visual and mathematical toolset that could help automate effortful analysis tasks. Visualization helps map the interactions and the relationships among actors as well as the full network of interactions. Mathematical analysis offers "measures" for quantifying students' position or importance in the networks known as centrality measures (Borgatti et al., 2009). Since importance can be viewed in different ways, there are several centrality measures that quantify, e.g., volume of interactions (degree centrality), distance to collaborators (closeness centrality), connecting collaborators (betweenness centrality), or strength of connected collaborators (eigenvector centrality) (Borgatti & Brass, 2019; Saqr et al., 2022). As such, researchers have capitalized on SNA potentials to identify roles within CSCL environments using visualization and centrality measures. For instance, Aviv et al. (2003) used Eigenvector centrality, network equivalence and network visualization to identify students' roles in a business ethics course. The authors described bridging and triggering roles, guide roles, and lurker roles. A similar approach was taken by Temdee et al. (2006) who used degree, closeness, and betweenness centralities to identify student leaders. Such early trials have proven fruitful, leading to a constant stream of SNA research about roles (Kim & Ketenci, 2019; Saqr & López-Pernas, 2021b; Xie et al., 2018).

To identify roles using SNA, researchers have used several methods that include visualization, manual classification, and algorithmic methods, e.g., clustering. Classifying students according to their interaction patterns using manual thresholds or range of values set by the researcher(s) has been the object of several studies (Chen & Chang, 2014; Marcos-García et al., 2015; Saqr et al., 2018; e.g., Stuetzer et al., 2011, 2013). Such an approach is prone to subjectivity and less likely to be automated. To address the subjectivity and automation challenges, structural equivalence (e.g., students with similar interaction profile) with SNA centralities have been explored to group students into similar groups of roles (Laghos & Zaphiris, 2007; Medina et al., 2016). Recently, algorithmic identification of roles by means of unsupervised clustering such as k-means have gained momentum with the increasing adoption of learning analytics methods and the interest in automation and scalability (Dowell et al., 2019; Kim & Ketenci, 2019; Saqr & López-Pernas, 2021b; Saqr & Viberg, 2020; Xie et al., 2018). Unsupervised clustering is commonly performed by using students' centralities as an input for an algorithm. Thereafter, the algorithm tries to identify the roles based on shared similarities or the lack thereof among the included students (Dowell et al., 2019; Kim & Ketenci, 2019).

Compared to roles identified with qualitative coding, the roles identified through SNA are less contextual and less diverse. Three common roles are prevalent in the SNA literature. First, a (1) leader role, where the student invests significant effort and plays an important role in driving the interactions, exists under several labels: influencers, full participants, guides, and hubs. An (2) isolate role is also commonly identified, where the student invests minimum effort in the collaborative process. Such a role exists in several studies under different labels which depict similar characteristics e.g., lurker, peripheral, missing, or detached. Lastly, an (3) intermediate role that acts as a mediator or moderator exists in many studies with different labels as well (e.g., arbitrators, facilitator, connectors, relays). See Table 1 for a review of the existing literature on role identification in CSCL where we list each study, the roles identified, the used centrality measures and the methods of identification.

### *2.1.3. Studies addressing role dynamics using computational methods*

Studies addressing the dynamics of roles with computational methods are rare and have mostly studied the dynamics of roles within the same course. For instance, Laat et al. (2007) used SNA —as well as qualitative analysis— to study the dynamics of teachers taking the leading roles versus students leading the conversations across time. Skrypnyk et al. (2015) studied the dynamics of roles in a Massive Open Online Course (MOOC) where they reported the variability of centrality over the duration of a MOOC, some MOOC users assumed the facilitator's role, and others maintained their assigned roles. Additionally, there was marked variability at the end of the course. A similar approach using variations of centrality measures to track leadership changes was reported by Xie et al. (2018). The authors found different clusters of leadership changes, where some students maintained high leadership across the course, and others assumed leadership roles only when assigned the role of a moderator. Socio-semantic block modeling was used by Hecking and Chounta (2017) to study different clusters of roles in a MOOC, the authors reported the progression of roles with the means of sequence mining, the results showed that most of the subgroups of roles were active only during the early weeks, and one active group maintained activity at all the time. A similar approach that used sequence mining to track the progression of roles was also taken by Boroujeni et al. (2017). The authors reported high fluctuation between active roles and passive roles. Saqr et al. (2018) studied the changes of roles across two midterms as a response to an intervention and reported more students taking active roles as a response to the intervention. A common pattern among the previous studies is they either tracked the variability of roles within a course, or studied more than a course with limited tracking of the progression of roles across the courses, e.g., 2 courses (Aviv et al., 2003), 3 courses (Saqr et al., 2018); (Kim & Ketenci, 2019), or 69 courses (Saqr & López-Pernas, 2021b). This study aims at filling this gap, i.e., track the changes of roles across courses, the longitudinal trajectories of roles, as well as the characteristics of each trajectory and the predictors thereof.

Table 1 presents search results within the *Scopus* database using the keywords "social network analysis" and "role" and "collaborative learning" OR "CSCL" within the titles, abstracts, or keywords. The results were limited to original articles in English. A total of 110 articles were retrieved. The two authors scanned them based on titles and abstracts, resulting in 36 articles included. A full text analysis returned 18 papers that were tabulated according to SNA measures, methods of identification, and the reported roles along with the number of students and courses.

## 3. Methods

### *3.1. Context*

The study was based on a healthcare program which integrates basic and clinical sciences in all subjects. While each course focuses on a different subject, all courses share a common pedagogical foundation based on PBL and are taught using a blended learning environment (online and face-to-face). They all have similar activities (lectures, seminars, practical sessions and PBL sessions) as well as similar evaluation methods. The courses are arranged sequentially and therefore referred to as blocks, i.e., each course starts after the previous course ends with no overlap. The PBL process has three components: the students, a teacher, and a small group. The group is composed of an average of ten students with a teacher. As the name suggests, PBL is built around students discussing a *problem.* The *problem* is open-ended and ill-structured by design, based on complex life situations that aim to trigger students' learning, rather than solving the problem. For instance, a problem that is designed to stimulate students learning about bleeding could include a middle-aged pregnant woman with a bleeding problem, who was hit by a car and required surgery and anesthesia upon arrival at the hospital. Students have to address the complexities of bleeding, pregnancy and anesthesia drugs as well as reflect upon the priorities of treatment during *problem* discussions.

The PBL process is structured into three phases. First, students attend a two-hour face-to-face session at the beginning of the week where they read the *problem*, discuss new terminology, and together identify the learning objectives that they have to study during the week. Right after the first face-to-face session, i.e., on the first day of the week, students start working online. In the online PBL, students engage in interactions regarding the problem, discuss learning goals, share learning resources, suggest solutions and alternatives, and together try to arrive at final conclusions. Students are also expected to reflect on their learning, group processes and each other's contributions. This phase is continuous all over the week and is facilitated by a teacher who is supposed to be on the side if the group is engaging in productive discussion and may intervene to encourage participation or help resolve a conflict. Students assume roles based on their disposition or attitude toward the PBL process and therefore, the roles in our context are emergent roles. On the last day of the week, there is a face-to-face session where students conclude their week by discussing their learning and reflecting on the group processes. As such, the data for the online PBL were retrieved from the LMS from all courses with an online PBL module in the healthcare program. We sought to include students who completed 10 courses in the program, and we found 329 students who fulfilled such inclusion criteria. Students come from multiple batches; yet the courses, teachers and curriculum remain essentially the same.

### *3.2. Data collection and preparation*

Data for all forum posts in these ten courses were retrieved from the learning management system database. The collected data included post ID, thread ID, group ID, course ID, student ID, teacher ID, forum thread ID, parent thread ID, timestamps for each post as well as course enrollments. The retrieved data were used to create a post-reply network where an edge (A- > B) was constructed from the post author or source (A) to the replied-to or target (B). All such edges were aggregated to build a weighted directed network for each group (i.e., each group had a separate network) where the weight represented the number of interactions from A to B (Poquet & Jovanovic, 2020; Saqr et al., 2020).

### *3.3. Data analysis*

#### *3.3.1. Network analysis*

The analysis was performed using the R programming language version 4.02 (R Core Team, 2018). The centrality measures were computed using the *igraph* and *centiserve* R packages (Csardi & Nepusz, 2006, p. 1695; Jalili et al., 2015). Since groups differed in number, centrality measures were normalized (divided by group size - 1) (Saqr et al., 2020).

The centrality measures selected for this study were based on the previous literature that relied on SNA to identify roles (see Table 1). Three aspects were captured using SNA: the participation dimension of CSCL (Weinberger & Fischer, 2006), the social dimension (Kreijns et al., 2013), and the diffusion and uptake of ideas (Saqr & López-Pernas, 2021b). The centrality measures were also based on the frameworks proposed by Strijbos and De Laat (2010) to capture the high and low effort dimensions, and by Driskell et al. (2017) to capture the sociality, dominance and task engagement dimensions. Therefore, three types of centralities were selected: (1) degree centrality reflecting students participation (e.g., Chen & Chang, 2014; Kim & Ketenci, 2019; Marcos-García et al., 2015), (2) closeness, betweenness and eigenvector centralities reflecting social positioning, influence, and involvement in the discussions (Marcos-García et al., 2015; Ouyang & Chang, 2019; Temdee et al., 2006), and (3) diffusion and cross clique centralities reflecting the diffusion, uptake, and reply-worthiness of posts (Saqr & López-Pernas, 2021b). The neighborhood size was also calculated to measure the sphere of interactions in terms of the number of collaborators that the student helped engage. The full details of these centralities, methods of measurement and operationalization are as follows:

- **Weighted Degree**: Total number of interactions (posts or forum contributions) that a student posts or receives. Weighted degree is operationalized as students' participation, effort and contribution to the collaborative discourse, and social positioning (Borgatti, 2005; Vignery, 2022; Wise & Cui, 2018).
- **Closeness centrality:** Inverted distance between a student and all other collaborators in the network. The shorter the distance, the higher the closeness, reachability, proximity of access to learning resources and diversity (Bae & Kim, 2014; Hernández-García et al., 2015; Vignery, 2022).
- **Betweenness centrality:** Number of times a student connected two other students (lied on the shortest path). It is operationalized as mediating interactions, bridging threads, and communities as well as access to different perspectives (Bae & Kim, 2014; Borgatti & Brass, 2019; Saqr et al., 2022).
- **Eigen centrality:** While degree centrality counts the immediate interactions, Eigen centrality considers the strength of connected collaborators and, therefore, reflects the student's network strength. It is operationalized as a reflection of the strength of the social network, connectedness, and range of influence (Hernández-García et al., 2015; Vignery, 2022; Wise & Cui, 2018).
- **Diffusion degree:** Sum of the diffusion probability of a student's contributed posts (i.e., interaction, replies or uptake by others) as well as collaborators' diffusion probabilities. Influential, reply-worthy, or posts with promising ideas will stimulate more discussions and involvement of collaborators (Banerjee et al., 2013; Saqr & Peeters, 2022).
- **Cross-clique Connectivity:** Number of cliques that the student belongs to. Cliques reflect strong ties, rich replies, and contributions from close collaborators. Higher values reflect involvement, ability to influence, and strength of collaborators (Joksimovic et al., 2016; Saqr & López-Pernas, 2021b).
- **Neighborhood:** Size of connected collaborators, reflecting the range of reach or influence and the extent of reply-worthy contributions. A student with valuable contributions can help engage other collaborators and longer threads of interactions (Liu et al., 2016; Saqr & López-Pernas, 2021b; Wang et al., 2017).
- **Leverage centrality:** Difference between the teacher's number of connections and the average number of connected students. This centrality was calculated for the teacher only to reflect their relative interactivity. A teacher with positive leverage has more interactions than an average student in their group. (Joyce et al., 2010).

For each group (network), we calculated the network properties that reflect group size, interaction intensity, and cohesion of the group using the *igraph* R package (Fig. 1A). We report network size in terms of number of nodes (i.e, participants), number of edges (i.

e., interactions or posts), as well as the mean degree (number of interactions divided by the number of collaborators) as an indicator of group activity. We also computed the network density, which is the number of connections divided by the maximum possible. Higher density reflects diverse collaborations, inclusion of collaborators, and group cohesion (Vignery, 2022). Similarly, we computed the centralization degree which reflects the dominance of collaborators (Driskell et al., 2017). Higher values of centralization reflect the presence of few central collaborators and lack of inclusion or participation of group members (Marcos-García et al., 2015).

### *3.3.2. RQ1: identification of roles*

To identify the different roles from the SNA centrality measures of students (Fig. 1B), Latent Profile Analysis (LPA) was performed using the R package *tidyLPA* (Rosenberg et al., 2018). LPA is a finite mixture model (similar to latent class analysis) that enables grouping of individuals into homogeneous subgroups based on their attributes, i.e., centrality measures in our case. LPA assigns individuals a probability of belonging to each subgroup using maximum likelihood. Then, subgroup membership is determined based on the highest probability (Spurk et al., 2020). LPA has some advantages over traditional clustering methods. LPA does not require assumptions regarding linearity or normality of the data. Being model based, the probabilities of belonging to each cluster are estimated directly from the model. Additionally, LPA has several parameters for the selection and evaluation of the model, which helps make model estimation and evaluation rigorous (Nylund-Gibson & Choi, 2018; Weller et al., 2020).

SNA centralities (degree, closeness, betweenness, eigen, diffusion, cross clique, and neighborhood centralities) were used as an

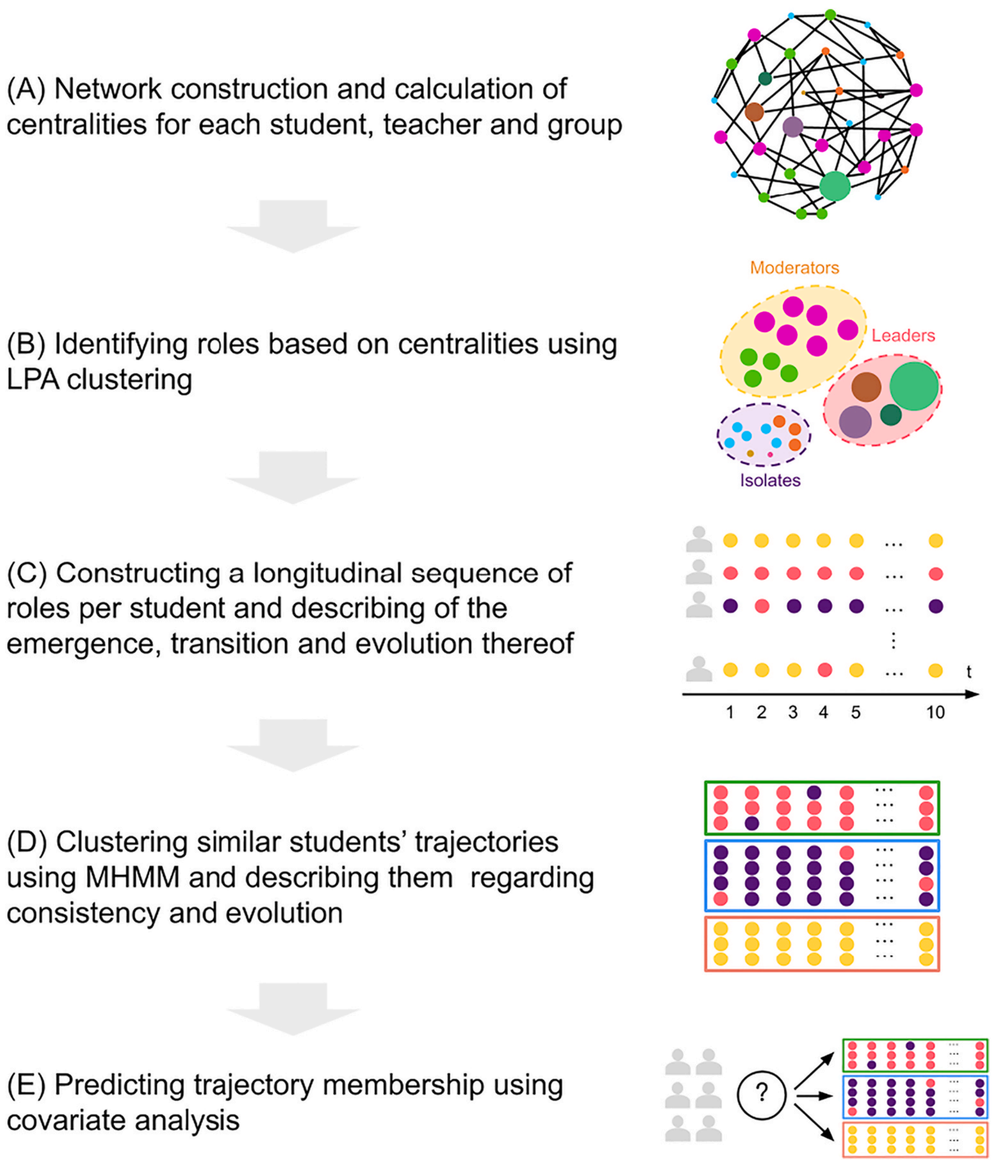


**Fig. 1.** Summary of the data analysis process.

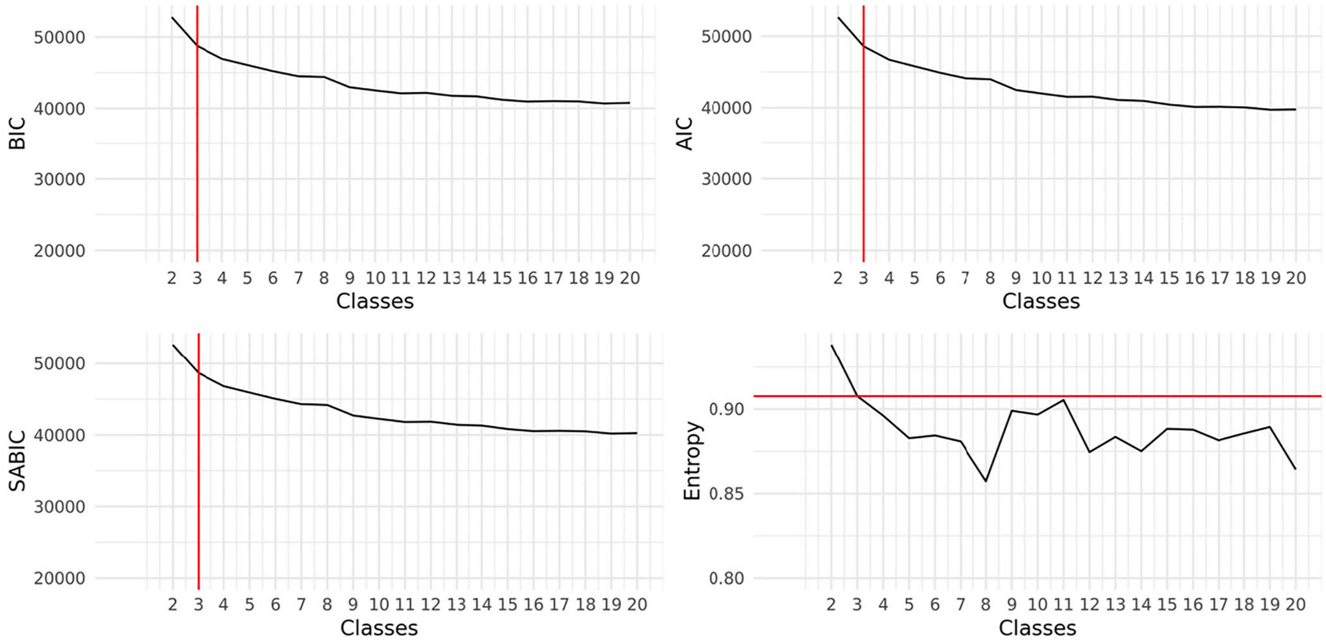


**Fig. 2.** A plot of information criteria (BIC, AIC, SABIC and Entropy) showing largest drop at 3 classes and flattening thereafter.

input to LPA after being standardized (mean subtracted and divided by standard deviation). Standardization was performed since different measures come from different courses and have different measurement scales. The number of classes was chosen based on the latest recommendations by Nylund-Gibson and Choi (2018) and Weller et al. (2020), i.e., by relying on the following information criteria: Akaike Information Criterion (AIC), Bayesian Information Criterion; (BIC), Sample Size-adjusted Bayesian Information Criterion (SABIC), entropy, relative size of the largest class (*Nmax*), and relative size of the smallest class (*Nmin*). Initially, 10 models were estimated. However, as Table S1 (in the appendix) shows, information criteria (AIC, BIC and SABIC) continued to decrease when increasing the number of classes, and the bootstrapped likelihood test (BLRT) was statistically significant for all the tested models. We further estimated an additional 10 models, where information criteria continued to decrease (i.e., there was no global minimum). Therefore, we used the "elbow" method recommended by Nylund-Gibson and Choi (2018), where the number of classes is determined based on the largest drop in information criteria. We chose three classes where AIC, BIC and SABIC had the largest decrease (see Fig. 2). The resulting model had an entropy of 0.91 (well above 0.8) indicating a very good classification of students into classes (Clark & Muthén, 2009). The proportion of the smallest class size was 0.25 within acceptable levels (it is greater than 0.05 and the class has more than 50 cases) (Weller et al., 2020). The average posterior probabilities (*AvePP*) for each class were (0.94, 0.98, 0.96), indicating very high class assignment certainty (Nylund-Gibson & Choi, 2018).

A means comparison test, i.e., Analysis of Variance (ANOVA), was performed to test how different the classes are from each other as well as to get an idea about the separation between the clusters (Clark & Muthén, 2009). ANOVA assumptions were checked and fulfilled: homogeneity of variances was tested using Levene's test, and Shapiro-Wilk's normality test was used to check the normality of variables (Vieira, 2011). Both tests were statistically insignificant. Eta-squared ($\eta^2$) was used to measure the effect size, where $\eta^2 = 0.01$ indicates a small effect; $\eta^2 = 0.06$ indicates a medium effect; $\eta^2 = 0.14$ indicates a large effect (Jacob Cohen, 1988). Pairwise comparisons were performed using the Scheffé test with Holm's correction for multiple testing (Holm, 1979). An important assumption of LPA is conditional independence —oftentimes referred to as local independence— which implies the latent class explains the correlation among the observed outcomes, and therefore, it is expected that items within the same class are uncorrelated, i.e., have no residual correlation (Nylund-Gibson & Choi, 2018). We checked each class for all possible pairwise correlations and the results were either statistically insignificant or below recommended values (Dormann et al., 2013).

### 3.3.3. RQ2: trajectories of roles across the program

The identified roles (RQ1) were used to build a time-ordered sequence according to the sequence of ten courses included in this study (Fig. 1C). As such, each student in our dataset has a sequence of ten roles: one for each course in which they were enrolled. An example for a hypothetical student who was a leader at course 1, leader at course 2, mediator at course 3, mediator at course 4, and so on would look as follows:

Leader - Leader - Mediator - Mediator - Isolate - Mediator - Leader - Isolate - Isolate - Leader.

The ordered roles were used to build a state sequence object using the TramineR package (Gabadinho et al., 2011). TramineR is an open source software with a rich toolset for the analysis and visualization of sequence data, longitudinal events, and life trajectories in social sciences (Gabadinho et al., 2011; Helske et al., 2018), and education (López-Pernas et al., 2021; Matcha et al., 2020; Saqr & López-Pernas, 2021a). The resulting sequence was plotted using an index plot, which visualizes each student's sequence of roles as color-coded stacked bars corresponding to the temporal succession of roles over the program duration (Gabadinho et al., 2011).

To identify possible "latent trajectories", i.e., subgroups of students who share a similar longitudinal pathway (i.e., ordered succession) of roles over the program (Fig. 1D), we used the Mixture Hidden Markov Model (MHMM) implemented in the R package *seqHMM* (Helske & Helske, 2019). MHMM enables clustering of longitudinal life-events data (Helske & Helske, 2019) and educational data (López-Pernas & Saqr, 2021). MHMM has three advantages: being model-based; allowing covariates and suited for the temporal nature of the data. Ten MHMM models were estimated, and the model with the lowest BIC values (three classes) was selected. See Table S2 (in the appendix) for the complete set of BIC values. Each model was estimated 1000 times to make sure we obtain the global optimum (Helske & Helske, 2019).

Each of the resulting classes (or trajectories) was plotted using index and distribution plots, described, and labeled according to the roles' distribution. Each of the trajectories was analyzed according to the frequency of roles, the transitions between them, and the consistency thereof. The consistency of the trajectories across students was studied with between-student entropy, which measures

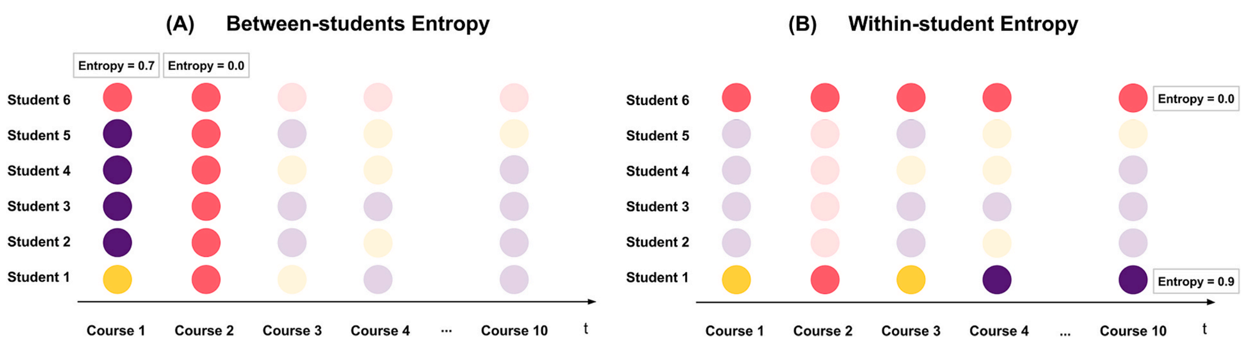


**Fig. 3.** Between-students entropy vs. within-student entropy.

how variable the roles are at each time point (course). The values of between-student entropy are expected to be 0 when all roles are similar in the trajectory (Fig. 3A).

### 3.3.4. RQ3: the unfolding of roles within each trajectory

To answer the question of how long a student assumes a role, consistently stays in such role and how they conclude the program, three measures were computed, namely the mean duration in each role (mean spell duration); the within-student entropy, i.e., how consistent each student was in assuming roles along the program compared to oneself (Fig. 3A), and the *integrative capacity,* or capability to assume a favorable role (i.e., *Leader* or *Mediator*) and continue to assume this role until the end. The trajectories were compared regarding these measures (spell duration, entropy and integrative capacity) using the Kruskal–Wallis test (Ostertagova et al., 2014) since ANOVA assumptions were violated. Epsilon-squared ($\epsilon^2$) effect size was calculated and interpreted according to Cohen's (1992) guidelines (very small: $\epsilon^2 < 0.02$; small: $0.02 \leq \epsilon^2 < 0.13$; medium: $0.13 \leq \epsilon^2 < 0.26$; large: $\epsilon^2 \geq 0.26$).

### 3.3.5. RQ4: trajectory predictors

Variables that could explain group membership were used as covariates in the model, i.e., group-related variables and teacher-related variables (Fig. 1E). For the group variables we included: (1) the mean degree, i.e., the number of interactions by all group members divided by group count (excluding the given student) to reflect group interactivity; (2) centralization of degree to reflect the distribution of interactions and dominance of group members; (3) network density to reflect group cohesion, inclusion of members and range of diversity of participation, and (4) node count to control for the number of students in the group. At the teacher level, we used the variables that reflect teacher interactivity (i.e., mean degree); range of influence of interactions and how students interacted with the teacher posts (i.e., diffusion degree), balance of teacher interactions to the group members (i.e., leverage degree), and neighborhood which reflects the number of students that the teacher interacted with. The covariate analysis followed the latest recommendation of Nylund-Gibson and Choi (2018).

## 4. Results

The full dataset included 183,916 interactions. The data were filtered to include only students who completed 10 sequential courses ($n = 329$) corresponding to 4 years of education (3290 course enrollments). The total number of interactions included in the study was 84,597 (70,296 by students and 14,283 by teachers). The mean number of posts sent or received per course by each of the students was 41.27 and by each teacher 80.13 (see Table 2). Each student interacted with a mean number of 5.28 others (SD = 3.31), while each teacher interacted with a mean of 5.31 students. The mean number of members in each group (students and teachers) was 11.13, with a mean number of 237.23 interactions, and a mean density of 0.42. See Table S3 in the appendix for the number of interactions per course by students, and Fig. S1 for the histogram of message length distribution.

### 4.1. RQ1: identification of roles

Clustering of students using centrality measures resulted in three distinct profiles, each representing a different role (Fig. 4). The three identified roles were labeled according to the centrality values as "leaders", "mediators" and "isolates". Summary statistics of the centrality measures for each identified role can be seen in Table S5 in the appendix. The detailed profile of these roles was as follows:

- **Leaders** (n = 807, 24.5%): *Leaders* had the highest values of degree centrality (mean degree = 71.1) indicating that they contributed and received the highest number of contributions, and that they invested the highest effort across the three roles. They communicated with more students (mean neighborhood = 8.7) and were close to more students (mean closeness = 0.9). They had

**Table 2**
Summary statistics of students' centrality measures, teachers', and group networks. Normalized statistics can be found in Table S4 in the appendix.

| | Measure | Mean | SD | 25% | Median | 75% |
|---|---|---|---|---|---|---|
| Students | Degree | 41.27 | 35.87 | 16.00 | 31.00 | 54.00 |
| | Closeness | 0.69 | 0.21 | 0.55 | 0.69 | 0.88 |
| | Betweenness | 0.03 | 0.05 | 0.00 | 0.01 | 0.04 |
| | Diffusion Degree | 0.18 | 0.16 | 0.07 | 0.14 | 0.25 |
| | Eigen | 0.39 | 0.31 | 0.13 | 0.30 | 0.60 |
| | Neighborhood | 5.28 | 3.31 | 2.00 | 5.00 | 8.00 |
| | Cross Clique Connectivity | 0.07 | 0.16 | 0.00 | 0.02 | 0.06 |
| Teachers | Degree | 80.13 | 109.94 | 17.00 | 39.00 | 92.00 |
| | Leverage | 0.05 | 0.37 | −0.19 | 0.04 | 0.33 |
| | Diffusion Degree | 0.21 | 0.17 | 0.09 | 0.18 | 0.29 |
| | Neighborhood | 5.31 | 3.56 | 2.00 | 6.00 | 8.00 |
| Groups | Node count | 11.13 | 1.65 | 10.00 | 11.00 | 12.00 |
| | Edge count | 237.23 | 166.65 | 117.75 | 198.00 | 307.00 |
| | Network Density | 0.42 | 0.24 | 0.23 | 0.39 | 0.59 |
| | Centralization Degree | 0.39 | 0.23 | 0.20 | 0.39 | 0.56 |

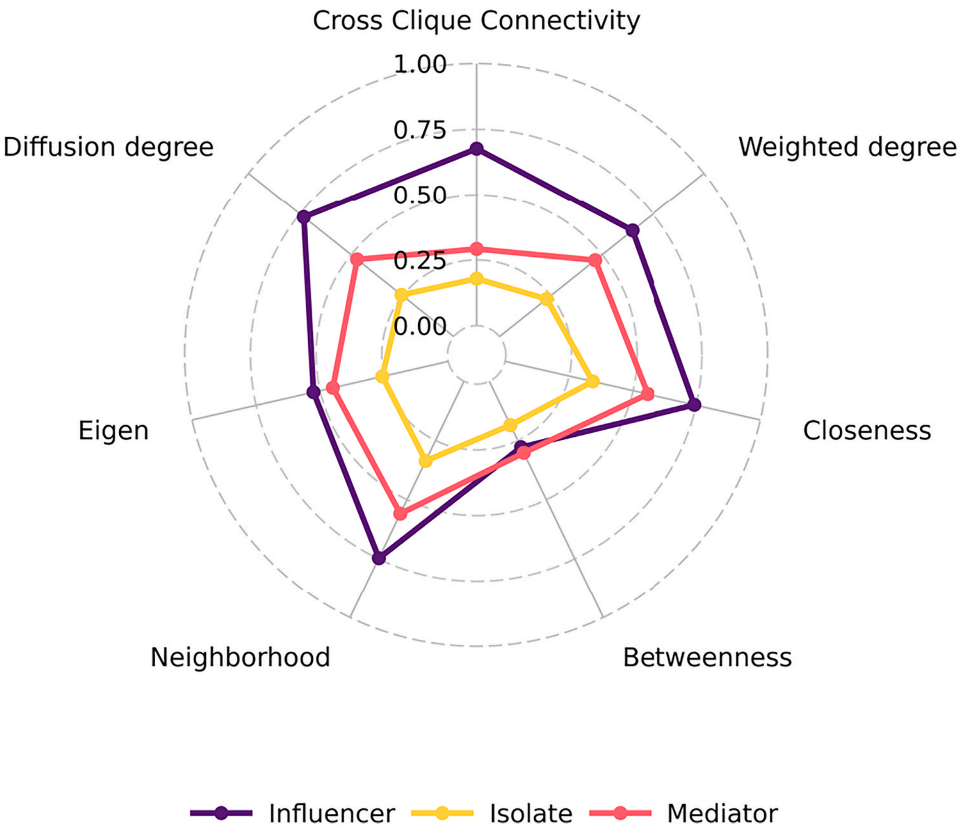


**Fig. 4.** Radar plot of the mean rescaled centrality values for each of the roles.

higher diffusion degrees (mean = 0.328) indicating their contributions were influential and more likely to be endorsed by their peers. However, *Leaders* had comparable betweenness (mean = 0.035) and eigen centrality (mean = 0.58) values to the *Mediators*.

- **Mediators** (n = 1019, 31%): *Mediators* had intermediate degree and diffusion values (mean degree = 45.9), mean diffusion = 0.197). They were close to an average number of students (mean neighborhood = 5.7; mean closeness = 0.7) and had intermediate eigen centrality values (mean closeness = 0.7). *Mediators* had the highest betweenness centrality values (mean = 0.046), indicating they "bridged" or relayed more interactions between students compared to the two other roles.
- **Isolates** (n = 1464, 44.5%): As the name may imply, *Isolates* had the lowest centrality values compared to the two other roles indicating low effort. *Isolates* contributed the least (mean degree = 21.6), interacted with a lower number of students, and were close to fewer students (mean neighborhood = 3; mean closeness = 0.6). *Isolates* were less likely to have influential posts and their contributions had the lowest values of diffusion (mean = 0.098) and betweenness (mean = 0.017).

A means comparison test (ANOVA) was performed to compare roles, see Table S3 for detailed statistics, and showed that difference among the roles regarding strength centrality was statistically significant and had a large effect size ($F (2, 3287) = 732.97$, $p < 0.001$; $\eta^2 = 0.31$, 95% CI [0.29, 1.00]). So were the differences among roles regarding closeness ($F (2, 3287) = 1296.77$, $p < 0.001$; $\eta^2 = 0.44$, 95% CI [0.42, 1.00]), Eigen centrality ($F (2, 3287) = 516.66$, $p < 0.001$; $\eta^2 = 0.24$, 95% CI [0.22, 1.00]), neighborhood ($F (2, 3287) = 1350.99$, $p < 0.001$; $\eta^2 = 0.45$, 95% CI [0.43, 1.00], and cross clique centrality ($F (2, 3287) = 702.81$, $p < 0.001$; $\eta^2 = 0.30$, 95% CI [0.28, 1.00]). However, the differences regarding betweenness centrality were statistically significant but small ($F (2, 3287) = 100.82$,

**Table 3**
Results of the ANOVA comparison of mean centrality measures across identified roles.

| Centrality | Sum of Squares | df | Mean square | F | p | $\eta^2$ | $\omega^2$ |
|---|---|---|---|---|---|---|---|
| Strength | 1,305,054.984 | 2 | 652,527.492 | 732.975 | <.001 | 0.308 | 0.308 |
| Betweenness | 0.539 | 2 | 0.270 | 100.823 | <.001 | 0.058 | 0.057 |
| Closeness | 66.498 | 2 | 33.249 | 1296.773 | <.001 | 0.441 | 0.441 |
| Eigen | 74.059 | 2 | 37.029 | 516.659 | <.001 | 0.239 | 0.239 |
| Diffusion degree | 27.757 | 2 | 13.878 | 807.937 | <.001 | 0.330 | 0.329 |
| Neighborhood | 16264.967 | 2 | 8132.483 | 1350.990 | <.001 | 0.451 | 0.451 |
| Cross Clique Connectivity | 23.975 | 2 | 11.988 | 702.814 | <.001 | 0.300 | 0.299 |

$p < 0.001$; $\eta^2 = 0.06$, 95% CI [0.05, 1.00]). Pairwise comparisons using the Scheffé test were all statistically significant with Holm's p-value adjustment. The previous results confirm that the clustering of students' centrality measures resulted in well-separated roles, in which the differences were large and statistically significant on most of the measures. See Table 3 for ANOVA results and effect sizes.

### 4.2. RQ2: trajectories of roles across the program

The roles identified in the previous step were used to construct a state sequence object in which the students' roles were stacked sequentially as a pathway (i.e., succession of roles) across the ten courses and visualized using an index plot (Fig. 5). In the index plot, each student's pathway is represented as a horizontal sequence of ten colored blocks (where each block represents a role at a given course).

During the ten courses, *Isolates* were the largest subgroup of students —ranging from 41% to 48% of all students at any time point—, followed by the *Mediators* (29%–35%) and the *Leaders* were the lowest represented subgroup (21%–29%). The sequences of students' roles in Fig. 6 show a pattern where the top of the plot is mostly dominated by *Leaders*; *Isolates* dominate at the bottom, and *Mediators* occupy the intermediate area in between. Therefore, we applied clustering to discover the subgroups (i.e., clusters) of students who had similar pathways of roles and, therefore, can be considered a distinct "latent trajectory" that shares similar succession of roles. The best MHMM model suggested three clusters which we plotted, labeled, and described below based on their defining characteristics..

- ***Intense* trajectory** (n = 103, 31.3%): Students in the *Intense* trajectory were most of the time *Leaders* (64%), less likely to be *Mediators* (29%), and very unlikely to be *Isolates* (7%) at any given course. As the distribution plot shows in Fig. 6, the lowest percentage of *Leader* roles was at the first course, in which 53% of the students assumed the leader role. The ratio of *Leaders* continued to increase, reaching 79% at the second course, and later decreased during the last four courses. Students in the *Intense* trajectory were more likely to continue to assume a *Leader*'s role between consecutive courses (70% probability), transition to *Moderators* (24%), and were unlikely to plummet to *Isolates* (6%). Between-student entropy —a measure of consistency and homogeneity among the roles in each course of the trajectory— had a mean of 0.74, indicating above average consistency of the trajectory (Fig. 9A).
- ***Fluctuating* trajectory** (n = 93, 28.3%): Students in the *Fluctuating* trajectory were likely to have a *Mediator* role (52%), become *Isolate* (34%), and unlikely to assume the role of *Leaders* (14%). Whereas these ratios look seemingly stable in the distribution plot (Fig. 7 - left), this trajectory had the highest variability, evident in the highest entropy values (mean = 0.9), frequent transitions between roles, and an obvious heterogeneity in the index plot; hence the label *Fluctuating*. Students who had the *Mediator* role transitioned to *Isolates* (27%), to *Leaders* (13%), or remained *Mediators* (60%).

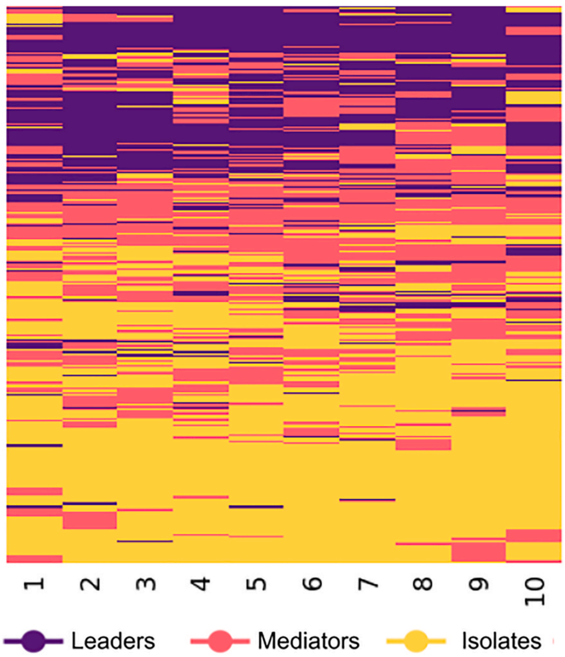


**Fig. 5.** Index plot representing the sequence of roles of the 329 students, each horizontal line represents a single student, each color represents a role, and the x-axis represents the order of courses (1:10). The sequences were sorted using similarity between students based on the longest common subsequence (LCS). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

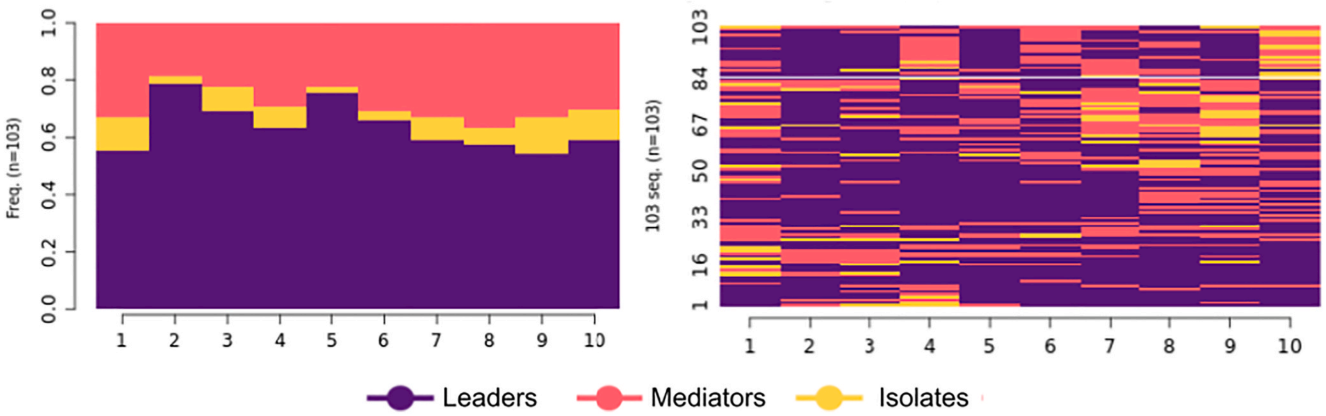


**Fig. 6.** Distribution plot (left) and index plot (right) of the sequence of roles taken by students in the *Intense* trajectory.

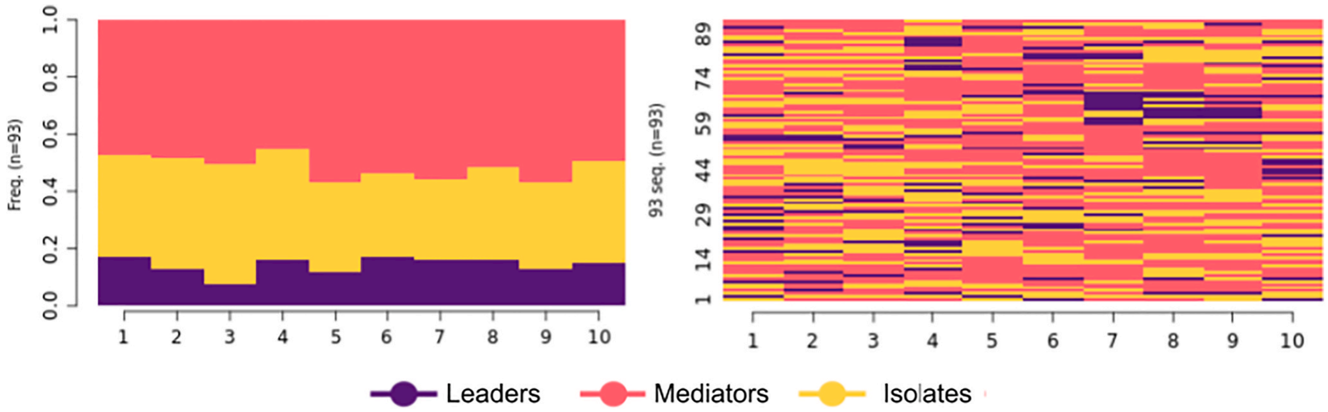


**Fig. 7.** Distribution plot (left) and index plot (right) of the sequence of roles taken by students in the *Fluctuating* trajectory.

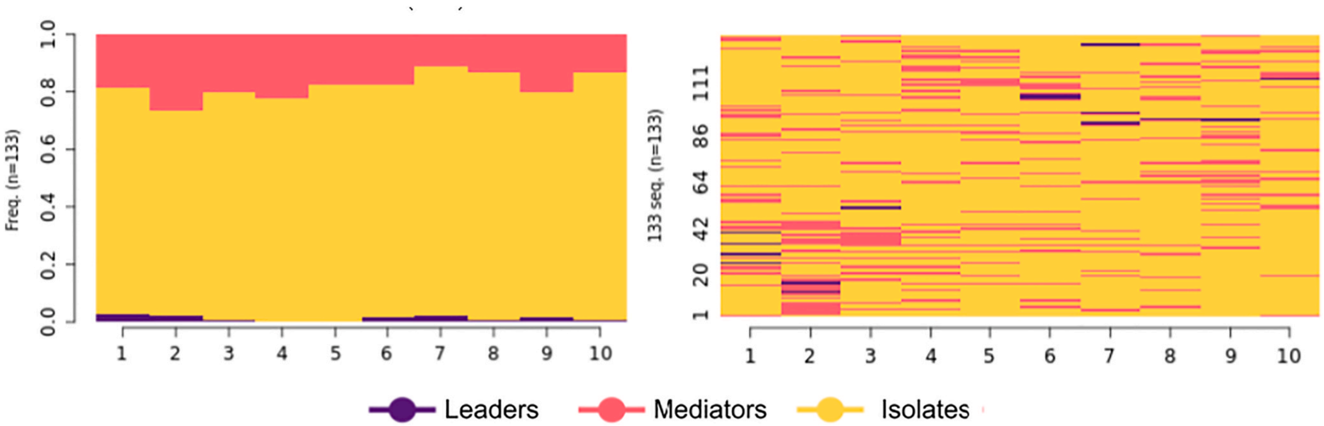


**Fig. 8.** Distribution plot (left) and index plot (right) of the sequence of roles taken by students in the *Wallowing-in-the-mire* trajectory.

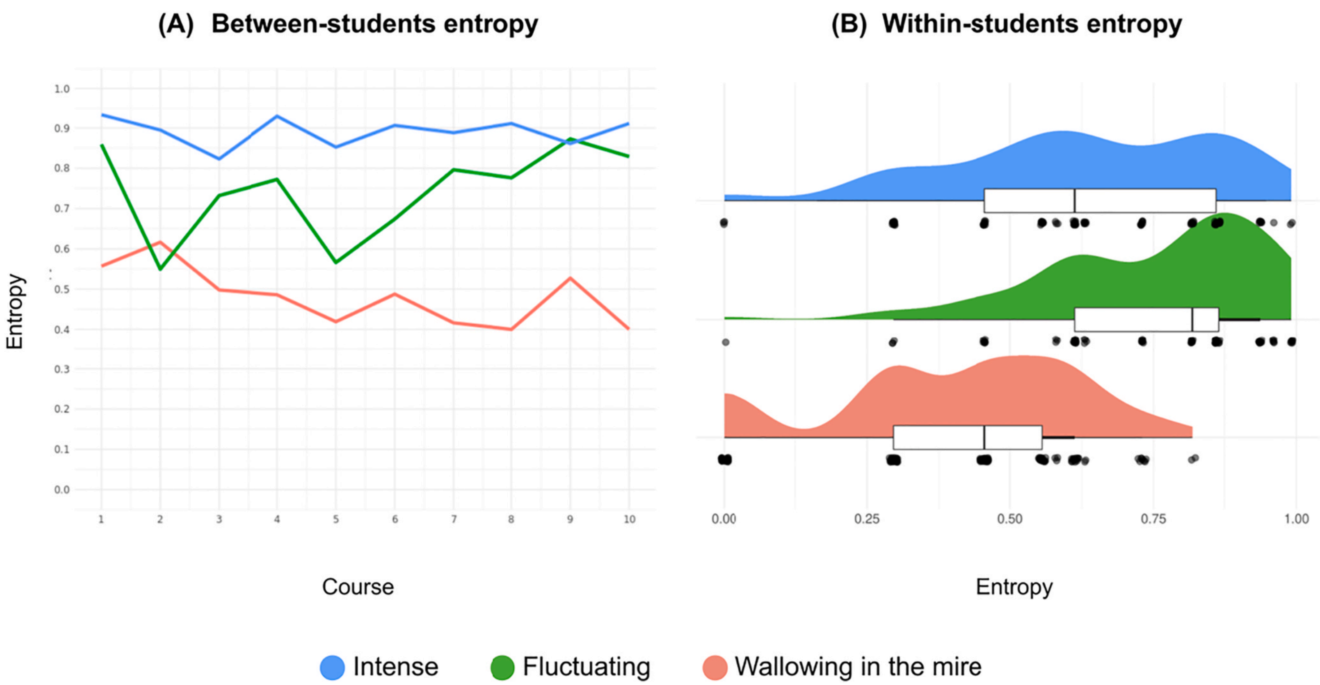


**Fig. 9.** (A) Between-students entropy and (B) Within-students entropy of each trajectory.

- ***Wallowing-in-the-mire* trajectory** (n = 133, 40%): Students in this trajectory —as the name implies— were stuck in the unfavorable role of *Isolate* (Fig. 8). Most of the students in this trajectory were *Isolates* (81%), some were *Mediators* (18%), and very few were *Leaders* (1%) at some point. This trajectory was the most consistent with infrequent transitions between roles as well as the lowest mean between-student entropy value (mean entropy = 0.4). At any given course, *Isolates* were the vast majority of roles with a ratio ranging from 71% to 87%. *Isolates* remained so between courses (84%), infrequently transitioned to mediators (15%), and rarely transitioned to leaders (1%).

### 4.3. RQ3: the unfolding of roles within each trajectory

Students in the *Wallowing-in-the-mire* trajectory were more likely to have a longer "spell duration", i.e., spend longer times "stuck" with an *Isolate* role, with a mean duration of 4 courses (MED = 3.3, SD = 2.8) compared to 2.6 (MED = 2.0, SD = 1,7) in the *Intense* trajectory, and 2.2 (MED = 2.0, SD = 1.2) in the *Fluctuating* trajectory, who showed the highest instability. The difference between the three trajectories regarding mean spell duration was statistically significant with a medium effect size ($\chi^2$ (2) = 59, p < 0.001; $\varepsilon^2$ = 0.18).

Students in the *Wallowing-in-the-mire* trajectory had the lowest within-student entropy (Fig. 9B) with a mean of 0.4 (MED = 0.46, SD = 0.22), compared to 0.75 (MED = 0.84, SD = 0.19) in the *Fluctuating* trajectory, and 0.63 (MED = 0.61, SD = 0.23) in the *Intense* trajectory. These results indicate the *Wallowing-in-the-mire* trajectory was the most consistent with infrequent role changes, whereas the *Fluctuating* trajectory was the most inconsistent. According to the Kruskal-Wallis test, such differences among the three trajectories were statistically significant with a large effect size ($\chi^2$ (2) = 114.566, p < 0.001; $\varepsilon^2$ = 0.35).

As expected, students in the *Intense* trajectory had the highest *integrative capacity*, i.e., the capability to integrate the favorable role "*Leader*" and end up in such a role compared to the two other trajectories: the integrative capacity of the *Intense* trajectory was 0.92 (SD = 0.1), for the *Fluctuating* trajectory it was 0.67 (SD = 0.21), and for the *Wallowing in the mire* it was 0.18 (SD = 0.15). The Kruskal-Wallis test showed that such differences were statistically significant with a very large effect size ($\chi^2$ (2) = 254.5, p < 0.001; $\varepsilon^2$ = 0.78).

### 4.4. RQ4: covariates of trajectory membership

To test which factors may explain why a student belongs to a certain trajectory, two main groups of variables (relevant to the collaborative process in PBL) were tested as covariates: (1) the small group in which the students participate and (2) the teacher who acts as the facilitator. The results of the covariate analysis are shown in Table 4. The *Intense* trajectory was the reference trajectory.

Compared to the *Intense* trajectory, students in the *Wallowing-in-the-mire* trajectory were more likely to be in *centralized* groups with a higher centralization degree (Est. = 25.58, p < 0.01). A centralized group is dominated by few students and therefore is less participatory. These students were in groups with lower density values (Est. = -67.53, p < 0.001), i.e., interactions were not distributed among participants. Their teachers had the lowest diffusion degree values (Est. = -29.09, p < 0.001) —i.e., the teacher posts were not discussed or taken up by the students—, lower leverage centrality values (Est. = -24.67, p < 0.001) —i.e., had relatively fewer connections—, and a slightly higher degree centrality (Est. = 0.08, p = 0.01). In summary, students in the *Wallowing-in-the-mire* trajectory were more likely to be in less collaborative groups where few students dominated, and to have teachers whose posts are less discussed and who have fewer students.

Students in the *Fluctuating* trajectory shared most of the factors with the *Wallowing-in-the-mire* trajectory; however, with lower magnitudes, lying in-between them and the *Intense* trajectory. They were likely to be in *centralized* groups with a high centralization degree (Est. = 21.52, p < 0.01) and low density (Est. = 21.52, p < 0.01). Their teachers had low diffusion degree (Est. = -15.76, p < 0.01), low leverage (Est. = -24.02, p < 0.001) and high mean degree (Est. = 0.08, p = 0.01).

## 5. Discussion

Our study analyzed the temporal evolution of roles across a full program (four years) using an innovative method that combined

**Table 4**
Results of the covariate analysis.

| | | Fluctuating | | | | Wallowing-in-the-mire | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Est. | SE | z | p | Est. | SE | z | p |
| | (Intercept) | 9.20 | 6.10 | 1.51 | 0.13 | 15.95 | 6.48 | 2.46 | 0.04* |
| Group | Centralization degree | 21.52 | 4.76 | 4.52 | 0.00*** | 25.58 | 5.14 | 4.98 | 0.00*** |
| | Mean degree | 0.18 | 0.10 | 1.81 | 0.07 | 0.32 | 0.11 | 2.84 | 0.01*** |
| | Network density | −47.76 | 12.18 | −3.92 | 0.00*** | −67.53 | 12.96 | −5.21 | 0.00*** |
| | Node count | −0.13 | 0.42 | −0.32 | 0.75 | −0.41 | 0.44 | −0.92 | 0.36 |
| Teacher | Diffusion degree | −15.76 | 7.50 | −2.10 | 0.04* | −29.09 | 8.15 | −3.57 | 0.00*** |
| | Leverage | −24.02 | 5.34 | −4.50 | 0.00*** | −24.67 | 5.43 | −4.55 | 0.00*** |
| | Neighborhood | 0.27 | 0.36 | 0.73 | 0.46 | 0.57 | 0.40 | 1.42 | 0.16 |
| | Degree | 0.08 | 0.03 | 2.90 | 0.00*** | 0.08 | 0.03 | 2.80 | 0.01** |

*Note*: Est. = Estimate; SE = Standard error. *p < 0.05, ***p < 0.001, **p < 0.01.

computational methods for the detection of roles (SNA), sequence analysis for charting the pathway of roles, and MHMM to group the roles into homogenous trajectories, study the characteristics of such trajectories, and find out the factors that explain why students belong thereto.

Our first research question aimed at discovering the types of roles that can be identified using LPA from a full program data. The identified roles were principally aligned with the theoretical frameworks proposed by Strijbos and De Laat (2010) and Driskell et al. (2017), especially at either side of the spectrum (high vs. low effort), task engagement, and dominance. We identified a leader role with high effort, influence, and strong social connections, in line with most prior research (e.g., Jimoyiannis et al., 2013; Marcos-García et al., 2015; Medina et al., 2016; Saqr et al., 2018; Xie et al., 2018). Similar roles with high effort related to the leader role were reported under different labels, e.g., influencers (Ouyang & Chang, 2019; Saqr & López-Pernas, 2021b), influentials (Dowell et al., 2019), and full participants (Kim & Ketenci, 2019). On the opposite side of the spectrum of effort, lies the isolate role which almost every study has reported under different labels, e.g., lurkers (Aviv et al., 2003; Dowell et al., 2019; e.g., Jimoyiannis et al., 2013), peripheral (Jimoyiannis et al., 2013; Kim & Ketenci, 2019; Marcos-García et al., 2015; Medina et al., 2016; Ouyang & Chang, 2019), satellites (Saqr & Viberg, 2020), detached (Dowell et al., 2019), or missing (Medina et al., 2016). In contrast to the relative agreement among studies regarding the leader and isolate roles, there was a broad intermediate role that researchers classified differently from one study to the other. Most of the of studies —similar to ours— described a role where students connect, bridge, relay or help others collaborate. That role was labeled as broker (Stuetzer et al., 2011), connector (Jimoyiannis et al., 2013), coordinator (Marcos-García et al., 2015; Medina et al., 2016; Saqr et al., 2018), mediator (Ouyang & Chang, 2019), arbitrator (Saqr & Viberg, 2020) or relay (Turkkila & Lommi, 2020). Nonetheless, there exist several intermediate roles that may reflect contextual peculiarities, e.g., source (Chen & Chang, 2014; Turkkila & Lommi, 2020), guide (Marcos-García et al., 2015) or regulator (Ouyang & Chang, 2019). Other intermediate roles represented degrees of activity that were neither intensely engaged to be leaders nor averagely participating to be labeled as mediators; for instance, the active role reported by (Medina et al., 2016; Saqr et al., 2018). As such, our results concur with prior research regarding the main types of roles. Yet, our range of centrality measures covered several dimensions based on the theoretical frameworks and previous research. In particular we stressed the dimension of influence, diffusion of ideas, and uptake to go beyond simple counts of posts (Saqr & López-Pernas, 2021b; Saqr & Viberg, 2020). Whereas previous research relied on manual thresholds (Chen & Chang, 2014; Stuetzer et al., 2011), structural equivalence (Laghos & Zaphiris, 2007; Medina et al., 2016) or clustering (Kim & Ketenci, 2019; Xie et al., 2018), our study introduced latent profile analysis as a method for identifying roles. As a mixture method, latent profile analysis does not require a certain distribution, is less likely to be influenced by outliers (a common problem with most students' data) as well as offers a rich toolset of verification methods that allows robust modeling (Spurk et al., 2020; Weller et al., 2020). Since Saqr and López-Pernas (2021b) have reported a large number of courses, their clustering results are worth comparing to ours. The leaders in our study account for a slightly higher percentage 4%; the mediators are 8% less, and the isolates are 4% more. While such differences are small, an explanation could be that our study had an inclusion criteria that students had to be enrolled for ten successive courses, used a different clustering method (LPA), and we used a slightly different set of centrality measures (i.e., we did not include coreness).

Our second research question —the prime motivation of this study— aimed at examining the temporal nature of roles that students assume across a full program. Strijbos and De Laat (2010) suggested that roles exist as a *stance* that drives participatory behavior in a task or a group of tasks. Given that the stance explains a rather limited timespan, our study investigated whether a pervasive attitude (i.e., a *disposition*) exists, and therefore drives students to assume similar roles across the program, i.e., leader students continue to emerge as leaders in most collaborative settings anytime in the program. Since students are heterogenous, they are expected to have different dispositions, i.e., some would always assume the active participatory role of a leader, and some would assume the role of a mediator. Such a dispositional nature of roles was confirmed by our study, i.e., emergent roles followed a repetitive pattern within each student across most courses of the program. Clustering the pathways of students (their successive roles) resulted in three homogeneous trajectories: *intense* (leaders most of the time), *fluctuating* (moderators most of the time), and *wallowing-in-the-mire* (isolate most of the time).

The *wallowing-in-the-mire* trajectory was the most stable trajectory in which students remained in the same role for the longest periods of time (longest spells of successive roles) and had the fewest changes or transitions (lowest within-student entropy or variability). This trajectory was the most homogenous (lowest between-student entropy) and such students were the least likely to end up in a favorable role (leader or mediator). A longitudinal trajectory of roles has not been previously described in the literature and, therefore, parallels cannot be directly drawn from previous research. Using raw values of centrality measures, some studies examined the variability of students' activity on a daily basis. Such results reflected fluctuations in interaction intensity rather than changes of roles within the task as a whole (Skrypnyk et al., 2015; Xie et al., 2018). The sequences of roles —similar to ours— were studied by Boroujeni et al. (2017) in a MOOC, where the authors reported weekly variability of active and inactive roles. Nonetheless, the temporal granularity (weekly) and the nature of the MOOC does not allow a head-to-head comparison with our study. To that end, our finding that a stable trajectory in which students opt to assume an inactive role in most collaborative courses is novel, yet alarmingly ominous.

The second trajectory we identified was the *fluctuating* trajectory which showed a predominantly moderator role, higher churn rate (frequency of changing roles), and variability within and between students' pathways. While —as previously described— longitudinal research is lacking, one can draw parallels between our study and that of Xie et al. (2018), who observed that some of the students who were assigned the role of moderators tended to assume a leadership role manifested as a higher leadership index. Some of these students transitioned to inactive roles, and some remained stable. The last trajectory we identified was the *intense* trajectory in which students predominantly opted for leader roles with infrequent transitions to mediators. This trajectory was moderately stable with fewer transitions and infrequent between-student variations. Such findings are indicative of a subgroup of students who are actively engaged in most collaborative tasks and willing to invest significant time and effort regardless of the course or the collaborative task.

While the study of trajectories is emerging in education (e.g., Saqr & López-Pernas, 2021a), research on the factors that explain the trajectory of membership is almost non-existent. Therefore, our study looked into the covariates that may explain such trajectories using the two relevant groups of variables relevant to the PBL process, i.e., the group factors and the teacher factors. Students in the wallowing-in-the-mire trajectory were more likely to be in productive groups that have a higher number of interactions than their counterparts and whose teachers have a higher number of interactions as well. However, such groups were dominated by few students who interact with each other, and the teacher's posts were less likely to be discussed by the students. Whereas inferring causation cannot be made here, these factors point to a picture of inactive students in interactive groups, i.e., they chose not to be involved in the collaborative process.

### *5.1. Role as a disposition*

The stance —as described by Strijbos and De Laat (2010)— concerns a rather limited time or situational "orientation towards the group task" or the extent to which a student "wants to (or can) engage in this task". Our results have shown that roles re-emerged consistently regardless of the task or the course over extended periods of time and in a predictable manner. For instance, isolates "assumed" such a role in almost all of their courses over the four years of study. The covariate analysis further supports this argument: isolates were in active groups, with active teachers, however, they interacted less and with few collaborators. Therefore, we use the word "assumed" which expresses the volition and agency of students in role-taking. Our findings of the longitudinal consistency support the conclusion that roles follow a disposition rather than re-emerge based on a transitory attitude towards certain tasks. To that end, we propose an extension of Strijbos and De Laat's framework (2010) where a fourth dimension of roles exists which represents the pervasive attitude towards collaborative work in general (Fig. 10). Such a longitudinal attitude can be referred to as a collaborative role as a ***disposition***. Our argument and proposal are supported by a large volume of data (329 students: 84,597 interactions) over four continuous years, along with a robust statistical analysis. Nonetheless, our conclusions may need to be put to the test in other contexts, before drawing firm conclusions. While the disposition —described here— may explain the majority of role taking, we emphasize that other factors may also come into play when students assume their roles, e.g., task, teacher, and group, however, these may not be the primary factors.

Conceptualizing roles as a disposition has important implications for research and practice. Given that collaboration, communication skills, critical thinking and learning to learn are important skills in their own right as well as essential qualities of today's professions, it stands to reason those unfavorable roles (e.g., *isolates*) should be a matter of concern that require educators' attention. Most theoreticians posit that dispositions are malleable to intervention: strategies like facilitation, role rotation, and scripting could offer potential solutions (Bevan, 2019; De Wever & Strijbos, 2021). While curricula always comprise knowledge, skills, dispositions, or attitude as basic elements, dispositions are rarely stressed despite their importance to learning and professional practice (Bair, 2017; Katz, 1993). Group composition has always relied on similarities of demographics, abilities or interests (Lin et al., 2010; Wilkinson & Fung, 2002). We believe that the results presented in our study indicate that students' prior roles are relevant as a factor in future role assignment and should be considered as a method for improving collaborative skills. Of course, not all students are comfortable with

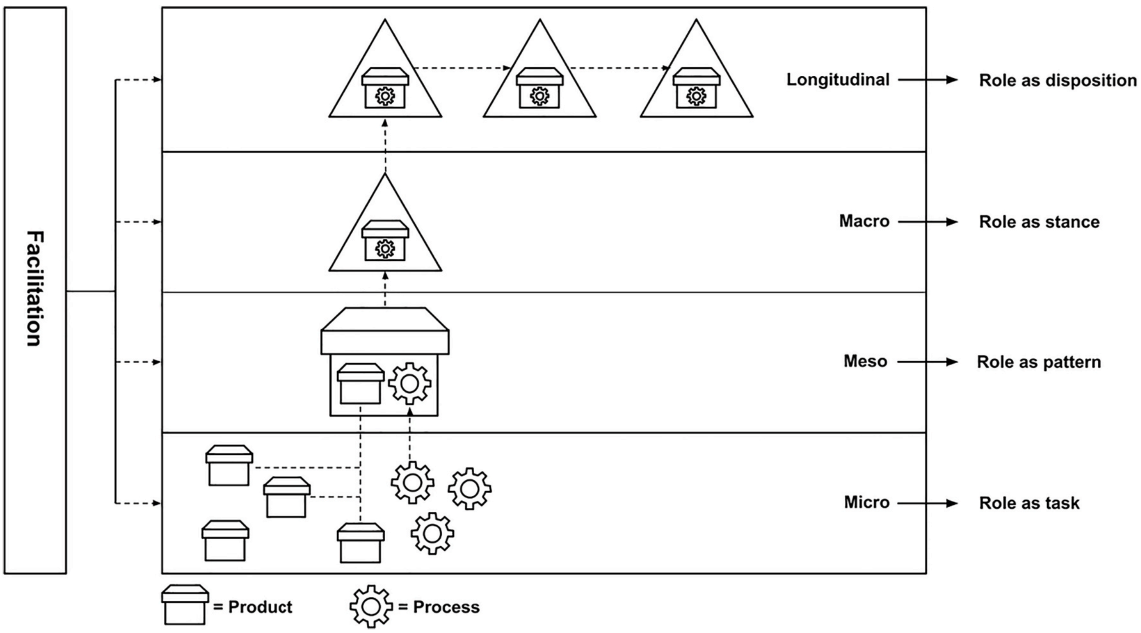


**Fig. 10.** Extension of the framework by Strijbos and De Laat (2010) where students repeat their role oftentimes at the longitudinal level.

socializing or engaging in demanding interactions and, therefore, benefits must be weighed against potential distress.

Self-regulated learning emphasizes agency, control, self-directedness —and so does PBL— and goal orientation as essential qualities and dispositions that students should possess to succeed (Järvenoja et al., 2013). Such qualities guide students' motivation and behavior. It seems that students with poor self-regulatory skills choose a non-effective trajectory where they assume an isolate role repeatedly. Therefore, support is needed to improve students' awareness and enhance their motivation which, in turn, could lead to better self and group performance (Hadwin et al., 2017; Järvelä & Hadwin, 2013; Malmberg et al., 2022). Looking at the group-level, when the collaborative group had a shared-regulation load (more distributed interactions, less dominance, and less teacher involvement) students were more likely to have active roles (i.e., leaders), and leaders were more likely to continue to emerge as leaders in such jointly working groups. This is consistent with the socially shared regulation view that sustains that when group members engage in joint activities, work together, and coordinate, they are more likely to be productive, engaged, and succeed in the execution of their tasks (Dillenbourg et al., 2009; Hadwin et al., 2017; Järvelä et al., 2021).

Longitudinal research is rather rare; yet, it is much needed to understand the temporal evolution of learners and their behavior (Saqr & López-Pernas, 2021a). The methods used in this study —summarized in Fig. 1— can be applied to a wide range of educational contexts where multiple time points exist as well as across a wide range of constructs. While we have used the course as a time unit, the methods can be applied to several time scales, e.g., days, weeks, and tasks. For instance, the same method can be used to trace the sequence of students' engagement in individual tasks.

## 6. Limitations of this study

As a longitudinal study, our inclusion criteria constrained participants to those who completed ten courses to fulfill the longitudinal condition and enable comparison. In doing so, the longitudinal condition creates a survivorship bias, i.e., limits the study to students who were able to "survive" or be retained for the full research period. A common problem of survivorship bias is that it creates a more optimistic view of the outcome (Brown et al., 1992; Carpenter & Lynch, 1999). Since our study aimed to automatically analyze the collected data, our reliance on SNA data and LPA can be considered as a limitation, i.e., not including a qualitative method for content coding to get an in-depth view of the students' discourse. There is of course a tradeoff: manually coding a massive number of interactions, like the ones analyzed in our study, was impractical. However, given that SNA has been established as a modeling method for almost two decades and repeatedly validated with other studies with qualitative analysis, our methods are evidently appropriate. The accuracy of the calculated centrality measures depends on the configuration of the forum platform, e.g., some students do not reply to the intended post but to the first post in the thread which could create inaccuracies, e.g., betweenness centrality. We believe that using several centrality measures has greatly mitigated this problem, as evidenced by the LPA fit statistics. As with any clustering method, there are possible instances where students are classified as a role that they do not belong to, i.e., classifying a leader as a moderator. The misclassification problem is particularly important for the intermediate class (i.e., mediators) and at the intermediate area between classes where the clustering algorithm classifies a student as, for instance, 48% probability leader and 52% probability moderator. However, the percentage of students at either side of class (with probabilities between 40% and 60%) were around 1% indicating high homogeneity of classes and that the misclassification was not a concern. Our study was performed in a PBL healthcare program and therefore generalizability to other disciplines and contexts remains to be tested. Including achievement in the covariates would have added to our understanding of roles. However, due to the large dataset and large number of course enrollments (which was 3290), course grades were not available for many students. Since our results rely on interactions and collaboration profiles, we believe that the absence of grades from the analysis has not affected our conclusions.

## 7. Conclusions

Our study analyzed the longitudinal progression of roles across a full program (four years) using an innovative method where we combined SNA for role detection, sequence analysis for charting the pathway of roles, and MHMM to cluster the roles into trajectories, study the characteristics of such trajectories, and find out the factors that explain why students belong thereto. We identified three roles that were comparable to previous research: a *leader* role (active and influential), a *mediator* (moderately active who moderates posts), and an *isolate* (the least active). Most importantly, at the program level, we found three distinct longitudinal trajectories: an *intense* trajectory which includes mostly leaders, a *fluctuating* trajectory which includes mostly mediators, and a *wallowing-in-the-mire* trajectory which includes mostly isolates. Our results show that roles re-emerge consistently regardless of the task or the course over extended periods of time and in a predictable manner. For instance, isolates "assumed" such a role in almost all of their courses over four years. Our findings of the longitudinal consistency support the conclusion that roles follow a ***disposition*** rather than re-emerge based on a transitory attitude towards certain tasks. To that end, we propose an extension of Strijbos and De Laat's framework (2010) where a fourth dimension of roles exists which represents the pervasive attitude towards collaborative work in general. Such a longitudinal attitude can be referred to as a collaborative role as a ***disposition***. While the disposition —described here— may explain the majority of role-taking, we emphasize that other factors may come into play when students assume their roles, e.g., task, teacher, and group. Nonetheless, these may not be the primary factors.

## Credits

MS and SLP have contributed to the idea conceptualization, research design, and planning. MS has performed data collection. MS and SLP have contributed to the methods, data analysis and reporting of results and visualization. Visualizations have been

conceptualized by MS and SLP and created by SLP. MS and SLP have contributed to manuscript writing and revision. The authors read, revised, and approved the final manuscript.

## Acknowledgments

The paper is co-funded by the Academy of Finland for the project TOPEILA, Decision Number 350560 which was received by the first author.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compedu.2022.104581.

## References

Ajzen, I. (1987). Attitudes, traits, and actions: Dispositional prediction of behavior in personality and social psychology. In L. Berkowitz (Ed.)*, Vol. 20. Advances in Experimental social Psychology* (pp. 1–63). Academic Press.
Aviv, R., Erlich, Z., & Ravid, G. (2003). Cohesion and roles: Network analysis of CSCL communities. *Proceedings 3rd IEEE International Conference on Advanced Technologies*, 145–149.
Bae, J., & Kim, S. (2014). Identifying and ranking influential spreaders in complex networks by neighborhood coreness. *Physica A: Statistical Mechanics and its Applications, 395*, 549–559.
Bair, M. A. (2017). Identifying dispositions that matter: Reaching for consensus using a delphi study. *The Teacher Educator, 52*(3), 222–234.
Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2013). The diffusion of microfinance. *Science, 341*(6144). https://doi.org/10.1126/science.1236498
Bevan, N. D. (2019). *Understanding factors that influence the Effectiveness of construal level on self-control success*. Ohio State University.
Borgatti, S. P. (2005). Centrality and network flow. *Social Networks, 27*(1), 55–71.
Borgatti, S. P., & Brass, D. J. (2019). Centrality: Concepts and measures. In *Social networks at work* (pp. 9–22).
Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science, 323*(5916), 892–895.
Boroujeni, M. S., Hoppe, H. U., Hecking, T., & Dillenbourg, P. (2017). Dynamics of MOOC discussion forums. *ACM International Conference Proceeding Series*, 128–137.
Brown, S. J., Goetzmann, W., Ibbotson, R. G., & Ross, S. A. (1992). Survivorship bias in performance studies. *Review of Financial Studies, 5*(4), 553–580.
Car, L. T., Myint Kyaw, B., Dunleavy, G., Smart, N. A., Semwal, M., Rotgans, J. I., Low-Beer, N., & Campbell, J. (2019). Digital problem-based learning in health professions: Systematic review and meta-analysis by the digital health education collaboration. *Journal of Medical Internet Research, 21*(2), 1–12.
Carpenter, J. N., & Lynch, A. W. (1999). Survivorship bias and attrition effects in measures of performance persistence. *Journal of Financial Economics, 54*(3), 337–374.
Chen, C.-M., & Chang, C.-C. (2014). Mining learning social networks for cooperative learning with appropriate learning partners in a problem-based learning environment. *Interactive Learning Environments, 22*(1), 97–124.
Clark, S. L., & Muthén, B. (2009). *Relating latent class analysis results to variables not included in the analysis*. https://www.statmodel.com/download/relatinglca.pdf.
Claxton, G., & Carr, M. (2004). A framework for teaching learning: The dynamics of disposition. *Early Years: An International Journal of Research and Development, 24*(1), 87–97.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge.
Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.
Csardi, G., & Nepusz, T. (2006). *The igraph software package for complex network research*. InterJournal, Complex Sy.
De Wever, B., Schellens, T., Van Keer, H., & Valcke, M. (2008). Structuring asynchronous discussion groups by introducing roles: Do students act in line with assigned roles? *Small Group Research, 39*(6), 770–794.
De Wever, B., & Strijbos, J.-W. (2021). Roles for structuring groups for collaboration. In U. Cress, C. Rosé, A. F. Wise, & J. Oshima (Eds.), *International Handbook of Computer-supported collaborative learning* (pp. 315–331). Springer International Publishing.
Dillenbourg, P., & Fischer, F. (2007). Computer-supported collaborative learning: The basics. *Zeitschrift Für Berufs-Und Wirtschaftspädagogik, 21*, 111–130.
Dillenbourg, P., Järvelä, S., & Fischer, F. (2009). The evolution of research on computer-supported collaborative learning. In *Technology-enhanced learning* (pp. 3–19). Springer.
Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography, 36*(1), 27–46.
Dowell, N. M. M., Nixon, T. M., & Graesser, A. C. (2019). Group communication analysis: A computational linguistics approach for detecting sociocognitive roles in multiparty interactions. *Behavior Research Methods, 51*(3), 1007–1041.
Driskell, T., Driskell, J. E., Burke, C. S., & Salas, E. (2017). Team roles: A review and integration. *Small Group Research, 48*(4), 482–511.
Gabadinho, A., Ritschard, G., Mueller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software, 40*(4), 1–37.
Hadwin, A., Järvelä, S., & Miller, M. (2017). Self-regulation, co-regulation, and shared regulation in collaborative learning environments. In *Handbook of self-regulation of learning and performance* (pp. 83–106). Routledge.
Hare, A. P. (1994). Types of roles in small groups: A bit of history and a current perspective. *Small Group Research, 25*(3), 433–448.
Hecking, T., & Chounta, I. A. (2017). *Role modelling in MOOC discussion forums. Of Learning Analytics*. http://lak13.learning-analytics.info/journals/index.php/JLA/article/view/5149.
Helske, S., & Helske, J. (2019). Mixture hidden Markov models for sequence data: The seqHMM package in R. In *, Vol. 88*. *Journal of statistical software*. https://doi.org/10.18637/jss.v088.i03. Issue 3.
Helske, S., Helske, J., & Eerola, M. (2018). Combining sequence analysis and hidden Markov models in the analysis of complex life sequence data. In *Life course research and social Policies* (pp. 185–200). https://doi.org/10.1007/978-3-319-95420-2_11
Hernández-García, Á., González-González, I., Jiménez-Zarco, A. I., & Chaparro-Peláez, J. (2015). Applying social learning analytics to message boards in online distance learning: A case study. *Computers in Human Behavior, 47*(PG-68–80), 68–80.
Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, Theory and Applications, 6*(2), 65–70.
Jalili, M., Salehzadeh-Yazdi, A., Asgari, Y., Arab, S. S., Yaghmaie, M., Ghavamzadeh, A., & Alimoghaddam, K. (2015). CentiServer: A comprehensive resource, web-based application and R package for centrality analysis. *PLoS One, 10*(11), Article e0143111.
Järvelä, S., & Hadwin, A. F. (2013). New frontiers: Regulating learning in CSCL. *Educational Psychologist, 48*(1), 25–39.
Järvelä, S., Malmberg, J., Sobocinski, M., & Kirschner, P. A. (2021). Metacognition in collaborative learning. In *International Handbook of Computer-supported collaborative learning* (pp. 281–294). Springer International Publishing.

Järvenoja, H., Volet, S., & Järvelä, S. (2013). Regulation of emotions in socially challenging learning situations: An instrument to measure the adaptive and social nature of the regulation process. *Educational Psychology, 33*(1), 31–58.
Jeong, H., & Hmelo-Silver, C. E. (2016). Seven affordances of computer-supported collaborative learning: How to support collaborative learning? How can technologies help? *Educational Psychologist, 51*(2), 247–265.
Jimoyiannis, A., Tsiotakis, P., & Roussinos, D. (2013). Social network analysis of students' participation and presence in a community of educational blogging. *Interactive Technology and Smart Education, 10*(1), 15–30.
Joksimovic, S., Manataki, A., Gaševic, D., Dawson, S., Kovanovic, V., & De Kereki, I. F. (2016). Translating network position into performance: Importance of centrality in different network configurations. In *ACM International Conference Proceeding Series, 25-29-Apri(PG-314-323)* (pp. 314–323).
Joyce, K. E., Laurienti, P. J., Burdette, J. H., & Hayasaka, S. (2010). A new measure of centrality for brain networks. *PLoS One, 5*(8), Article e12200.
Katz, L. G. (1993). *Dispositions as Educational Goals. ERIC Digest*. https://eric.ed.gov/?id=ED363454.
Kim, M. K., & Ketenci, T. (2019). Learner participation profiles in an asynchronous online collaboration context. *The Internet and Higher Education, 41*, 62–76.
Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts – a conceptual analysis. *Educational Psychology Review, 18*(2), 159–185.
Kreijns, K., Kirschner, P. A., & 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.
Kreijns, K., Kirschner, P. A., & Vermeulen, M. (2013). Social aspects of CSCL environments: A research framework. *Educational Psychologist, 48*(4), 229–242.
Laat, M. D., Lally, V., Lipponen, L., & Simons, R.-J. (2007). Online teaching in networked learning communities: A multi-method approach to studying the role of the teacher. *Instructional Science, 35*(3), 257–286.
Laghos, A., & Zaphiris, P. (2007). Social network analysis of self-taught e-learning communities. *International Journal of Knowledge and Learning, 3*(4–5), 465–482.
Lin, Y.-T., Huang, Y.-M., & Cheng, S.-C. (2010). An automatic group composition system for composing collaborative learning groups using enhanced particle swarm optimization. *Computers & Education, 55*(4), 1483–1493.
Liu, Y., Tang, M., Zhou, T., & Do, Y. (2016). Identify influential spreaders in complex networks, the role of neighborhood. *Physica A: Statistical Mechanics and its Applications, 452*, 289–298.
López-Pernas, S., & Saqr, M. (2021). Bringing synchrony and clarity to complex multi-channel data: A learning analytics study in programming education. *IEEE Access, 9*, 166531–166541.
López-Pernas, S., Saqr, M., & Viberg, O. (2021). Putting it all together: Combining learning analytics methods and data sources to understand students' approaches to learning programming. *Sustainability, 13*(9), 4825.
Ludvigsen, S., Lund, K., & Oshima, J. (2021). A conceptual stance on CSCL history. In *International Handbook of Computer-supported collaborative learning* (pp. 45–63). Cham: Springer.
Lu, J., Lajoie, S. P., & Wiseman, J. (2010). Scaffolding problem-based learning with CSCL tools. *International Journal of Computer-Supported Collaborative Learning, 5*(3), 283–298.
Malmberg, J., Saqr, M., Järvenoja, H., & Järvelä, S. (2022). How the monitoring events of individual students are associated with phases of regulation. *Journal of Learning Analytics, 9*(1), 77–92.
Marcos-García, J.-A., Martínez-Monés, A., & Dimitriadis, Y. (2015). Despro: A method based on roles to provide collaboration analysis support adapted to the participants in CSCL situations. *Computers & Education, 82*, 335–353.
Matcha, W., Gašević, D., Jovanović, J., Uzir, N. A., Oliver, C. W., Murray, A., & Gasevic, D. (2020). Analytics of learning strategies: The association with the personality traits. *ACM International Conference Proceeding Series*, 151–160.
Medina, E., Vega, D., Meseguer, R., Medina, H., Ochoa, S. F., & Magnani, M. (2016). Using indirect blockmodeling for monitoring students roles in collaborative learning networks. In *2016 IEEE 20th International Conference on computer supported Cooperative work in design (CSCWD)* (pp. 164–169).
Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. In *, Vol. 4. Translational Issues in Psychological science* (pp. 440–461). https://doi.org/10.1037/tps0000176. Issue 4.
Ostertagova, E., Ostertag, O., & Kováč, J. (2014). Methodology and application of the Kruskal-Wallis test. *Applied Mechanics and Materials, 611*, 115–120.
Ouyang, F., & Chang, Y.-H. (2019). The relationships between social participatory roles and cognitive engagement levels in online discussions. *British Journal of Educational Technology: Journal of the Council for Educational Technology, 50*(3), 1396–1414.
Poquet, O., & Jovanovic, J. (2020). Intergroup and interpersonal forum positioning in shared-thread and post-reply networks. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 187–196.
R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org.
Roethlisberger, F. J., & Dickson, W. J. (1939). *Management and the worker* (Vol. 615). https://psycnet.apa.org/fulltext/1940-00509-000.pdf.
Rosenberg, J., Beymer, P., Anderson, D., van Lissa, C.j., & Schmidt, J. (2018). tidyLPA: An R package to easily carry out latent profile Analysis (LPA) using open-source or commercial software. *Journal of Open Source Software, 3*(30), 978.
Saqr, M., Elmoazen, R., Tedre, M., López-Pernas, S., & Hirsto, L. (2022). How well centrality measures capture student achievement in computer-supported collaborative learning? – a systematic review and meta-analysis. *Educational Research Review, 35*, Article 100437. https://doi.org/10.1016/j.edurev.2022.100437
Saqr, M., Fors, U., Tedre, M., & Nouri, J. (2018). How social network analysis can be used to monitor online collaborative learning and guide an informed intervention. *PLoS One, 13*(3), Article e0194777.
Saqr, M., & López-Pernas, S. (2021a). The longitudinal trajectories of online engagement over a full program. *Computers & Education, 175*, Article 104325.
Saqr, M., & López-Pernas, S. (2021b). Modelling diffusion in computer-supported collaborative learning: A large scale learning analytics study. *International Journal of Computer-Supported Collaborative Learning*. https://doi.org/10.1007/s11412-021-09356-4
Saqr, M., & Peeters, W. (2022). Temporal networks in collaborative learning: A case study. *British Journal of Educational Technology: Journal of the Council for Educational Technology*. https://doi.org/10.1111/bjet.13187
Saqr, M., & Viberg, O. (2020). Using diffusion network analytics to examine and support knowledge construction in CSCL settings. In *, Vol. 12315. Addressing global challenges and quality education. EC-TEL 2020* (pp. 158–172). Springer.
Saqr, M., Viberg, O., & Vartiainen, H. (2020). Capturing the participation and social dimensions of computer-supported collaborative learning through social network analysis: Which method and measures matter? *International Journal of Computer-Supported Collaborative Learning, 15*(2), 227–248.
Skrypnyk, O., Joksimović, S., Kovanović, V., Gassević, D., & Dawson, S. (2015). Roles of course facilitators, learners, and technology in the flow of information of a CMOOC. *International Review of Research in Open and Distance Learning, 16*(3), 188–217.
Spurk, D., Hirschi, A., Wang, M., Valero, D., & Kauffeld, S. (2020). Latent profile analysis: A review and "how to" guide of its application within vocational behavior research. *Journal of Vocational Behavior, 120*, Article 103445.
Stahl, G., Koschmann, T., & Suthers, D. (2014). *Cambridge Handbook of the learning sciences. Computer-supported collaborative learning: An historical perspective* (pp. 409–426). Cambridge Handbook of the Learning Sciences.
Strijbos, J.-W., & De Laat, M. F. (2010). Developing the role concept for computer-supported collaborative learning: An explorative synthesis. *Computers in Human Behavior, 26*(4), 495–505.
Strijbos, J.-W., Kirschner, P. A., & Martens, R. L. (2006). *What We Know about CSCL: And implementing it in higher education*. Springer Science & Business Media.
Stuetzer, C. M., Carley, K. M., Koehler, T., & Thiem, G. (2011). The communication infrastructure during the learning process in web based collaborative learning systems. *Proceedings of the 3rd International Web Science Conference*, 1–8.
Stuetzer, C. M., Koehler, T., Carley, K. M., & Thiem, G. (2013). Brokering" behavior in collaborative learning systems. *Procedia - Social and Behavioral Sciences, 100*, 94–107.
Temdee, P., Thipakorn, B., Sirinaovakul, B., & Schelhowe, H. (2006). Of collaborative learning team: An approach for emergent leadership roles identification by using social network analysis. *Technologies for E-Learning and Digital Entertainment*, 745–754.
Turkkila, M., & Lommi, H. (2020). Student participation in online content-related discussion and its relation to students' background knowledge. *Education in Science: The Bulletin of the Association for Science Education, 10*(4), 106.

Vieira, V. A. (2011). *Experimental designs using ANOVA* (Vol. 15, pp. 363–365). Belmont, CA: Thomson/Brooks/Cole.
Vignery, K. (2022). From networked students centrality to student networks density: What really matters for student performance? *Social Networks, 70*, 166–186.
Wang, J., Hou, X., Li, K., & Ding, Y. (2017). A novel weight neighborhood centrality algorithm for identifying influential spreaders in complex networks. *Physica A: Statistical Mechanics and its Applications, 475*, 88–105.
Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education, 46*(1), 71–95.
Weinberger, A., Stegmann, K., & Fischer, F. (2010). Learning to argue online: Scripted groups surpass individuals (unscripted groups do not). *Computers in Human Behavior, 26*(4), 506–515.
Weller, B. E., Bowen, N. K., & Faubert, S. J. (2020). Latent class Analysis: A guide to best practice. *Journal of Black Psychology, 46*(4), 287–311.
Wilkinson, I. A. G., & Fung, I. Y. Y. (2002). Small-group composition and peer effects. *International Journal of Educational Research, 37*(5), 425–447.
Wise, A. F., & Cui, Y. (2018). Unpacking the relationship between discussion forum participation and learning in MOOCs: Content is key. *ACM International Conference Proceeding Series*, 330–339.
Xie, K., Di Tosto, G., Lu, L., & Cho, Y. S. (2018). Detecting leadership in peer-moderated online collaborative learning through text mining and social network analysis. *The Internet and Higher Education, 38*, 9–17.