International Journal of Educational Technology in Higher Education

RESEARCH ARTICLE

Open Access

Leveraging the Louvain algorithm for enhanced group formation and collaboration in online learning environments

Minkyung Lee[1*] and Priya Sharma[1]

*Correspondence: minklee010@gmail.com

[1] Department of Learning and Performance Systems, Pennsylvania State University, University Park, PA, USA

Abstract

This study explores the dynamics of student interaction networks within an online asynchronous discussion forum, focusing on both whole group discussions and subgroup interactions distinguished by the Louvain algorithm, a renowned community detection method. Analyzing 2481 posts from 154 undergraduate students across three sections of a communications course centered on discussions about movie clips or social phenomena to enhance media literacy, this research aims to interpret the interaction patterns in these virtual spaces. Traditional methods of group formation, such as teacher intervention and self-selection, often fail to create balanced and effective groups, especially in large online courses. The Louvain algorithm, known for its efficiency in modularity optimization, identifies clusters based on actual student interaction patterns. By leveraging both global and local network analyses, this study provides a comprehensive understanding of interaction structures. The global network analysis offers a macro view of overall interaction structures, revealing diverse patterns despite identical course designs, suggesting the influence of unique group dynamics. The local analysis, focusing on the intricacies of node and edge connections, underscores that the Louvain algorithm's classifications correlate with heightened cohesiveness and collaborative potential. The results indicate that algorithmically detected groups exhibit strong internal communication and cohesiveness, as evidenced by high clustering coefficients, density values, and weighted degrees. These findings underscore the potential of network analysis to optimize online student interactions, providing valuable insights for refining educational design to promote student engagement and collaborative problem-solving. This research highlights the transformative potential of integrating advanced data-driven techniques in educational technology to improve group formation and collaborative learning outcomes, offering empirical insights for educators to enhance online interactions and expand pedagogical understanding.

Keywords: Louvain algorithm, Group formation, Asynchronous discussion, Cohesiveness, Network analysis, Online learning, Community detection, Collaborative learning, Educational technology, Media literacy, Students engagement

Introduction

The rise of online education has introduced new challenges in effectively forming groups that encourage participation and enhance the learning experience. Unlike traditional classrooms, online environments lack physical presence and spontaneous interactions, making it difficult for educators to facilitate student engagement. This study investigates the use of the Louvain algorithm, a method from network analysis, to improve group formation and collaboration in online learning environments.

Group projects and peer interactions in traditional classrooms support deeper comprehension and critical thinking (Springer et al., 1999). However, translating these benefits to an online domain, with its geographical barriers, requires innovative approaches. Group discussions in online courses serve as crucial spaces for student interaction. Yet, forming effective groups in online courses, especially those with extensive enrollments, remains a challenge due to its impact on collaborative learning outcomes (Kozlowski & Bell, 2003). This challenge is a critical research agenda in Computer Supported Collaborative Learning (CSCL) (Ouyang & Scharber, 2017; Rodríguez et al., 2011; Weinberger et al., 2005).

The existing literature primarily explores traditional group formation methods, such as teacher intervention and self-selection, which have limitations in ensuring balanced and effective groups (Felder & Brent, 2001). While some studies have attempted algorithmic solutions for automatic group formation (Müller et al., 2022), these have largely been based on students' demographic and psychological characteristics. However, there is a lack of research exploring the utility of network science algorithms, like the Louvain algorithm, for group formation based on actual interaction patterns.

The Louvain algorithm, renowned for its efficiency in modularity optimization in networks (Blondel et al., 2008) is employed here to discern clusters among students based on their interaction patterns. This study hypothesizes that these algorithmically detected groups correspond to naturally occurring collaborative units. By analyzing interaction data from an undergraduate communications course, this study aims to assess the algorithm's efficacy in identifying cohesive and collaborative student groups.

In addition to facilitating group formation, the Louvain algorithm provides analytical insights into collaborative dynamics within digital educational spaces. These insights are essential for refining educational design to promote student engagement and collaborative problem-solving (Rodríguez et al., 2011). Accordingly, this work outlines our methodology, presents findings from the Louvain algorithm's application, and then discusses its potential to enhance the online educational experience. The goal is to demonstrate the value of network analysis in educational technology, offering a fresh perspective on optimizing online student interactions for improved educational outcomes.

Theoretical background

Online learning communities

The notion of "community" is integral to effectiveness and satisfaction in online learning environments. Rovai (2002) and Wenger (1998) have emphasized that a sense of community in online courses enhances the students' overall learning experience and positively impacts their academic outcomes. Within the expanding domain of online education, fostering a sense of community has gained considerable attention from researchers. Much research ascertained that a strong sense of community in online learning environments elevates the quality of educational experiences and positively influences academic outcomes (Shea et al., 2006; Singh et al., 2022). This perspective gains extra significance in the context of asynchronous online discussions. The asynchronous nature of these platforms such as discussion boards, can improve the role of community, as the absence of real-time interaction necessitates other forms of rich, meaningful discussion (Wise et al., 2012).

These discussions serve as more than only informational exchanges; they evolve into environments for developing skills such as critical thinking and problem-solving. The Community of Inquiry (CoI) framework (Garrison et al., 2000) highlights that such interactions are vital for fostering an online learning environment where knowledge is co-constructed. Within such communities, the collaborative engagement among peers serves as a medium for social, cognitive, and teaching presence, the core elements of the CoI framework. A study by Xie et al. (2014) indicated that well-facilitated online communities could lead to higher grades, increased engagement, and a deeper understanding of course content. This becomes increasingly significant in the situation where student retention is a concern. Rovai (2007) and, more recently, Peacock and Cowan (2019) suggest that a sense of belonging, often nurtured through strong community interactions, can be crucial in retaining students in online courses. However, the process of forming these communities, or discussion groups, remains an open question.

Traditional methods like teacher intervention and self-selection have been critiqued for their limitations in ensuring balanced and effective groups (Felder & Brent, 2001). While some have attempted algorithmic solutions for automatic group formation (Müller et al., 2022), these have primarily been based on students' demographic and psychological characteristics.

Herein lies the research gap this study aims to fill. While the existing literature provides a foundation for the importance of online learning communities and suggests the necessity for effective group formation, there is little research exploring the utility of network science algorithms like the Louvain algorithm for this purpose (Lee & Sharma, 2023). The Louvain algorithm, known for its efficiency in modularity optimization in networks (Blondel et al., 2008), offers a data-driven approach to cluster students based on their actual interactions, thereby promising more cohesive and naturally formed groups.

Group formation

Group formation is essential for effective collaborative learning, as it relies on student interaction (Cohen, 2023). The composition of groups is determined by the characteristics of their members, which significantly impact the efficiency and effectiveness of group processes (Müller et al., 2022). According to Bell (2007), group composition variables can be categorized into two levels: surface-level and deep-level attributes. Surface-level attributes include demographic factors such as age, race, education level, and organizational tenure, while deep-level attributes refer to psychological traits, such as personality, values, and attitudes. Bell suggests that these team composition variables can be used to compose teams and improve team performance effectively, considering contextual factors.

Many studies on small group composition focus on examining heterogeneous versus homogeneous groups in collaborative situations. For example, Donovan et al. (2018) found that low competence students achieved better learning outcomes in heterogeneous biology groups. Similarly, Jensen and Lawson (2011) demonstrated that students in heterogeneous biotechnology lab groups, composed of undergraduates and graduates, earned better grades than those in homogeneous groups. However, Miller et al. (2012) found no differences in learning outcomes between homogeneous and heterogeneous groups in physics courses. This variability indicates that group composition can affect students' learning experiences in different ways. Again, student preferences for collaboration, interaction styles and working with certain peers can influence group success as Müller et al. (2022) emphasized. Some students may prefer working with familiar peers, while others may be more open to collaborating with new individuals, influencing the dynamics and outcomes of group work (Rienties et al., 2014). These preferences are often reflected in students' interaction patterns within online discussions, which provide insight into their engagement choices.

Human factors such as student motivation and participation barriers also shape the dynamics and success of collaborative groups, particularly in online settings. According to Deci & Ryan's Self-Determination Theory (2000), students who feel intrinsically motivated are more likely to engage meaningfully in group tasks. Conversely, external barriers, such as time zone differences, technology readiness or social identification can hinder effective group collaboration (Rienties et al., 2012; Wilkins et al., 2023). Participation barriers, including a lack of confidence, anxiety, or previous negative experiences in group work, can also lead to passive participation (Capdeferro & Romero, 2012; Hilliard et al., 2020). These human factors highlight the need for a more delicate approach to group formation that goes beyond surface-level attributes, taking into account deeper motivational and emotional factors.

In addition to individual human factors, contextual variables such as course content and platform features heavily influence group dynamics in online learning. Courses with collaborative or problem-solving-focused content tend to generate more meaningful interactions compared to those emphasizing individual learning. For example, a meta-analysis by Xu et al. (2023) found that collaborative problem solving enhances students' critical thinking and cognitive skills, particularly in settings where teaching type, group size, and learning scaffolds are key factors. Learning Management Systems (LMS) also play a vital role in shaping interation quality and frequency. Features such as peer feedback mechanisms, collaborative tools, or real-time communication channels can either facilitate or hinder effective group work (Arbaugh, 2014). Muñoz-Carril et al. (2021) identified perceived usefulness, ease of use, and enjoyment as crucial factors influencing students' satisfaction and learning impact in LMS environment. Some LMS platforms offer tools that promote social presence and engagement, while others may limit interactions, potentially affecting group cohesion (Rourke et al., 2001). Acknowledging these contextual variables reinforces the understanding that group dynamics in online learning environments are shaped by a combination of individual and external factors including course structure and the technological environment.

Given the various factors that influence group work, several methods can be used to form groups: random selection, self-selection, teacher selection, and automatic selection (Felder & Brent, 2001). Random selection, the simplest method, involves randomly assigning students to groups. While quick to implement, this method may result in unbalanced groups with varying skills and interests, which may not be optimal for collaboration and teamwork. Teacher selection involves manually assigning students to groups based on criteria, such as skills, interests, or learning needs. While this method can ensure balanced groups, it can be time-consuming, particularly for larger student cohorts. Moreover, the teacher may lack comprehensive knowledge, resulting in some students being placed in groups that are not optimal for their needs.

Self-selection allows students to choose their group mates based on shared interests or prior relationships. While this approach promotes collaboration and engagement, it can also lead to groups that lack diversity or are dominated by a few individuals. Students who struggle to find a group may feel excluded. Finally, automatic selection uses algorithms to assign students to groups based on specific criteria such as skills, interests, or learning needs. This method can be both effective and economically superior, generating optimal groupings (Muller et al., 2022; Vallès-Català & Palau, 2023). For example, Vallès-Català and Palau (2023) created an approach based on complex network theory to design an algorithm called Minimum Entropy Collaborative Groupings (MECG) to form heterogeneous groups more effectively. They found that the groups created with MECG were more effective with less uncertainty in comparison to the randomized groups.

The implementation of an automatic approach which would consider a range of learning needs and employ intelligent techniques for the structure of optimal student groups, is essential for promoting effective collaborative learning. Muller et al. (2022) highlight that automatic group formation based on deep level personality traits, such as extraversion and conscientiousness, can also lead to better satisfaction and group performance compared to other group formations. Krouska and Virvou (2019) emphasized genetic algorithms focused on three dimensions, academic, social, and cognitive dimensions for heterogeneous group formation and social learning.

Likewise, the criteria for algorithm-based group formation can vary by researcher and purpose, and there is no general or evidence-based recommendation for selecting consistent criteria. Additionally, it may not be possible to create an equally successful distribution for all groups simultaneously, as contextual trade-offs between different criteria often occur. As noted by Pai et al. (2015) and Muller et al. (2022), developing user-friendly online applications that facilitate the selection and weighting of relevant criteria could help address this challenge. These applications would allow for a more interactive and intuitive approach to grouping students, resulting in more accurate and comprehensive groupings. Yassine et al. (2022) also emphasized the importance of automating community detection techniques in online learning environments to optimize group formation.

Louvain algorithm: group formation by frequency of interaction

Interactions serve as a multifaceted metric that offers invaluable insights into various aspects of the educational experience in online learning environment. They operate as a tangible measure for student engagement, allowing educators to distinguish between active and passive learners for targeted support (Moore & Kearsley, 2011). Interactions also reveal naturally forming learning clusters within an online course, which are crucial for peer support and collaborative learning (Dawson, 2008; Rovai, 2002). By analyzing the flow of interactions, educators can identify key nodes or students who significantly influence information dissemination and discussion facilitation (Cela et al., 2015; Haythornthwaite, 2001). Understanding these interaction patterns helps form strategic groupings for assignments or projects, leveraging existing dynamics or encouraging new connections (Saqr et al., 2020). Moreover, both the quality and quantity of interactions have been shown to correlate with academic outcomes, making them a predictive tool for student success in numerous research studies (Arbaugh, 2008; Wise et al., 2014). Thus, interactions are not just incidental exchanges but critical metrics that can shape, assess, and enhance the online learning experience.

In educational contexts, Social Network Analysis (SNA) has been used extensively to capture how relationships within a network affect learning outcomes. SNA typically assesses the extent of interaction and collaboration within learning communities (Froehlich et al., 2020). Here, communities are typically defined as nodes with similar affiliations that are different from the rest of the network (Yang & Liu, 2010) or as cohesive network structures with the possibility of separation (Newman, 2018). This notion of cohesiveness is central to effective collaboration, as studies show cohesion positively influences team effectiveness and outcomes (Beal et al., 2003; LePine et al., 2008). Tekleab et al. (2009) highlight the group's united pursuit of goals as a key element of cohesion. Recently, Braun et al. (2020) defined it as "team members' shared commitment or attraction to their task/goal and to one another." In the context of community detection, especially in SNA, cohesion typically refers to the degree to which members within a particular community or group are more closely interconnected. A highly cohesive community is one where its members have many connections among themselves, but fewer connections to nodes outside the community. Thus, community can be separated when the internal links are larger than the external links, so a strong community has more internal links than external links (Wasserman & Faust, 1994).

Several community detection algorithms are widely used in SNA, including the Girvan-Newman, Label Propagation, K-means clustering and Louvain algorithms. Each offers distinct advantages and limitations within educational settings. The Girvan-Newman algorithm identifies communities by removing edges with high "betweenness" (Girvan & Newman, 2002). While effective for detecting hierarchical structures, it is computationally intensive, making it less suitable for large datasets typical in online learning environments. Label Propagation, on the other hand, is computationally efficient and can handle large networks by assigning labels to nodes and propagating labels to nodes until consensus is reached (Raghavan et al., 2007). However, its sensitivity to initial conditions can lead to inconsistent results, reducing its reliability for forming stable learning groups. K-means clustering (Lloyd, 1982) is another popular method, particularly for partitioning nodes into k predefined clusters by minimizing the distance between nodes and their assigned cluster centroid. While it is simple and computationally efficient, K-means requires the number of clusters to be specified in advance and assumes that clusters are spherical and non-overlapping, which can be a limitation in complex networks where learning groups naturally overlap or vary in size and shape.

A few studies have applied community detection algorithms, revealing their value in identifying group structures within online learning environments. For example, Jan and Vlachopoulos (2019) used the Integrated Methodological Framework with SNA to identify communities in online learning. They found that community-based learning and structural similarity between networks and communities make SNA a natural choice for deeper understanding of group interaction. Their method effectively identified the structure of both Community of Inquiry (CoI) and Community of practice (CoP). Yassine et al. (2021) also applied both Label Propagation algorithm and Louvain algorithm to learning networks confirming the utility of these community detection algorithms in detecting meaningful communities based on interaction patterns. Khaled et al. (2019) used K-means clustering in a social learning environment to identify 18 learner communities based on engagement patterns. The study showed how clustering can enhance personalized recommendations by matching learners with relevant content, thereby improving engagement. Similarly, Salas et al. (2016) applied K-means clustering in an LMS to group students into high- and low-performance clusters, helping educators tailor instructional strategies to better address student needs based on engagement and performance.

The Louvain algorithm (Blondel et al., 2008), a widely used and approved community detection method (Menczer et al., 2020; Traag et al., 2019), is interactive for detecting communities in networks through modularity optimization. The algorithm consists of two main steps that are repeated until modularity can no longer be improved, ultimately determining the partition with the highest modularity. Initially, the algorithm assigns each node to its own community, and examines neighboring communities for potential modularity gains if a node were to be reassigned. Nodes are then moved to the neighboring community that results in the maximum modularity gain or they remain in their current community if no gain is possible. Following this, the network is simplified by replacing communities with single super-nodes. Links between distinct communities are transformed into weighted links between corresponding super-nodes, with weights based on the sum of original link weights. Additionally, internal community links are converted into self-loops for super-nodes, where self-loop weights are equal to the sum of the original internal link weights. Upon completing these two steps, the algorithm iterates again with the new, smaller weighted network generated in the second step. This process of reassigning nodes to communities and aggregating communities into super-nodes continues until no further modularity improvement is possible. When the algorithm converges, the final partition of nodes into communities represents the network's structure with the highest modularity.

This algorithm has emerged as a prominent method due to its efficiency in scalability and optimizing modularity (Yassine et al., 2021). Yassine et al.'s systematic review study found that the Louvain method was widely used across different educational platforms, including LMSs and social learning environments, to detect communities based on learner interactions. For example, Adraoui et al. (2018) applied the Louvain algorithm in LMS forums to identify at-risk learners through their participation patterns. Similarly, Gottardo and Noronha (2012) utilized the algorithm to analyze community structures in distance education courses, offering insights into how students form learning clusters within asynchronous discussion forums. The algorithm's ability to optimize modularity and efficiently detect communities makes it particularly suited for analyzing interaction networks in structured online learning environments, regardless of the scale of the dataset (Blondel et al., 2008). Its effectiveness extends beyond MOOCs, demonstrating versatility in various educational contexts such as LMS platforms, where interaction data can provide meaningful insights into learner behaviors and engagement patterns (Mengoni et al., 2018).

The Louvain algorithm was selected for this study due to its scalability, efficiency, and suitability for educational settings where group dynamics evolve over time. The algorithm is renowned for its ability to optimize modularity, which ensures the detection of communities with the highest internal cohesiveness relative to external connections. This makes the Louvain algorithm particularly effective for analyzing collaboration dynamics in online educational environments, where interaction patterns among students vary significantly. Unlike other methods, which may rely on demographic or psychological traits, the Louvain algorithm operates purely on actual interaction data, making it both data-driven and adaptive to the natural communication tendencies of students. A key advantage of using an algorithm is that it not only reflects natural interaction preferences but also systematizes group formation by using student interactions as the primary input. In this sense, it functions as a hybrid between traditional self-selection and more structured automatic selection methods: it exploits students' natural preferences, much like self-selection, while avoiding the imbalances often associated with student-chosen groups. By ensuring high internal cohesion and a balanced distribution of roles within the learning community, the algorithm facilitates the creation of groups that are both collaboratively efficient and productive. This approach, using the Louvain algorithm with the SNA tool Gephi, offers a deeper understanding of how students naturally interact, enhancing the design of group learning activities and supporting peer collaboration at scale.

Purpose of the study

This study began with the primary objective of examining student interaction patterns across different online courses. Originally focused on identifying and comparing interaction patterns (RQ1), the study's scope expanded as initial findings suggested the potential utility of the Louvain algorithm for classifying these interactions. The Louvain algorithm is particularly effective at capturing naturally emerging interaction preferences, by revealing how students group themselves organically based on their interactions. This data-driven approach helps form groups based on real engagement patterns, aligning group formation with students' preferred interaction behaviors rather than arbitrary criteria like demographics or teacher assignments. Therefore, the study led to a deeper exploration of how the algorithm could effectively categorize interactions and influence our understanding of collaborative dynamics within student groups.

This adaptive, exploratory approach reflects the emergent nature of the research (Maxwell, 2012), leading to the development of additional research questions that address both the initial insights and the subsequent findings. The research questions are structured to reflect this progression:

1. Initial exploration of interaction patterns: How do interaction patterns among students vary across different online courses?
 a. Are there discernible patterns that differentiate courses based on interaction metrics?
 b. What global metrics (e.g., average degree, diameter, density) highlight the nature and quality of student interactions in these courses?
2. Emergent focus on classification effectiveness: Following the detection of significant patterns in initial analyses, how effectively does the Louvain algorithm classify students based on their interaction patterns?
 a. How do local metrics (such as clustering coefficient, density, weighted degree, and the presence of triangles) vary across the groups formed by the Louvain algorithm?
 b. What patterns within these metrics does the Louvain algorithm identify to distinguish one group from another?
 c. How can the correlations between local clustering coefficient, density, weighted degree, and triangles deepen our understanding of group collaboration and cohesiveness within student networks?
 d. Implications for Collaborative Dynamics: Given the insights from the Louvain algorithm's application, how does the algorithm's classification reflect the collaborative nature and cohesiveness of these groups?

This structure explicitly acknowledges the exploratory and iterative nature of the research, starting from a broad inquiry into interaction patterns and progressively focusing on specific analytical tools and their implications for understanding group dynamics in online learning environments. This phased approach underlines the flexibility and responsiveness of the research process to emerging data and insights, characteristic of exploratory studies.

Methods

Course selection

The selection criteria were to include courses with sufficient interaction data to provide robust analytical outcomes while ensuring the representativeness of various interaction dynamics within the online learning environment. Three sections of the same course were selected for this study based on minimum course duration, the amount of interaction and the number of participants (Table 1).

The course in online undergraduate communications provided a rich and structured dataset that is relevant for exploring the research theme of group composition and interaction patterns. The course emphasizes media literacy, critical thinking, and analytical viewing skills, incorporating lectures, readings, and film screenings. It focuses

Table 1 Course information

	Students	Posts	Topics
C4	53	808	9
C5	50	764	9
C6	51	839	9

on how films tell stories and make arguments through visual language, allowing students to interpret and analyze films to understand the cultural conditions that produced them. The Lesson Discussion Forums consisted of general question prompts to initiate discussion. There were eight lessons in total. Students were encouraged to engage in the discussion by drawing upon what they learned from the lesson. They were required to respond to the question thoughtfully and engage the ideas of their classmates by responding to at least two other students' posts. Respectful and constructive interaction was emphasized, and the total discussion participation average accounted for 20% of the final grade. This context is ideal for examining student interactions as it involves active engagement, critical discussions, and collaborative learning. By analyzing how students interact and form groups in the context of this course, the research can contribute to a broader understanding of collaborative learning in online environments.

Data collection and processing

The data for this study were provided from an undergraduate communications course conducted during the Spring 2021 semester. The course aimed to enhance students' media literacy through discussions centered on films and visual media. The anonymized dataset was provided by the institution's Canvas Learning Management System (LMS) team in 2022. Prior to the researchers receiving the dataset, all personally identifiable information was removed by an external data analyst who was not part of the research team. This step ensured full compliance with ethical standards, resulting in an IRB classification of the study as non-human subjects research, and waived the need for student consent. Before analysis could commence, the data underwent a series of preparatory steps to ensure it was in a suitable format for network analysis.

Data cleaning and transformation The first phase involved cleaning the raw data to remove any irrelevant or redundant information, such as administrative messages or incomplete entries. This streamlined the dataset, making it more focused and representative of actual student interactions. Next, the cleaned dataset was transformed into a format appropriate for network analysis. In this step, each anonymous student ID was treated as a node, and a reply to another student's post was treated as a directed edge between two nodes. For example, if Student A responded to Student B, an edge from Student A to Student B was created. This transformation effectively converted the text-based discussion into a graph-based network representation, where interactions among students could be quantitatively analyzed.

To map out the student interactions in a structured format, adjacency lists were compiled. These adjacency lists captured the connections between students, specifically noting who replied to whom, based on the columns present in the dataset, which included:

- Discussion entry id: A unique identifier for each post.
- Student name: An anonymized identifier for the student.
- Parent discussion entry id: Identifies the post to which a student replied.

- Created at: Timestamp showing when the post was made.
- Discussion message: The content of the student's post.

The adjacency lists served as the foundational structure for network visualization. Once the adjacency lists were completed for each section of the course, the next step was visualizing these interactions using network analysis tools.

The transformed data, including the adjacency lists, was imported into Gephi, a widely-used tool for network visualization and analysis. In Gephi, each student's interactions were represented as nodes (students) and edges (replies or interactions). This allowed for the creation of detailed sociograms, visually representing how students interacted with one another throughout the course. Several key metrics were calculated to analyze the interaction patterns within the network, including Weighted degree, Graph density, Diameter and Clustering coefficient. To ensure the accuracy and reliability of the data, the network statistics calculated in Gephi were cross-validated using statistical software, including R and SPSS.

Application of the Louvain algorithm A key part of the analysis was the detection of communities within the student interaction network. For this purpose, the Louvain algorithm—a modularity optimization technique built into Gephi—was used to identify naturally forming student subgroups based on their interaction patterns. The Louvain algorithm works by maximizing the modularity of the network, which essentially measures the strength of division between the identified communities. It starts by assigning each node to its own community, then iteratively evaluates whether moving a node into a neighboring community will increase the modularity score. Once modularity can no longer be improved, the communities are finalized. In this study, the algorithm allowed for the automatic detection of cohesive groups of students based on their actual interaction patterns—reflecting the frequency and intensity of their discussions. These communities were then further analyzed to understand their internal dynamics and cohesiveness. By following this process, the study was able to derive meaningful insights into how students interacted within an online learning environment, and how the Louvain algorithm could be used to identify and analyze naturally occurring communities based on their interactions.

Analytical approach

The analysis was conducted in two main stages to thoroughly explore student interaction patterns and the effectiveness of community detection algorithms in an educational setting. Initially planned as an analysis of student interaction patterns, the research evolved into an exploratory investigation into the utility of the Louvain algorithm for group composition. This shift occurred as initial data analysis revealed the algorithm's potential in uncovering hidden patterns in student interactions, providing a practical application of network analysis in collaborative learning. The study's methodology emphasizes the potential of network analysis to enhance the educational experience by optimizing collaborative learning dynamics and identifying natural group formations.

- Global network analysis: The study first examined the networks at a global level to understand their overall structure and interaction patterns. Metrics such as graph density, centralization, and clustering coefficient were calculated to provide insights into the topology and connectivity of the network. This global perspective helps to set the stage for a more detailed local analysis by identifying overarching patterns in student interactions.

- Local network analysis: This stage focused more deeply on the communities identified by the Louvain algorithm, a key component in our methodological approach. Here's how we conducted this analysis:
- Community detection: Using Gephi's built-in modularity function, which employs the Louvain algorithm, the study segmented the network based on patterns of connectivity among nodes. This process optimizes the modularity score to gauge the strength of divisions into distinct communities. The study kept the modularity resolution setting in Gephi at its default to ensure consistency across all datasets, based on preliminary tests that confirmed its effectiveness in accurately reflecting community structures within our data.
- Analysis of local metrics: the study examined several local metrics within the identified communities, such as the local clustering coefficient, within-community density, and the roles of central nodes. These metrics helped us to understand the cohesiveness and collaborative potential of the groups formed within the network. Specifically, the study explored:
- How local metrics vary across groups formed by the Louvain algorithm.
- What patterns within these metrics that the Louvain algorithm identifies to distinguish one group from another.
- How correlations between local clustering coefficient, density, weighted degree, and triangles contribute to our understanding of group collaboration and cohesiveness within student networks.

This detailed examination of local network dynamics enriches our understanding of how students interact within smaller, more defined groups and provides insights into the structural and functional aspects of these interactions. It highlights the utility of the Louvain algorithm in identifying meaningful interaction patterns and naturally occurring student communities.

Results

Networks can be analyzed from two perspectives: local and global. The local level examines specific nodes or edges, while the global level provides an overview of the entire network (Miele et al., 2019). Morrison et al. (2022) noted the importance of using both types of analysis for a thorough understanding. Following this idea, our study leverages both types of analysis to gain a comprehensive understanding of interaction dynamics. Our study began by assessing the networks globally to understand the overarching structure and subsequently focused on the group cohesion at the local level identified by the Louvain algorithm.

Global metrics (RQ1)

Global metrics offer understanding into the overall characteristics, topology, and structure of networks (Morrison et al., 2022). Metrics such as density, centralization,

Fig. 1 Networks of C4, C5, and C6 (from left)

Table 2 Global Network Metrics for three sections of an online course

Metric	Description	C4	C5	C6
Number of Nodes	Total number of students participating in the forum	53	50	51
Number of Edges	Total number of interactions between students	314	261	340
Graph Density	Ratio of actual connections to possible connections, higher values indicate more interconnectivity	0.228	0.213	0.267
Weighted Degree	Average level of connectivity per student, indicating interaction intensity	16.679	14.960	17.294
Path Length	Average shortest path between any two nodes, lower values indicate quicker information spread	2.111	1.996	1.807
Diameter	Longest of all the shortest paths in the network, smaller values suggest closer connectivity	5	4	4
Global Clustering Coefficient	Degree to which nodes cluster together, higher values suggest a tighter community	0.345	0.282	0.372

and clustering coefficient provide a holistic view of a network's structure, and particularly, density and diameter are often used to compare the structures and extents of different networks.

For the research question 1 as an initial exploration of interaction patterns, the study focuses on how interaction patterns among students vary across different online courses. The first sub-question asking if there are discernible patterns that differentiate courses based on interaction metrics, the study found there are clear visual differences in the connectivity patterns across courses from the network diagrams even with the same course content and instructor; thus the differences were based only on students participating the course discussion. Three networks of the courses C4, C5, and C6 are shown in Fig. 1.

These networks represent the structures of communication networks within different course environments. Each node represents a student, with its size representing its weighted degree, indicating the intensity of interactions. Nodes with darker colors indicate hubs or students with higher connectivity, emphasizing relative connectivity rather than just the number of connections. These different networks depict varied student participation and interaction behaviors. For instance, C6 appears denser with more connections than C4 and C5. This observation is substantiated when considering the following global metrics in Table 2.

The second sub-question, which focuses on identifying global metrics that can highlight the nature and quality of student interactions across the sections, evaluated several key indicators. The network for C4, depicted in the left diagram, reveals a relatively sparse structure with fewer connections per node. This observation aligns with the quantitative data, which shows a graph density of 0.228. This indicates that interactions among students are not as frequent. The average weighted degree of 16.679 suggests moderate interaction levels, but these interactions are not well-distributed throughout the network. The network diameter of 5 and path length of 2.111 indicate that information or interactions need to traverse more nodes to reach others, implying less direct communication pathways. Additionally, the global clustering coefficient (CC) of 0.345 points to a moderate tendency for students to form clusters, suggesting some level of group collaboration but not as cohesive as might be ideal.

The middle diagram for C5 shows a network with a moderate level of connectivity. The graph density is slightly lower than C4, at 0.213, suggesting fewer connections relative to the number of possible interactions. The average weighted degree is 14.960 further supports this observation, indicating that a relatively moderate frequency of interactions among students when compared to the higher interaction levels in C4. However, despite the lower overall density and weighted degree, the network structure of C5 reveals a more efficient configuration of connections. This is evidenced by a diameter of 4 and a shorter path length of 1.996, which are indicative of a more compact network. That is, although C4 has a slightly higher graph density (0.228) compared to C5 (0.213), it also has a higher diameter of 5 compared to 4 in C5. This contrast of higher density but larger diameter can seem counterintuitive initially, as typically, a higher density would suggest a more compact network, potentially leading to a smaller diameter.

This phenomenon can be explained with a few aspects in network. First, although a higher density indicates that students in C4 have more connections than C5, these connections might not be optimally distributed to minimize the longest path between nodes. If these connections in C4 are more clustered within smaller groups, the overall network might still have longer paths between distant nodes, resulting in a higher diameter. Thus, higher clustering within certain subgroups can lead to localized high density but still maintain a high diameter if these clusters are not well interconnected. In other words, students within a cluster might be very well connected but there could be a few connections between clusters, necessitating longer paths to travel between students in different clusters. Also, this could mean the lack of central hubs as seen in Fig. 1. Central hubs play a crucial role in reducing the diameter of a network by providing shortcuts that reduce the number of steps needed to connect any two nodes. C5 might have more effective central hubs that reduce the overall diameter, even with a lower density, whereas C4 had a higher density but lacked these central hubs, resulting in a less efficient overall network structure. The global clustering coefficient of 0.282, the lowest among the three courses, further suggests the presence of fewer tightly-knit groups, and that students were less likely to form cohesive collaborative clusters in C5.

C6, with the highest density of 0.267, the highest weighted degree of 17.294 and a diameter of 4, demonstrates the most robust as well as frequent interaction patterns.

The high density and lower diameter reflect a well-integrated network where students are not only connected but are part of a closely-knit community. The shortest path length of 1.807 further supports this, indicating efficient communication and interaction pathways throughout the course. The high global clustering coefficient of 0.372 signifies strong group cohesiveness, suggesting that students in C6 engage in frequent interactions, potentially enhancing collaboration.

These observations highlight the nuanced nature of network dynamics in online courses, where simply increasing the number of connections (density) does not necessarily lead to a more efficiently connected network (lower diameter). Strategic interventions to enhance inter-cluster connectivity and the presence of central hubs can significantly impact the overall structure and effectiveness of student interactions.

Local metrics (RQ2)

Local network analysis focused more deeply on the communities identified by the Louvain algorithm, a key component in our methodological approach. Using Gephi's built-in modularity function, which employs the Louvain algorithm, the study segmented the network based on patterns of connectivity among nodes. This process optimizes the modularity score to gauge the strength of divisions into distinct communities. The study kept the modularity resolution setting in Gephi at its default to ensure consistency across all datasets, based on preliminary tests that confirmed its effectiveness in accurately reflecting community structures within our data. The study examined several local metrics within the identified communities, such as the local clustering coefficient, within-community density, and the roles of central nodes. These metrics helped us to understand the cohesiveness and collaborative potential of the groups formed within the network. Specifically, the study explored how local metrics vary across groups, patterns within these metrics, how correlations between them contribute to our understanding, and finally reflecting on potential implications for collaboration and cohesiveness based on the algorithm's classification.

These examinations of local network dynamics enriched our understanding of how students interact within smaller, more defined groups and provides insights into the structural and functional aspects of these interactions. The study also highlights the utility of the Louvain algorithm in identifying meaningful interaction patterns and naturally occurring student communities. The Louvain algorithm helps to classify groups based on the actual interaction patterns as a communication detection method to uncover high modularity groups within large networks. It basically optimizes the modularity of a network by aggregating nodes into communities in a way that maximizes the density of links inside communities compared to links between communities. For this research question, the study analyzed the groups after the Louvain algorithm had identified groups as a post-classification analysis to understand their structural characteristics and how each metric can be used post-classification to interpret the nature of the groups formed in Table 3.

This table provides a clear framework for understanding the role and interpretation of local metrics in analyzing the structure of groups identified by Louvain algorithm.

Table 3 Local level metrics in groups

Course	Group	LCC	Triangles	Density	W. Degree
C4	Group0	0.485	2.455	0.291	5
	Group1	0.429	0.857	0.381	4.25
	Group2	0.818	5.625	0.607	7.75
	Group3	0.771	7.286	0.81	15.7
	Group4	0.658	4.500	0.607	8
	Group5	0.470	4.500	0.364	5
C5	Group6	0.619	4.714	0.714	9.143
	Group7	0.463	3.700	0.467	5.8
	Group8	0.528	2.500	0.667	9
	Group9	0.595	3.000	0.619	7.714
	Group10	0.385	1.000	0.417	4
	Group11	0.015	0.091	0.236	3.6
C6	Group12	0.571	3.000	0.619	6.286
	Group13	0.562	3.429	0.619	6
	Group14	0.562	4.000	0.5	4.667
	Group15	0.381	2.143	0.524	4.286
	Group16	0.622	3.222	0.417	4.667
	Group17	0.576	3.857	0.667	6.571
	Group18	0.900	4.200	0.9	9.6

Table 4 The role of local metrics in post-Louvain classification analysis

Metric	Purpose	Post-classification use	Interpretation of collaborative nature and cohesiveness
Clustering Coefficient (LCC)	Indicates how close nodes in a network tend to cluster together	Higher values suggest tightly knit members within a community	High LCC indicates strong internal communication and cohesion
Density	Measures the proportion of actual connections to possible connections within a group	Higher density suggests a well-connected group	High density implies effective communication and collaboration
Triangles	Counts the number of closed triplets within the network	Measures local cohesiveness and strong subgroup relationships	Many triangles indicate robust mutual connections within subgroups
Weighted Degree	Reflects the sum of the weights of the edges connected to a node	Identifies influential or central nodes within each group	High weighted degree indicates potential key individuals crucial for information flow

Variation in local metrics (RQ2_a)

To evaluate the Louvain algorithm's classification effectiveness, the study analyzed the local metrics within the groups it identified. Table 4 presents these metrics for groups in courses C4, C5, and C6, highlighting variations in clustering coefficient, number of triangles, density, and weighted degree. The selection of these local metrics was strategic for understanding group cohesion in online learning environments. To be specific, these metrics can offer a comprehensive analysis of how students interact within their groups, the strength of their connections, and the roles of central nodes in facilitating collaborative learning.

The clustering coefficient indicates the likelihood that a node's neighbors are also connected. High LCC values suggest tight-knit groups where members are likely to form triangles. For example, in course C6, Group 18 exhibits an exceptionally high LCC of 0.900, indicating that nearly every student in this group forms a triangle with their peers, reflecting very high cohesiveness. High LCC values indicate that students are likely to engage in mutual interactions, forming strong triadic relationships. This cohesiveness is beneficial for collaborative learning as it fosters a supportive environment where students can easily share ideas and resources. On the other hands, Groups in C4 and C5 show more variability. To be specific, Group 2 in C4 has a high LCC of 0.818, while Group 11 in C5 has an extremely low LCC of 0.015, indicating a lack of cohesive clusters in that group. This variability can highlight areas where intervention is needed to foster better group cohesion and collaboration.

Density measures how many of the possible connections in a group are actually present. Higher density values suggest more interconnected and potentially more collaborative groups. Group 18 in C6 also shows the highest density at 0.9, indicating that almost all potential connections are present, suggesting an environment where students are highly interconnected. High-density groups are likely to have efficient communication channels, enhancing collaborative efforts. Density values vary significantly in other courses. For instance, Group 3 in C4 has a high density of 0.81, suggesting strong internal connectivity, while Group 11 in C5 has a much lower density of 0.236, indicating sparse connections.

The weighted degree reflects the total interaction volume a node (student) has. Higher values suggest that some students are particularly active and central within their groups. Group 3 in C4 shows a high weighted degree of 15.7, indicating that students in this group are very active and central to the group's interactions. In C6, Group 18 also exhibits a high weighted degree of 9.6, further demonstrating the high level of interaction within this group.

The number of triangles in a network indicates local clustering and the presence of cohesive subgroups. Groups in C6, such as Group 18 with 4.200 triangles, indicate robust mutual connections and strong subgroup formation. C4's Group 3 shows a high number of triangles (7.286), suggesting strong cohesiveness within the group. Conversely, C5's Group 11 has very few triangles (0.091), indicating weak subgroup formation.

Table 5 Descriptive statistics for the group cohesiveness

Metric	LCC	Density	Triangles	W. Degree
N (Valid)	19	19	19	19
Mean	0.548	0.549	3.373	6.686
Median	0.562	0.607	3.429	6
St. Deviation	0.189	0.173	1.704	2.875
Skewness	− 0.796	0.077	0.12	1.79
Kurtosis	2.853	− 0.315	0.645	4.389

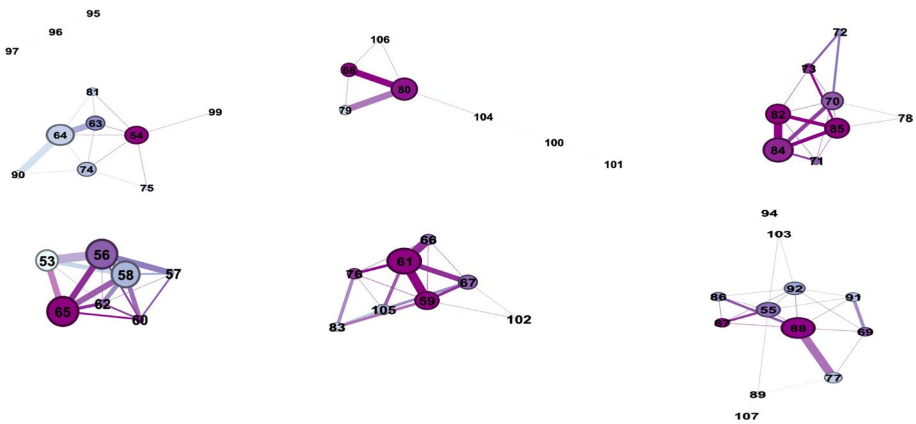

Fig. 2 C4 Group Networks (0–2 left to right, top row; 3–5 left to right, bottom row)

Patterns identified by the Louvain algorithm (RQ 2_b)

Table 5 presents descriptive statistics to analyze group cohesiveness, while Figs. 2 to 4 visualize the group networks of three sections of the same course, identified using the Louvain Algorithm. The average Clustering Coefficient (LCC) across all groups is 0.548, with a median of 0.562. This indicates a general tendency for students to form

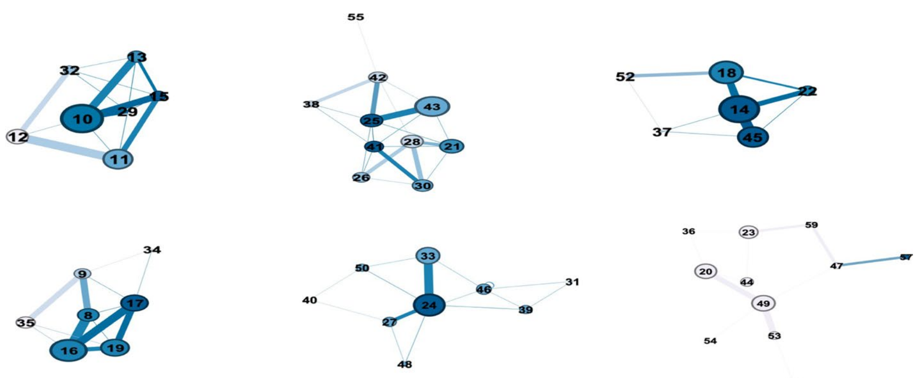

Fig. 3 C5 Group Networks (6–8 left to right, top row; 9–11 left to right, bottom row)

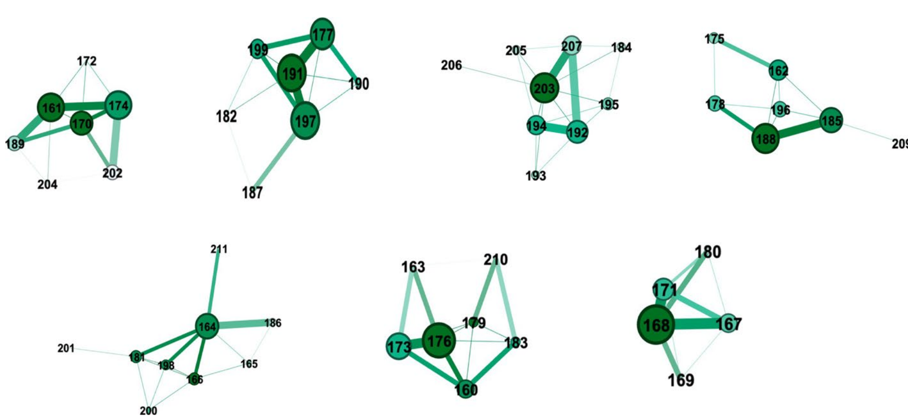

Fig. 4 C6 Group Networks (12–15 left to right, top row; 16–18 left to right, bottom)

moderately cohesive clusters. Groups with high LCC values, such as Group 18 in C6 (0.900) (Fig. 4) and Group 3 in C4 (0.771) (Fig. 2), are characterized by strong local clustering. These groups are likely to have robust internal interactions where students frequently form triadic relationships, enhancing collaborative learning, whereas groups with low LCC values, such as Group 11 in C5 (0.015) (Fig. 3), show minimal clustering, indicating less cohesive and more fragmented interactions. The mean Density of 0.549 and median of 0.607 suggest that, on average, groups maintain about half of their potential connections. High-density groups, like Group 18 (Fig. 4) in C6 (0.9) and Group 3 in C4 (0.81) (Fig. 2), exhibit strong interconnectedness among members, fostering efficient communication and collaboration whereas groups with lower density, such as Group 11 in C5 (0.236) (Fig. 3), are less interconnected, which may impede effective group interactions and collaboration.

The average number of Triangles is 3.373, with a median of 3.429. Triangles indicate the presence of closed triplets, which are crucial for local cohesion. Groups like Group 3 in C4 (7.286) (Fig. 2), and Group 6 in C5 (4.714) (Fig. 3), have a high number of triangles, suggesting strong subgroup formations and robust mutual connections while Groups with fewer triangles, such as Group 11 in C5 (0.091) (Fig. 4), lack these strong mutual connections, reflecting weaker subgroup cohesion.

The mean Weighted Degree is 6.686, with a median of 6. This indicates a moderate level of interaction volume per node. Groups with high weighted degrees, like Group 3 in C4 (15.7) (Fig. 2), and Group 6 in C5 (9.143) (Fig. 4), have central nodes with significant interaction volumes, suggesting that certain students play key roles in maintaining group dynamics while groups with lower weighted degrees, such as Group 11 in C5 (3.6) (Fig. 3), may have less active participation, indicating potential areas for intervention to boost engagement.

To summarize, the mean values for LCC (0.548), density (0.549), and triangles (3.373) across all groups suggest moderate overall cohesiveness and connectivity. These metrics indicate that, on average, groups are fairly cohesive, but there's variability. When considering standard deviation, the variability in LCC (0.189), density (0.173), and triangles (1.704) confirms that some groups are more cohesive and interconnected than others, as visually represented in the diagrams. Finally, negative skewness in LCC (-0.796) suggests a tail towards lower values, indicating some groups with very low cohesiveness. High kurtosis in weighted degree (4.389) points to the presence of outliers, likely groups with exceptionally high central node activity.

Table 6 Spearman's Rank-Order Correlations among Group Metrics

Metric	Triangles	Density	W. Degree
Local CC	0.762**	0.677**	0.801**
Triangles	–	0.531*	0.719**
Density	0.531*	–	0.860**
Weighted Degree	0.719**	0.860**	–

** p < .01, *p < .05

Correlational analysis of group dynamics (RQ 2_c)

The study further utilized Spearman's rank-order correlation coefficients to analyze the relationships between various structural metrics within student interaction networks, as identified by the Louvain algorithm. These metrics—Local Clustering Coefficient (LCC), Triangles, Density, and Weighted Degree—offer insights into the collaborative nature and cohesiveness of the hypothetical groups formed within the online courses (Table 6).

A strong positive correlation (0.762) was observed between the Local Clustering Coefficient (LCC) and the number of Triangles. This correlation suggests that groups with higher LCC, indicative of more cohesive units, also tend to have a greater number of closed triangles. Such a structure implies that these groups not only have tightly-knit connections among members but also exhibit a high level of mutual interconnectedness, facilitating robust subgroup formations within the network. Similarly, the LCC showed a significant positive correlation with Density (0.677), indicating that more cohesive groups are generally denser. This relationship underscores that as groups become more cohesive, the proportion of actual connections relative to possible connections increases, reflecting a more interconnected network where collaborative interactions are likely more frequent and substantial. The correlation between LCC and Weighted Degree (0.801) further highlights that groups characterized by higher cohesiveness also feature nodes with higher levels of interaction volume. This finding suggests that in groups where members are closely connected, there are typically one or more individuals who are central to maintaining and facilitating these interactions, thus enhancing the group's overall communicative activity.

Examining the relationships involving Triangles, a moderate positive correlation (0.531) with Density was found, supporting the notion that networks with more closed triangles tend to be denser. This correlation aligns with the idea that increased group interconnectivity, as evidenced by numerous triadic relationships, contributes to a denser overall network structure, potentially enhancing the collaborative environment. Furthermore, the strong positive correlation between Triangles and Weighted Degree (0.719) indicates that groups with numerous triangles not only have dense connections but also contain highly active nodes. This suggests that these groups not only form tight-knit communities but are also dynamic, with substantial engagement across members.

The most noticeable correlation was between Density and Weighted Degree (0.860), illustrating that denser groups typically feature nodes with higher weighted degrees. This robust relationship implies that in networks where connections are plentiful, individual nodes tend to engage more actively, contributing significantly to the network's liveliness and the efficacy of collaborative efforts.

These correlations collectively present a comprehensive picture of group dynamics within the studied networks. The strong interdependencies between cohesiveness (as measured by LCC and Triangles) and both the quantity (Density) and quality (Weighted Degree) of interactions suggest that more cohesive groups are not only well-connected but also highly active. This detailed understanding of network structures provides valuable insights into how different aspects of group interactions contribute to the collaborative potential and educational outcomes within online learning environments.

Reflection of collaborative nature and cohesiveness by the algorithm's classification (RQ2_d)

The Louvain algorithm's classification offers valuable insights into the collaborative nature and cohesiveness of student groups within the network. While these groups are hypothetical, their interaction patterns reveal important findings about potential group dynamics. By using the algorithm to identify these patterns, we can systematically validate the structural properties that signify effective collaboration and group cohesion.

Groups with high clustering coefficients and density values reflect strong potential for collaboration and mutual support. For example, Group 18 in C6 exhibits a high local clustering coefficient (LCC) of 0.900 and a density of 0.9. The visual representation of this group (Fig. 4) shows dense interconnections, suggesting that, if this group were real, students would likely form a tightly-knit network with frequent interactions. This high level of cohesion indicates a strong potential for effective collaboration and mutual support among group members. Such metrics and visual representations provide empirical evidence of cohesive group structures essential for collaborative learning environments. Conversely, groups with lower clustering coefficients and density values, such as Group 11 in C5 (LCC of 0.015 and density of 0.236), indicate weaker connections and less cohesive interactions. The corresponding network diagram (Fig. 3) shows sparse connections and isolated nodes. These metrics suggest that this group might struggle with collaboration due to fragmented interactions. The ability to systematically identify such groups using the Louvain algorithm highlights its utility in detecting areas that require pedagogical intervention to improve group connectivity and engagement.

However, a key insight emerging from this classification is the unevenness in group dynamics, which mirrors the natural variation in student participation in online learning environments. While some groups exhibit strong collaborative potential, others are less cohesive, and this disparity reflects the real-world challenge of uneven engagement. The Louvain algorithm, by design, captures these variations rather than enforcing artificial balance. This is a strength in that it reflects real student preferences and interaction tendencies, but it also reveals a limitation: not all groups will be equally active or cohesive, as student engagement varies naturally.

Groups with high weighted degrees often feature central nodes that play a significant role in maintaining communication and interaction within the group. For instance, Group 3 in C4 has a high weighted degree of 15.7, indicating substantial interaction volume per node. The network diagram (Fig. 2) shows several nodes with multiple connections, suggesting that key participants are driving the group's interactions. These central figures are crucial for facilitating communication and fostering a collaborative environment within the group. Identifying these key participants is vital for educators to support and leverage them in enhancing overall group effectiveness.

Additionally, the presence of triangles within the network highlights the formation of cohesive subgroups. Group 3 in C4, with a high number of triangles (7.286), demonstrates robust mutual connections. The visual diagram (Fig. 2) shows multiple interconnected triangles, indicating strong subgroup formations within the group. These subgroups reflect a high level of local cohesion, which is essential for effective collaboration and support within the larger group. This detailed analysis using the Louvain algorithm helps in understanding the underlying microstructures that contribute to overall cohesiveness of the group.

Despite these strengths, it is also important to recognize the challenge posed by less active students in online environments. The Louvain algorithm groups students based on their actual interactions, meaning that less active students may be grouped together, resulting in weaker overall group cohesion. This reflects the real-world dynamics of online learning, where not all students are equally engaged. While this may seem like a limitation, it provides an opportunity for educators to intervene. Identifying groups with low clustering coefficients and density values offers a chance to provide additional support or adjust group structures to enhance collaboration.

Balanced metrics across different groups provide a comprehensive view of their potential for effective collaboration. Group 6 in C5, with a moderate to high LCC of 0.619, density of 0.714, and weighted degree of 9.143, represents a healthy hypothetical group. The network diagram (Fig. 3) shows a well-connected network, suggesting that students in this group would likely engage and support each other effectively. Such balanced metrics indicate strong collaborative potential and robust group dynamics. The algorithm's ability to detect these patterns validates its use in educational settings for fostering optimal group formations.

The correlation analysis further reinforces the understanding of these group dynamics. The strong positive correlation between LCC and Triangles (0.762) indicates that groups with higher clustering coefficients tend to have more closed triangles, suggesting robust mutual connections and cohesive subgroup formations. The correlation between LCC and Density (0.677) shows that more cohesive groups are also denser, reflecting a higher proportion of actual connections relative to possible connections. The very strong correlation between LCC and Weighted Degree (0.801) suggests that highly cohesive groups also have members with high interaction volumes, highlighting the importance of active participants in maintaining group connectivity. Additionally, the moderate correlation between Triangles and Density (0.531) and the strong correlation between Triangles and Weighted Degree (0.719) indicate that groups with more interconnected triads tend to be denser and have members with higher levels of engagement. The very strong correlation between Density and Weighted Degree (0.860) underscores that denser groups feature more active nodes, contributing significantly to the network's vibrancy and collaborative potential.

Ultimately, the classification of student groups by the Louvain algorithm offers a systematic reflection of the collaborative nature and cohesiveness of these hypothetical groups. High values in these metrics correspond to more cohesive and potentially collaborative groups, as evidenced by dense, interconnected network diagrams. Conversely, low values indicate fragmented groups with weak interactions, highlighting areas needing targeted interventions. This analysis enhances our understanding of group dynamics, offering insights that can inform the design of more effective collaborative learning environments.

Implications

The application of the Louvain algorithm within online learning environments provides valuable insights and practical applications in two primary areas: group formation and as an analytical tool for understanding collaborative dynamics.

Group Formation The Louvain algorithm offers a powerful post-process for forming groups in educational environments where students interact continuously over time. One example is cohort-based programs where students progress through multiple courses together. In these settings, the Louvain algorithm can be employed to analyze interaction patterns from previous courses or earlier stages of a program to form cohesive, balanced groups for subsequent collaborative projects. This is particularly valuable in cohort-based undergraduate programs, where the same group of students is enrolled in a sequence of courses over several years. The algorithm optimizes group formation by leveraging actual student interactions, much like self-selection, but adds a data-driven layer of insight. It systematizes the process by ensuring groups are formed based on real interaction dynamics, leading to high-cohesion groups that are more likely to effectively collaborate and support one another.

Another highly applicable context for the Louvain algorithm is within semester-long courses that include both individual and group-based learning. For instance, many engineering or STEM programs follow a structure where the first 4–6 weeks are focused on individual learning and lab work, occasionally coupled with weekly online discussions. As the course transitions to collaborative projects later in the semester, the Louvain algorithm can be used to analyze early discussion data and optimize group formation for these final projects. This method ensures that students are grouped based on natural interaction patterns that have already developed in the first half of the course, promoting effective teamwork and improving learning outcomes in the latter half. The algorithm's ability to form cohesive groups based on real interactions makes it a practical tool for enhancing collaboration and peer support during critical project phases.

Analytical Tool for Collaborative Dynamics Beyond its use for group formation, the Louvain algorithm serves as a powerful analytical tool for examining collaborative dynamics in online learning environments. By clustering students based on natural interaction patterns, the algorithm allows educators to explore the structural properties of these groups. A key advantage of using the Louvain algorithm is its ability to uncover natural affinities between students by detecting interaction frequencies and patterns.

Although the groups formed by the algorithm are hypothetical, they reflect real student preferences and interaction tendencies. This can reveal insights into the underlying structure of collaboration, such as which students engage more actively with others and how tightly-knit or fragmented their clusters are. By analyzing these algorithmically generated groups, educators can identify potential imbalances in group cohesion and participation. For instance, groups with high clustering coefficients and density values may indicate strong, cohesive clusters, while groups with lower metrics could suggest fragmented or isolated students who may need additional support. This identification enables educators to spot trends in group collaboration that may not be immediately visible, providing opportunities to intervene where necessary to foster more inclusive and supportive group interactions.

Additionally, the Louvain algorithm can track changes in group dynamics over time, allowing educators to assess how student interactions evolve. This temporal analysis of group dynamics offers valuable insights into how collaboration develops, revealing whether groups become more cohesive or disintegrate over time. Such insights are critical for educators seeking to implement real-time interventions that enhance participation and balance within groups. Therefore, the Louvain algorithm provides a data-driven approach to analyzing collaborative learning that goes beyond surface-level metrics. By focusing on naturally occurring interaction patterns, the algorithm enables a deeper understanding of the structural aspects of student collaboration, helping to refine instructional strategies that promote more equitable and effective teamwork in online learning environments.

Through both its applications in group formation and collaborative analysis, the Louvain algorithm offers educators a robust tool for enhancing and understanding collaborative learning dynamics, ultimately contributing to more effective and cohesive learning experiences.

Conclusion

Few existing studies have explored the use of algorithms like the Louvain algorithm in educational contexts, and those that have primarily focused on identifying communities or groups, rather than using the algorithm as a tool for group composition. This study contributes to the field by shifting the focus toward the application of the Louvain algorithm as a means of actively forming student groups based on real-time interaction patterns. By leveraging the algorithm's capacity to detect naturally occurring clusters of interactions, this research introduces a data-driven approach to group formation that enhances collaboration and peer support in online learning environments. In doing so, this study addresses a critical gap in the literature by extending the utility of the Louvain algorithm beyond community detection, demonstrating its practical value in composing more cohesive and collaborative groups. The findings provide educators with a powerful tool for optimizing group dynamics in a way that traditional methods of group formation—such as random selection or self-selection—cannot achieve. This novel application adds a new dimension to the existing body of knowledge and offers a practical solution for enhancing collaborative learning in online environments. This research lays the groundwork for future studies on the use of network science algorithms in education.

However, there are several limitations to consider in this exploratory research. First, the study is built on the premise that frequent interactions serve as a strong indicator of potential group cohesion and collaboration. The Louvain algorithm uses this premise to form groups based on actual interaction patterns, often reflecting students' natural preferences to engage with certain peers. In many cases, this result in groups with high internal cohesion, which enhances the likelihood of collaboration. This reliance on organically formed student interactions, rather than arbitrary or surface-level grouping methods, is a key strength of the algorithm. However, it is important to note that frequent interactions alone do not always guarantee the depth of collaboration. Some students may engage in frequent but superficial exchanges, which might lead to the groups that appear cohesive but lack substantial collaborative dynamics. Despite this, the study highlights that cohesive groups formed through frequent interactions, provide a solid starting point for fostering collaboration.

Secondly, while the Louvain algorithm effectively identifies clusters of students based on their interaction patterns, it does not always ensure that the groups formed are equally active or cohesive. In some cases, the algorithm may produce groups where certain individuals dominate interactions, leading to imbalanced participation. This limitation arises from the algorithm's reliance on interaction frequency as the primary metric. As a result, some groups may exhibit high activity and collaboration, while others remain less engaged like the real world dynamics of online learning, where not all students are equally engaged.

Third, this study relied on rich interaction data, which may not be available in all learning environments. In contexts with sparse or uneven interactions, the Louvain algorithm may struggle to form balanced and effective groups. Furthermore, the algorithm's exclusive focus on interaction patterns overlooks student attributes, such as academic background and learning preferences—factors that can potentially influence group cohesion and performance, as noted in previous research. Without these contextual data points, the algorithm's ability to fully optimize group formation remains limited.

Finally, while the Louvain algorithm is known for its scalability across datasets of various sizes, its application to smaller datasets requires caution. Although the algorithm performed well for mid to large networks, its modularity optimization may be less effective in very small or exceptionally large datasets. Future research should aim to replicate these findings across diverse educational contexts and explore the longitudinal effects of algorithm-assisted group formation on student engagement and achievement. Additionally, combining interaction patterns with other metrics, such as content or discourse analysis, or incorporating qualitative data, would enhance group formation and ensure more balanced and effective groupings. This approach would mitigate the effects of unequal group activity and provide a more comprehensive understanding of collaboration dynamics. Such efforts will help refine the role of technology in improving the quality of interaction in online learning communities.

The insights gained from this study represent a step toward creating online educational environments where collaboration and interaction are enhanced through data-driven strategies. This research lays the foundation for future studies that could further explore how technology can be used to improve the quality of interaction and collaboration in online learning spaces.

Funding
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Declarations

Competing interests
We have no known conflict of interest to disclose.

Received: 18 June 2024 Accepted: 19 November 2024

References

Adraoui, M., Retbi, A., Idrissi, M. K., & Bennani, S. (2018). Network visualization algorithms to evaluate students in online discussion forums: A simulation study. *2018 international conference on Intelligent Systems and Computer Vision (ISCV)* (pp. 1–6). Piscataway Township: IEEE.

Arbaugh, J. B. (2008). Does the community of inquiry framework predict outcomes in online MBA courses? *International Review of Research in Open and Distributed Learning*. https://doi.org/10.19173/irrodl.v9i2.490

Arbaugh, J. B. (2014). System, scholar or students? Which most influences online MBA course effectiveness? *Journal of Computer Assisted Learning, 30*, 349–362. https://doi.org/10.1111/jcal.12048

Beal, D. J., Cohen, R. R., Burke, M. J., & McLendon, C. L. (2003). Cohesion and performance in groups: A meta-analytic clarification of construct relations. *Journal of Applied Psychology, 88*, 989–1004. https://doi.org/10.1037/0021-9010.88.6.989

Bell, S. T. (2007). Deep-level composition variables as predictors of team performance: A meta- analysis. *Journal of Applied Psychology, 92*(3), 595–615. https://doi.org/10.1037/0021-9010.92.3.595

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 2008*(10), P10008.

Braun, M. T., Kozlowski, S. W. J., Brown, T. A. R., & DeShon, R. P. (2020). Exploring the dynamic team cohesion-performance and coordination-performance relationships of newly formed teams. *Small Group Research*. https://doi.org/10.1177/1046496420907157

Capdeferro, N., & Romero, M. (2012). Are online learners frustrated with collaborative learning experiences? *The International Review of Research in Open and Distributed Learning, 13*(2), 26–44. https://doi.org/10.19173/irrodl.v13i2.1127

Cela, K. L., Sicilia, M. Á., & Sánchez, S. (2015). Social network analysis in e-learning environments: A preliminary systematic review. *Educational Psychology Review, 27*(1), 219–246. https://doi.org/10.1007/s10648-014-9276-0

Cohen, E. G. (2023). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64*(1), 1–35.

Dawson, S. (2008). A study of the relationship between student social networks and sense of community. *Journal of Educational Technology and Society, 11*(3), 224–238.

Donovan, D. A., Connell, G. L., & Grunspan, D. Z. (2018). Student learning outcomes and attitudes using three methods of group formation in a nonmajors biology class. *CBE—Life Sciences Education, 17*(4), ar60.

Felder, R. M., & Brent, R. (2001). Effective strategies for cooperative learning. *Journal of Cooperation and Collaboration in College Teaching, 10*(2), 69–75.

Froehlich, D. E., Van Waes, S., & Schäfer, H. (2020). Linking quantitative and qualitative network approaches: A review of mixed methods social network analysis in education research. *Review of Research in Education, 44*(1), 244–268.

Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education, 2*(2–3), 87–105.

Girvan, M., & Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences, 99*(12), 7821–7826.

Gottardo, E., & Noronha, R. V. (2012). Social networks applied to distance education courses: analysis of interaction in discussion forums. In: Proceedings of the 18th Brazilian symposium on Multimedia and the web. 355–358.

Haythornthwaite, C. (2001). Social network analysis: An approach and technique for the study of information exchange. *Library and Information Science Research, 18*(4), 323–342.

Hilliard, J., Kear, K., Donelan, H., & Heaney, C. (2020). Students' experiences of anxiety in an assessed, online, collaborative project. *Computers and Education, 143*, 103675.

Jan, S. K., & Vlachopoulos, P. (2019). Social network analysis: A framework for identifying communities in higher education online learning. *Tech Know Learn, 24*, 621–639. https://doi.org/10.1007/s10758-018-9375-y

Jensen, J. L., & Lawson, A. (2011). Effects of collaborative group composition and inquiry instruction on reasoning gains and achievement in undergraduate biology. *CBE—Life Sciences Education, 10*(1), 64–73.

Khaled, A., Ouchani, S., & Chohra, C. (2019). Recommendations-based on semantic analysis of social networks in learning environments. *Computers in Human Behavior, 101*, 435–449.

Kozlowski, S. W., & Bell, B. S. (2003). Work groups and teams in organizations. In I. B. Weiner (Ed.), *Handbook of Psychology*. Wiley.

Krouska, A., & Virvou, M. (2019). An enhanced genetic algorithm for heterogeneous group formation based on multi-characteristics in social-networking-based learning. *IEEE Transactions on Learning Technologies, 13*(3), 465–476.

Lee, M. K., Sharma, P. (2023). Applying community detection algorithms to examine group formation in online discussions, AERA, Chicago.

LePine, J. A., Piccolo, R. F., Jackson, C. L., Mathieu, J. E., & Saul, J. R. (2008). A meta-analysis of teamwork processes: Tests of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychology, 61*, 273–307. https://doi.org/10.1111/j.1744-6570.2008.00114.x

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory, 28*(2), 129–137.

Maxwell, J. A. (2012). *A realist approach for qualitative research*. Sage.

Menczer, F., Fortunato, S., & Davis, C. (2020). *A first course in network science*. Cambridge University Press. https://doi.org/10.1017/9781108653947

Mengoni, P., Milani, A., & Li, Y. (2018). Community graph elicitation from students' interactions in virtual learning environments. In *Computational Science and Its Applications–ICCSA 2018: 18th International Conference, Melbourne, VIC, Australia, 2018, Proceedings, Part III*. Springer International Publishing. 18. 414–425.

Miele, V., Matias, C., Robin, S., & Dray, S. (2019). Nine quick tips for analyzing network data. *PLoS Computational Biology, 15*(12), e1007434.

Miller, H. B., Witherow, D. S., & Carson, S. (2012). Student learning outcomes and attitudes when Biotechnology lab partners are of different academic levels. *CBE-Life Sciences Education, 11*(3), 323–332.

Moore, M. G., & Kearsley, G. (2011). *Distance education: A systems view of online learning*. Cengage Learning.

Morrison, D., Bedinger, M., Beevers, L., et al. (2022). Exploring the raison d'etre behind metric selection in network analysis: A systematic review. *Appl Netw Sci, 7*, 50.

Müller, A., Bellhäuser, H., Konert, J., & Röpke, R. (2022). Effects of group formation on student satisfaction and performance: A field experiment. *Small Group Research, 53*(2), 244–273.
Muñoz-Carril, P. C., Hernández-Sellés, N., Fuentes-Abeledo, E. J., & González-Sanmamed, M. (2021). Factors influencing students' perceived impact of learning and satisfaction in computer supported collaborative learning. *Computers and Education, 174*, 104310.
Newman, M. (2018). *Networks*. Oxford University Press.
Ouyang, F., & Scharber, C. (2017). The influences of an experienced instructor's discussion design and facilitation on an online learning community development: A social network analysis study. *The Internet and Higher Education, 35*, 3. https://doi.org/10.1016/j.iheduc.2017.07.002
Pai, H. H., Sears, D. A., & Maeda, Y. (2015). Effects of small-group learning on transfer: A meta-analysis. *Educational Psychology Review, 27*, 79–102.
Peacock, S., & Cowan, J. (2019). Promoting sense of belonging in online learning communities of inquiry at accredited courses. *Online Learning, 23*(2), 67–81. https://doi.org/10.24059/olj.v23i2.1488
Raghavan, U. N., Albert, R., & Kumara, S. (2007). Near linear time algorithm to detect community structures in largescale networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics, 76*(3), 036106.
Rienties, B., Alcott, P., & Jindal-Snape, D. (2014). To let students self-select or not: That is the question for teachers of culturally diverse groups. *Journal of Studies in International Education, 18*(1), 64–83.
Rienties, B. C., Giesbers, S. J. H., Tempelaar, D. T., Lygo-Baker, S., Segers, M. S. R., & Gijselaers, W. H. (2012). The role of scaffolding and motivation in CSCL. *Computers and Education, 59*(3), 893–906. https://doi.org/10.1016/j.compedu.2012.04.010
Rodríguez, D., Sicilia, M. Á., Sánchez-Alonso, S., Lezcano, L., & García-Barriocanal, E. (2011). Exploring affiliation network models as a collaborative filtering mechanism in e- learning. *Interactive Learning Environments, 19*(4), 317–331.
Rourke, L., Terry Anderson, D., Garrison, R., & Archer, W. (2001). Methodological issues in the content analysis of computer conference transcripts. *International Journal of Artificial Intelligence in Education, 12*, 8–22.
Rovai, A. P. (2002). Sense of community, perceived cognitive learning, and persistence in asynchronous learning networks. *The Internet and Higher Education, 5*(4), 319–332.
Rovai, A. P. (2007). Facilitating online discussions effectively. *The Internet and Higher Education, 10*(1), 77–88. https://doi.org/10.1016/j.iheduc.2006.10.001
Salas, D. J., Baldiris, S., Fabregat, R., & Graf, S. (2016). Supporting the acquisition of scientific skills by the use of learning analytics. In Advances in Web-Based Learning–ICWL 2016: 15th International Conference, Rome, Italy, October 26–29, 2016, Proceedings. Springer International Publishing. 15 (281–293)
Saqr, M., Nouri, J., Vartiainen, H., & Tedre, M. (2020). Robustness and rich clubs in collaborative learning groups: A learning analytics study using network science. *Scientific Reports, 10*(1), 14445. https://doi.org/10.1038/s41598-020-71483-z
Shea, P., Li, C. S., & Pickett, A. (2006). A study of teaching presence and student sense of learning community in fully online and web-enhanced college courses. *The Internet and Higher Education, 9*(3), 175–190.
Singh, J., Singh, L., & Matthees, B. (2022). Establishing social, cognitive, and teaching presence in online learning—A panacea in COVID-19 pandemic, post vaccine and post pandemic times. *Journal of Educational Technology Systems, 51*(1), 28–45.
Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research, 69*(1), 21–51. https://doi.org/10.3102/00346543069001021
Tekleab, A. G., Quigley, N. R., & Tesluk, P. E. (2009). A longitudinal study of team conflict, conflict management, cohesion, and team effectiveness. *Group and Organization Management, 34*, 170–205. https://doi.org/10.1177/1059601108331218
Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. *Science and Reports, 9*, 5233. https://doi.org/10.1038/s41598-019-41695-z
Vallès-Català, T., & Palau, R. (2023). Minimum entropy collaborative groupings: A tool for an automatic heterogeneous learning group formation. *PLoS ONE, 18*(3), e0280604.
Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge University Press.
Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer supported collaborative learning. *Instructional Science, 33*(1), 1–30. https://doi.org/10.1007/s11251-004-2322-4
Wenger, E. (1998). Communities of practice: Learning as a social system. Systems thinker. *International Journal of Educational Management, 9*(5), 2–3. https://doi.org/10.1108/IJEM-10-2022-0412
Wilkins, S., Butt, M. M., Hazzam, J., & Marder, B. (2023). Collaborative learning in online breakout rooms: The effects of learner attributes on purposeful interpersonal interaction and perceived learning. *International Journal of Educational Management*. https://doi.org/10.1108/IJEM-10-2022-0412
Wise, A. F., Hausknecht, S. N., & Zhao, Y. (2014). Attending to others' posts in asynchronous discussions: Learners' online "listening" and its relationship to speaking. *International Journal of Computer-Supported Collaborative Learning*. https://doi.org/10.1007/s11412-014-9192-9
Wise, A. F., Saghafian, M., & Padmanabhan, P. (2012). Towards more precise design guidance: Specifying and testing the functions of assigned student roles in online discussions. *Educational Technology Research and Development, 60*, 55–82.
Xie, K., Yu, C., & Bradshaw, A. C. (2014). Impacts of role assignment and participation in asynchronous discussions in college-level online classes. *The Internet and Higher Education, 20*, 10–19.
Xu, E., Wang, W., & Wang, Q. (2023). The effectiveness of collaborative problem solving in promoting students' critical thinking: A meta-analysis based on empirical literature. *Humanit Soc Sci Commun, 10*, 16. https://doi.org/10.1057/s41599-023-01508-1
Yang, D., & Liu, J. L. (2010). *Discovering communities from social networks: Methodologies and applications* (pp. 331–346). Springer.

Yassine, S., Kadry, S., & Sicilia, M. A. (2021). Application of community detection algorithms on learning networks. The case of Khan Academy repository. *Computer Applications in Engineering Education, 29*(2), 411–424.

Yassine, S., Kadry, S., & Sicilia, M. A. (2022). Detecting communities using social network analysis in online learning environments: Systematic literature review. *Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery, 12*(1), e1431.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.