Beyond posting frequency: How network metrics and textual readability relate to engagement in online discussion

Minkyung Lee[1], Priya Sharma[2]

Department of Institutional Effectiveness, Univerisity of Houston, Downtown
[2] Department of Learning and Performance Systems, Pennsylvania State University

Minkyung Lee, Ph.D.
Director at the Department of Institutional Effectiveness
University of Houston, Downtown
Email: minklee010@gmail.com

Priya Sharma
Associate Professor at the Department of Learning and Performance Systems
Penn State University
Email: pus3@psu.edu

We have no known conflict of interest to disclose.

Corresponding concerning this article should be addressed to Minkyung Lee, minklee010@gmail.com.

ABSTRACT

This study investigates how student interaction patterns and the readability of their discussion posts relate to engagement in asynchronous online discussion. Data from an undergraduate online course were analyzed using social network analysis (SNA), focusing on metrics such as indegree, outdegree, betweenness, eigenvector, closeness, and clustering coefficient. The statistical results indicate that students who received more responses from peers tended to have higher eigenvector and betweenness centralities, reflecting stronger connections to influential peers and a greater role in bridging groups. However, frequent postings and participation were not strongly associated with broader engagement. Additionally, the readability assessment of the discussion posts rated as more accessible and clearly written were often associated with broader engagement. Those posts were associated with students occupying more central positions in the network, while more complex or dense writing corresponded to peripheral network positions and fewer responses. These findings suggest that both interaction structure and communication clarity are associated with patterns of engagement in online learning discussions. The study demonstrates the usefulness of combining network analytics and textual assessment for informing instructional practices and learning platform design and points to opportunities for further research using longitudinal approaches.

INTRODUCTION

Asynchronous discussions, a significant component of many Learning Management Systems (LMS), provide a flexible platform for student interactions, transcending the limitations of time and geographical constraints (Hrastinski, 2008). However, while the asynchronous format offers flexibility, it also introduces distinct challenges including the potential for diminished social presence and inherent delays in receiving feedback (Borup et al., 2012; Dennen, 2008; Kovanović et al., 2015). Beyond these issues, the dynamics of asynchronous discussions are influenced by various factors. For instance, the quality and timing of a post can significantly shape the level of engagement it receives (Wise & Cui, 2018; Koh et al., 2010; Gao et al., 2013). Additionally, the role of the instructor and the nature of discussion prompts can significantly impact either catalyzing or stifling student interactions (Dennen, 2008; Xie & Correia, 2023).

Despite these insights, there remains a gap in the literature regarding the intricate network dynamics inherent to asynchronous discussion, particularly concerning the disparities in response rates among students. To bridge this gap, this study utilizes the concept of Social Learning Analytics (SLA) to identify patterns of participation, collaboration, and engagement, particularly by analyzing interactions and behaviors within digital networks. While traditional methods of gauging engagement often relied on self-reported scales, interviews, or teacher observations, SLA provides a behavior-based method to extract engagement features. This method can be pivotal in predicting dropout rates, flagging students at risk, and enabling timely interventions by educators (Dixson, 2015; Saqr & Lopez-Pernas, 2021). SLA includes various types of analyses—network analysis, discourse analysis, content analysis, disposition analysis, and context analysis (Ferguson & Shum, 2012). However, not all these types of analytics need to be applied in any given study. Researchers can tailor their approach based on the available data and the focus of their research.

For this study, the focus is on network analysis within SLA. Network analysis offers a powerful lens through which to examine the relational and interactional dynamics between students. Drawing inspiration from past research (Wise et al., 2014; Wise & Cui, 2018), this study utilizes anonymized discussion interaction data from an online business course . Through post-hoc analysis rooted in learning analytics, we aimed to explore how students' positions and roles within the discussion network are related to their level of engagement in asynchronous discussion using Social Learning Analytics (SLA). By leveraging network analysis, this research explores how students connect with one another, the role of certain individuals as central figures in discussions, and how these connections influence engagement.

This approach aligns with the extracted analytics perspective, which focuses on analyzing interaction data after the activity has taken place. Extracted analytics, as noted by Wise et al. (2014), provide valuable insights into student engagement by examining digital traces post-activity. These insights offer practical applications for improving the design of asynchronous discussions, fostering more balanced and inclusive interactions.

By utilizing SLA's network analysis, this study captures the social dynamics in asynchronous discussions, addressing gaps in understanding why certain students receive more responses than others. This research explores the network dynamics in asynchronous discussions to elucidate factors influencing response rates and unravel the complexities of student interactions within these forums. The insights derived from this research will offer practical, actionable knowledge to enhance engagement and interaction in asynchronous discussions, fostering more effective, inclusive, and balanced educational discussions.

Social Learning Analytics (SLA)

Social Learning Analytics (SLA) offers a powerful approach to understanding learning processes in collaborative environments, particularly by analyzing interactions and behaviors within digital networks. SLA focuses on the social dimension of learning, using digital traces from student activities to identify patterns of participation, collaboration, and engagement that influence learning outcomes (De Laat & Prinsen, 2014; Shum & Ferguson, 2012). This approach is particularly useful in higher education, where social learning environments such as online discussions and MOOCs are becoming increasingly important.

SLA differs from traditional learning analytics by placing emphasis on the collective nature of learning, rather than solely on individual performance. Through network analysis techniques, SLA can reveal how students build connections, contribute to knowledge construction, and engage in communities of practice (Kaliisa et al., 2022). In these environments, students' social roles and their positioning within the network can have significant effects on their learning experiences and the outcomes of their peers. Studies that apply SLA (e.g., De Laat & Prinsen, 2014) highlight the utility of these tools in raising awareness about students' social mobility and helping themselves make informed decisions about where to participate and whom to collaborate with.

Recent work, including the systematic review by Kaliisa et al. (2022), identifies key trends in SLA research, particularly its application in formal online learning settings. This review of 36 studies between 2011 and 2020 shows that SLA is predominantly used to understand students' collaborative learning behaviors through methods like social network analysis (SNA) and discourse analysis. However, the review also highlights gaps, such as the need for more sophisticated analytical techniques that integrate multiple methods, including epistemic network analysis and multimodal network analysis. These newer techniques allow for deeper insights into how knowledge and relationships evolve over time in collaborative learning environments. Furthermore, they underscore that while SLA research has focused on understanding student interactions, there is still limited use of these insights to support teaching practices. Few studies have shared SLA results with educators in ways that help inform instructional design or classroom interventions. This is an area where SLA can grow, particularly by developing tools that provide real-time feedback to teachers, enabling them to tailor their instructional strategies to promote more equitable participation and collaboration.

In this study, SLA's inherent social dimension, particularly through network analysis, is crucial in understanding why some students are more central in discussions than others. The use of SNA to analyze relationships between participants offers a powerful lens to investigate social learning processes, where those with higher centrality often play key roles in shaping the group's discourse and knowledge construction (Carvalho et al., 2016; Kaliisa et al., 2022).

The focus on SLA also aligns with evolving pedagogical trends in higher education, where collaborative and participatory learning models are becoming more prominent. SLA supports this shift by providing the tools to map out social connections and collaborative patterns in digital learning environments. These insights are particularly valuable in asynchronous online discussions, as they help to highlight the social structures and dynamics that can either facilitate or inhibit productive learning interactions.

Network analysis in education

While Social Network Analysis (SNA) has long been utilized in fields such as sociology and business, its application in educational research has seen more recent growth, particularly in online learning environments. Scholars like Haythornthwaite (2000) and Palonen & Hakkarainen (2000) were among the first to apply SNA in educational settings, recognizing its value for understanding relationships among students in digital learning platforms. These early studies revealed that SNA provides novel insights into interaction patterns and social structures that are otherwise difficult to capture using traditional observational methods (Russo & Koesten, 2005).

In educational contexts, centrality measures such as indegree, outdegree, betweenness, and eigenvector are critical for understanding the prominence and influence of specific students within a network. These measures help to dissect how certain students become focal points, receiving more responses and consequently exerting more influence within online learning environments. For instance, Saqr & Lopez-Pernas (2022) explored the use of centrality measures from SNA in collaborative learning settings. Their study concludes that degree and eigenvector centrality measures can serve as reliable predictors of student success in collaborative learning settings. However, short-path centralities (closeness and betweenness) were less consistent and should be used cautiously for interpretation. A systematic review and meta-analysis conducted by Saqr et al. (2022) focused on 19 studies and explored the utility of SNA in assessing students' interactions in Computer-Supported Collaborative Learning (CSCL) environments, identifying centrality measures as strong indicators of achievement. Notably, degree centralities consistently showed the strongest correlation with final course grades, while eigenvector centralities also demonstrated positive and significant correlations across the studies.

Indegree centrality focuses on the number of incoming links a student receives, which translates to the number of responses a student receives in online forums. Several studies have shown that high indegree centrality often correlates with perceived popularity or authority and prestige (Baldwin et al., 1997; Galikyan & Admiraal, 2019; Vlachopoulos, P., 2012). Students with high indegree centrality are often those who demonstrate subject matter expertise or take on mentoring roles within the forum, becoming central figures around whom discussions gravitate. Outdegree centrality considers the number of outgoing links a student initiates by interacting with others. This measure highlights active participants who contribute to various discussions, often driving the conversation and encouraging others to respond. Such students are crucial for sustaining dialogue and engagement within collaborative environments (Wise et aal., 2013). Betweenness centrality measures the extent to which a student acts as a bridge between other students in the network. Students with high betweenness centrality play a key role in connecting disparate discussion threads or groups. These "bridge builders" may not be the most active contributors, but they are pivotal in enriching dialogue and creating a cohesive learning community. They help ensure that knowledge and ideas are disseminated across different clusters of students (Saqr et al., 2020b). Eigenvector centrality builds on indegree by not only counting the number of connections a student has but also considering the influence of those connections. Students with high eigenvector scores are connected to other highly central individuals, which can signify a "rich club" phenomenon (Saqr et al., 2020b). This suggests that influential students are more likely to engage with other key players, thus reinforcing their central position within the network (De Laat, 2012). In educational settings, this could mean that students who interact with influential peers are more likely to become influential themselves.

Another important concept in network analysis is the clustering coefficient, which indicates the density of a student's ego network—how interconnected a student’s peers are. A high clustering coefficient suggests that a student’s contacts are well-connected to each other, forming a tightly-knit community (Saqr et al., 2020b). These ego networks consist of an ego (the individual), alters (contacts), and ties representing their relationships (Borgatti, 1998). While Borgatti’s work (2024) has laid the foundation for quantitatively analyzing ego networks—focusing on metrics such as size, density, and centrality—our interpretation incorporates a more contextual lens. Drawing from Hollstein (2011) and Edwards (2010), we explore the measurable connections as well as the dynamics behind these relationships. This approach deeply explores how individuals might perceive the importance of their ties and how their positions within the network influence engagement. Thus, although ego networks are often analyzed numerically, interpreting them through this contextual framework allows for a deeper understanding of the social dynamics and relational contexts that influence student interactions within online discussions.

Network analysis has become an invaluable method for examining how students engage in collaborative learning, particularly in online and blended learning environments. In such settings, social network analysis helps researchers track the development of social relationships and the formation of learning communities (Kaliisa et al., 2022). Studies utilizing network analysis consistently find that students with high centrality scores—especially high indegree and betweenness—often play crucial roles in group dynamics. They facilitate the flow of information, encourage collaboration, and act as knowledge brokers. For example, Bakharia & Dawson (2011) used SNA to investigate student engagement in online discussion forums, identifying students with high centrality measures as critical figures in fostering collaboration. These students served as hubs, connecting more peripheral participants to the larger group and ensuring that discussions remained dynamic and inclusive. Similarly, Haythornthwaite and De Laat (2010) demonstrated that students with high centrality are more likely to help others integrate into the learning community, making them essential for promoting group cohesion and knowledge sharing. Network analysis, especially through centrality measures, offers educators a detailed lens to understand how students interact, influence one another, and contribute to the learning process. By identifying central students and their roles, educational practitioners can design interventions to support more balanced and inclusive participation, enhancing the overall collaborative learning experience.

Asynchronous discussion and factors influencing response rates

Asynchronous discussions have become a vital component of many LMSs, fostering interaction among students without the constraints of time and place. These platforms provide flexibility, allowing students to engage in more thoughtful and reflective participation (Hrastinski, 2008). While much of the existing research has focused on enhancing learner engagement or deepening cognitive engagement, there is limited exploration into why certain contributions receive more responses than others, an area that warrants further investigation.

Post quality is a key determinant of response rates in asynchronous discussions. Posts that are well-structured, coherent, and intellectually stimulating are far more likely to attract responses. According to Hew & Cheung (2011), contributions that invite further exploration or challenge existing viewpoints tend to generate higher levels of engagement. Conversely, posts that lack depth or simply reiterate previous comments may restrain interaction and reduce the likelihood of receiving additional responses. Gao et al. (2013) further support this, noting that posts introducing fresh perspectives or posing thought-provoking questions are more likely to create meaningful exchanges, which in turn leads to higher response rates.

In further support of the importance of content-driven engagement, Wise and Cui's (2018) study on participation in Massive Open Online Courses (MOOCs) highlights the relationship between forum contributions and learning outcomes. Their findings show that students who actively engage in content-related discussions are more likely to succeed in the course compared to non-contributors, suggesting that content quality and relevance are crucial to meaningful participation. However, they also found that the quantity of posts related to course content slightly predicted higher grades, while measures of social network centrality—indicative of students' connectedness in discussions—did not add predictive value beyond the number of posts. This reinforces the argument that the substance of contributions, particularly those related to the course material, is more influential in determining engagement and outcomes than mere social presence or centrality within a network.

While the timing of posts may contribute to their visibility, the relationship between timing and response rates is less straightforward. Early posts may have the advantage of longer exposure, but Dringus and Ellis (2010) found that discussions often experience peaks and declines in activity, and late contributors—particularly those posted outside peak periods—are less likely to respond to earlier posts. However, this temporal aspect is only one part of the equation, and it is the quality of the content that remains the primary driver of engagement. Posts that engage participants with thoughtful insights or challenging perspectives are more likely to sustain ongoing discussion, regardless of when they are posted.

The role of the instructor and the design of discussion prompts also contribute to these dynamics. Instructors can stimulate discussions by offering insightful contributions or posing follow-up questions, but their participation must be carefully balanced. As highlighted by Xie and Correia (2024), while instructor presence and frequency can positively affect the quantity and quality of student participation, excessive involvement may shift the focus from peer-to-peer interaction to an instructor-centered discussion (Boling et al., 2012).Well-crafted discussion prompts that encourage students to connect with the material and each other, combined with appropriately timed instructor input, can lead to richer, more engaged conversations.

Moreover, the design of the technological platform used for discussions can affect student interaction. Platforms that provide intuitive navigation and notifications for new posts enhance participation by keeping students informed and prompting timely responses (Gros & García-Peñalvo, 2023; Rose & Ferschke, 2016). This helps maintain a continuous flow of communication, ensuring that posts are not overlooked due to design limitations or lack of notification.

Despite a range of studies exploring how to increase engagement or deepen cognitive participation, there remains a gap in understanding the network structures that underlie response rates in asynchronous discussions. Specifically, why some students consistently receive more responses than others, or why certain posts attract a disproportionately high number of replies, is not fully explained in the literature. By employing a network analytical approach, the current study seeks to decode the mechanisms driving these variations, focusing on the interaction patterns that contribute to differing response rates. This approach aims to uncover the social dynamics, providing insights into how certain students become central figures in discussions—receiving high levels of engagement, or high indegree—while others may remain on the periphery. These criteria reflect standard practices in quantitative and educational research for ensuring valid group comparisons and reducing the influence of statistical outliers (Bryman, 2016; Tabachnick & Fidell, 2007). Also, though the primary analysis focused on comparing students who received high and low numbers of responses (indegree), all network metrics for each included student were calculated within the context of the entire classroom network. That is, each student's position and pattern of interaction was analyzed in relation to the full set of classroom participants, not in isolation. Students with insufficient participation or highly atypical response/posting ratios were excluded to ensure their network positions could be meaningfully interpreted and compared within the broader classroom structure.

PURPOSE OF STUDY

The objectives of this research are:

1. To characterize the network positions and roles of students in asynchronous discussions by comparing participants with high and low levels of received responses, utilizing quantitative network metrics and a contextual analysis of ego networks to understand interaction dynamics.
2. To investigate associations between various network centrality measures and student response rates (indegree), and to describe how different network roles (e.g., bridges, connections to influential peers) relate to engagement patterns within the discussion forum.

To achieve these aims, the study employs a step-by-step approach. First, students are identified and compared using quantitative social network metrics. Next, both whole-network and ego network analyses are conducted to examine their structural characteristics and roles. Finally, selected ego networks are interpreted qualitatively to provide deeper insight into individual and group interaction patterns. This sequential process enables a comprehensive examination of both overall trends and specific cases.

METHODS

Context and data selections

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(full length semester), the amount of interaction, and the number of participants (Table 1).

Table 1. Course description

	Number of students	Number of posts
C4	53	808
C5	50	764
C6	51	839

The selected online course in 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 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.

From this dataset, we derived various network metrics, including indegree, outdegree, closeness, betweenness, eigenvector, clustering coefficient, and skewness with the igraph package built in R.

Following this, students in the first quartile by responses received (bottom 25%), and those in the third quartile by responses received (top 25%) were identified.

Participant data were excluded from students who missed over three of the eight discussion prompts. Additionally, to focus the analysis on participants with more conventional back and forth interaction styles within the discussion forums, we excluded students whose ratio of weighted indegree to weighted outdegree fell into the extreme quartiles (i.e., the bottom 25% and top 25% of this ratio). This step was taken to reduce the variance potentially introduced by participants with highly skewed interaction patterns such as those who posted many messages but received very few replies, or those who posted very little but received a disproportionately high number of replies relative to their own output. The final sample consisted of 38 students: 24 from the high response receiver category and 14 from the low response receiver category.

Data analysis

In this study, the analysis was carefully designed to explore the complexities of student engagement in asynchronous discussions, identifying patterns that differentiate highly engaged students from their less engaged counterparts and explore how broadly these dynamics contribute to fostering more effective online discussion. To achieve this, the study leveraged a range of analytical techniques, rooted in SNA, to uncover the roles that students play within the context of the entire classroom interaction network and how these roles impact their visibility and engagement. Importantly, all network metrics and comparisons, including those focused on high and low response receivers, were computed in relation to the structure of the whole group, ensuring the individual patterns are interpreted within the authentic social context of the virtual class.

The first step in the analysis involved constructing a network from the discussion data representing each student as a node, and mapping every response from one student to another as a directed edge. This structure allowed visualization and quantification of interaction flows, laying the groundwork for deeper analyses. Once the network was established, several key centrality metrics were computed to capture various aspects of students network positions. Indegree centrality, measuring the number of responses received, served as the primary indicator of student visibility or influence. Outdegree centrality was also calculated to assess individual activity levels based on message sent. Additional metrics included betweenness centrality reflecting bridging roles between network subgroups, eigenvector centrality to identify connections to influential peers, closeness centrality to measure interaction proximity, and clustering coefficient to evaluate the extent of tightly-knit group interactions.

To explore potential differences between students with high and low response rates, ANOVA was used to compare centrality metrics between the two groups. Subsequently, a multiple linear regression analysis was conducted to explore the statistical associations between various network metrics (outdegree, betweenness, eigenvector, closeness, and clustering coefficient) and indegree centrality. Recognizing inherent interrelations among these centrality measures (Haythornnthwaite & De Laat, 2010; Saqr et al., 2020), the primary goal was to understand how these different aspects of students' network positions and activity levels collectively related to the number of responses they received, rather than to isolate the independent predictive strength of each individual metric.

In addition to the quantitative network analyses, a qualitative analysis of three illustrative ego networks was conducted to offer a more granular perspective on individual students' interaction patterns. The selection of three ego networks was not intended to achieve statistical representativeness, but rather to serve as illustrative cases that provide context-rich insights into the diverse engagement patterns within the class. As Hollstein (2011) observes, "*a common strategy of analyzing qualitative data is to give a detailed account of individual cases by way of 'thick descriptions'... geared toward tracing how actions or events unfold and the impact they have in order to make them comprehensible... Understanding the individual case is the objective and, as such, an 'end in itself'*" (p. 413). This approach is consistent with qualitative social network research and allows for the exploration of peer interaction patterns that aggregate metrics alone cannot reveal (see also Nimmon & Atherley, 2021).

The qualitative assessment incorporated both structural metrics (e.g., size, density, centrality) and contextual interpretation, following established practice in educational SNA (Edwards, 2010; Macfadyen & Dawson, 2010). These cases complement group-level findings by illustrating how engagement manifests concretely within the classroom network.

Additionally, to further contextualize individual students' network positions and response rates, readability analyses were conducted on student contributions using established textual complexity measures. Readability scores, including Flesch Reading Ease (Flesch, 1948), Flesch-Kincaid Grade Level (Kincaid et al., 1975), Gunning Fog Index (Gunning, 1968), and SMOG Index (McLaughlin, 1969), were computed to quantify the textual accessibility of student posts. These readability measures are widely used in educational research to assess text complexity, influencing reader comprehension and engagement (Crossley et al., 2011). Integrating readability analyses allowed us to explore whether the textual accessibility of students' contributions was associated with their relational positions and visibility within the interaction network, providing a richer and more nuanced understanding of engagement dynamics in asynchronous discussions.

RESULTS

Analysis of Variance

Anonymized interaction data were extracted from the discussion forums of the online course. A key objective of this analysis is to understand the distinct characteristics that separate high-response-receiving students from their low-response-receiving peers. By applying ANOVA, we compared the two groups across various network metrics (e.g., indegree, eigenvector centrality) to test for statistically significant differences. This approach was motivated by the need to identify specific network factors contributing to differences in engagement levels.

Table 2 presents the ANOVA results for the various network metrics. For the Indegree metric, there is a highly significant difference between the groups, as indicated by an $F(1, 36) = 91.905$, $p < .001$. The effect sizes, $\eta^2 = .719$ and $\omega^2 = .705$, suggest a strong effect. Conversely, Outdegree does not show a significant difference between groups, $F(1, 36) = .237$, $p = .629$. Betweenness centrality showed a significant variation, $F(1, 36) = 11.174$, $p = .002$, with effect sizes of $\eta^2 = .237$ and $\omega^2 = .211$. Eigenvector centrality demonstrates a notable difference between the groups, $F(1, 36) = 36.245$, $p < .001$, with effect sizes of $\eta^2 = .502$ and $\omega^2 = .481$. Similarly, Closeness centrality is significantly different, $F(1, 36) = 9.225$, $p = .004$, with effect sizes of $\eta^2 = .209$ and $\omega^2 = .182$. The Clustering Coefficient does not differ significantly between the groups, $F(1, 36) = .830$, $p = .368$.

Based on the ANOVA results, indegree, eigenvector, betweenness, and closeness centralities show significant differences between students receiving high and low responses, with indegree and eigenvector centrality exhibiting the largest effect sizes (Table 2). Conversely, outdegree and clustering coefficient did not show significant differences between the two groups in this dataset, suggesting they were not strong differentiating factors for response rates in this context. These findings highlight which aspects of network centrality are most associated with variations in the number of responses students received in online discussions.

Table 2. ANOVA results for the network metrics

Metric	F	P (Sig.)	η^2	ω^2
Indegree	91.905	<.001***	.719	.705
Outdegree	.237	.629	.007	-.020
Betweenness	11.174	.002**	.237	.211
Eigenvector	36.245	<.001***	.502	.481
Closeness	9.225	.004**	.209	.182
Clustering coefficient	.830	.368	.023	-.005

Note. *p < .05. **p < .01. ***p < .001.

Multiple Linear Regression Analysis

Indegree centrality, representing the number of responses a student receives, serves as the outcome variable in a multiple linear regression analysis. This analysis aimed to explore the statistical association between indegree and other network metrics (outdegree, betweenness, eigenvector centrality, closeness, and clustering coefficient. Given the inherent conceptual and mathematical interrelationships among centrality measures within a network, the primary goal was to describe how these different facets of students' network positions and activities were related to the number of responses they received in this specific dataset, rather than to isolate the independent predictive strength of each individual metric. The findings are intended to offer insights into the structural characteristics associated with varying levels of received responses in online discussion.

The multiple linear regression explored the predictive power of various network metrics (outdegree, betweenness, eigenvector, closeness, and clustering coefficient) on the indegree centrality of students in an online discussion forum. The overall regression model showed a statistically significant relationship with indegree, $F(5, 31) = 148.659$, $p < .001$ (Table 3). The R^2 value was .960 (Adjusted R^2 =.954), indicating that approximately 96% of the variance in indegree was accounted for by the linear combination of the included network metrics. This exceptionally high R^2 value is largely attributable to the inherent intercorrelations among centrality measures when used to model another centrality measure from the same network. Thus, while the model demonstrates a strong descriptive fit to the data, the focus of the interpretation is on the nature of the associations of individual metrics with indegree, bearing in mind these interdependencies.

Within the predictors, Outdegree significantly negatively predicted Indegree ($\beta = -0.379$), $t(31) = -5.062$, $p < .001$, meaning that students who send out more messages tend to receive fewer responses. Betweenness Centrality was a significant positive predictor with $\beta = 0.134$, $t(31) = 2.432$, $p < .05$, meaning that students who act as bridges in the communication network (connecting disparate groups) tend to receive more responses. Eigenvector Centrality notably emerged as a strong positive predictor with $\beta = 1.021$, $t(31) = 13.684$, $p < .001$, indicating students connected to other influential students are more likely to receive numerous responses. However, Closeness Centrality did not significantly predict Indegree, $t(31) = 0.136$, $p > .05$, suggesting that the speed at which a student can reach or be reached by others in the network doesn't necessarily influence how many responses they receive. Lastly, Clustering Coefficient was observed to be a marginally significant negative predictor, with $\beta = -0.093$, $t(31) = -2.007$, $p \approx .05$, suggesting that students in tightly-knit clusters may receive slightly fewer responses, possibly due to the insular nature of such clusters which might limit wider engagement.

Table 3. Regression analysis results to predict indegree centrality from the network metrics

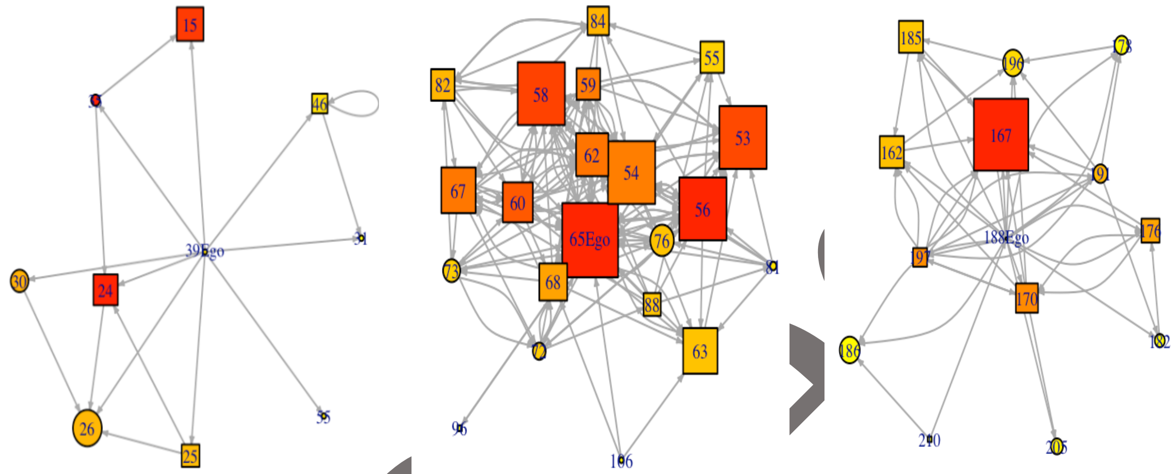
	B	SE B	β	t	p
(Constant)	5.560	4.579		1.214	.234
Outdegree	-0.832	0.164	-0.379	-5.062	***
Betweenness	0.019	0.008	0.134	2.432	*
Eigenvector	32.126	2.348	1.021	13.684	***
Closeness	80.471	590.573	0.009	0.136	
Clustering Coefficient	-7.340	3.658	-0.093	-2.007	.

Note. [a] $R = .980$, $R^2 = .960$, Adjusted $R^2 = .954$, $F(5, 31) = 148.659$, $p < .001$
*p < .05. **p < .01. ***p < .001.

Ego Networks

To provide a deeper understanding of how network positions manifest in individual student interactions, we selected three illustrative case studies for detailed ego network analysis :Student 65 from course C4 (representing a high-response receiver), Student 39 from C5 (a low-response receiver with zero indegree), and Student 188 from C6 (also a low-response receiver but with a slightly different interaction profile). These students were purposefully chosen to represent contrasting levels of engagement (indegree), as well as distinct positions regarding eigenvector and betweenness centralities, as identified in the ANOVA and regression analyses. Since all three courses were structured identically, these cases allowed us to examine individual interaction dynamics within a consistent educational context.

The ego networks of these students ((Figure 1; metrics detailed in Table 4) represent distinctive interaction patterns, further enriched by readability analyses of their posts (Table 5).



a.39 b.65 c.188

Figure1. Ego networks of 39 (low), 65 (high) and 188 (low) from the left

- Node size: In-degree centrality, larger nodes indicate greater number of responses received.
- Node color: Eigenvector centrality, darker shades indicate higher connectivity to influential peers.
- Node shape: Betweenness centrality, square shapes indicate high bridging roles above median; circles indicate below-median bridging roles.

Table 4. Normalized ego network metrics

Node	Level	Course	Indegree	Outdegree	Closeness	Betweenness	Eigenvector	Clustering coefficient
65	High	C4	0.5	0.31	0.0095	26.72	1	0.41
188	Low	C6	0.02	0.34	0.0103	11.67	0.31	0.28
39	Low	C5	0	0.18	0.0094	0	0.09	0.19

Readability Analysis

For the readability analysis, all relevant discussion posts were selected from these three students who reflected a range of network positions and engagement levels and four standard readability indices were computed using the textstat in Python. Each score was interpreted using established benchmarks to assess accessibility. Readability results(Table 5) were then compared with the students' network metrics, allowing for analysis of how textual clarity and network position may contribute to patterns of peer engagement in discussion. Each index evaluates readability by analyzing sentence length, word complexity, or syllable count, enabling a more comprehensive assessment of how accessible and engaging student posts may be for their peers.

Table 5. Readability scores of three cases

Student	Flesch Reading Ease	Flesch-Kincaid Grade Level	Gunning Fog Index	SMOG Index	Readability Interpretation
39	22.5	15.9	17.9	16.4	Highly academic, possibly too dense
65	54.4	11.9	14.4	13.0	Accessible, appropriate for college readers
188	46.6	11.9	14.2	13.6	Balanced, slightly formal but college-level readability

Student 65 (High engagement, Accessible text): Student 65 represents a case of high engagement, acting as a central figure in the discussion. Structurally, this is reflected in their high indegree, eigenvector, and betweenness centrality scores. Visually, their node is larger, darker, and square-shaped, indicating they received many responses, were connected to other influential peers, and served as a vital bridge connecting different parts of the discussion network. Beyond this advantageous network position, the textual quality of their post likely contributed to this high level of interaction. A readability analysis revealed their post was accessible and appropriate for college-level peers (Flesch Reading Ease = 54.4; Kincaid Grade Level = 11.9) (Table 5). This combination of a structurally central position and clear and accessible posts created a compelling contribution that was both highly visible and easy for peers to engage with, fostering their role as a discourse leader.

Student 39 (No engagement, Complex text): In contrast, Student 39 represents a case of non-engagement, having received zero replies (indegree of 0). Their peripheral network position is evident from their very low eigenvector and zero betweenness centrality, visualized as a small, circular node in the network. The reason for this complete lack of engagement becomes clearer when considering the post content. The readability metrics showed that the posts were far from being simplistic and had graduate level complexity, rated with an extremely low Flesch Reading Ease score of 22.5 and a Flesch-Kincaid Grade Level of 15.9 (Table 5). In this instance, the student's isolated network position was likely deepened by a significant communication barrier; the posts' high complexity may have intimidated peers or made it too difficult to understand and respond to easily.

Student 188 (Low engagement, Moderate readability): Student 188, who received minimal responses, demonstrates a more complex scenario. Structurally, their indegree and other centrality scores were low. However, they showed more outreach than Student 39 with a moderate outdegree, and their ego network included a notable tie to a highly active peer (node 167). The readability analysis characterized the posts as moderately complex yet college appropriate (Flesch Reading Ease = 46.6; Grade Level = 11.9). While the posts were textually appropriate for the academic level, it was slightly denser than the highly successful post from Student 65. This case suggests that for students who are not already in central network positions, even minor increases in textual formality or complexity, combined with a less central position, can result in limited peer interaction.

These illustrative cases reveal a complex interaction between students' structural network positions and their posts' textual readability. They reveal how Student 65 become central figures, facilitating dialogue not just through their position but through accessible communication, while others may remain on the periphery struggling to influence or benefit from the network's collaborative potential due to a combination of network isolation and communication styles that are not well-pitched for peer discussion. Collectively, these cases emphasize the importance of integrating clear textual communication with strateigies promoting structural centrality to maximize student engagement within asynchronous discussions. These insights help understanding the multifaceted nature of engagement in online learning environments and inform educational practices aiming to support inclusive and effective participation.

DISCUSSION

This study used social network analysis (SNA), statistical analyses, and textual readability assessments to examine student engagement in asynchronous online discussions. All network metrics and ego network visualizations in this study were calculated within the complete classroom network, ensuring that each student's engagement and network position reflected their standing among all peers, not in isolation. The high and low response receiver groups were defined for analytical comparison, but each individual's network patterns and illustrative ego networks were interpreted within the broader social context of the entire group. This maintains the network's structural validity and enables meaningful comparison between different engagement types.

Network structure and student engagement

At a network-wide level, the ANOVA revealed that students who received many responses (high indegree centrality) significantly differed from low-response receivers on critical network metrics—particularly indegree, eigenvector centrality, and betweenness centrality. These findings reinforce and extend existing literature that highlights the importance of structural centrality in predicting online interaction (Saqr et al., 2022; Lee & Sharma, 2025). Importantly, the finding that outdegree centrality (number of posts sent) and clustering coefficients (dense local networks) did not significantly enhance broader engagement underscores that quantity of participation alone does not guarantee meaningful interaction (Wise & Cui, 2018). This emphasizes the strategic quality and diversity of connections rather than mere posting frequency or local connectivity.

Complementing the ANOVA, the multiple regression analysis further elucidated how specific centrality measures jointly contributed to student visibility (indegree). Notably, eigenvector centrality, indicating influential peer connections, strongly predicted receiving responses, followed by betweenness centrality, highlighting bridging roles. Conversely, higher outdegree showed a negative association with indegree, suggesting that overly frequent posting without strategic relational positioning may actually dilute one's impact. This nuanced statistical analysis advances our understanding of network metrics in educational contexts, illustrating that strategic relational positioning, connecting to influential peers and bridging disparate groups, is crucial for fostering engagement.

Ego network and readability

Beyond network-level patterns, our ego network analysis illuminated individual differences in structural positioning and communicative accessibility. The selection of three illustrative ego networks (Students 65, 39, and 188) highlighted distinct scenarios reflecting high, low, and no engagement levels, respectively. This fine-grained qualitative analysis enabled the exploration of network structure at the individual level, aligning clearly with our quantitative findings.

The central actor: a synthesis of network position and communicative style: One of the important findings is that the most engaged student (Student 65) was not only structurally central but was also a highly effective communicator. This aligns well with the work of Ouyang and Chang (2018), who identified that socially active students in roles like "leaders" and "starters" contributed more significantly to the cognitive aspects of online discussions. Our study adds a potential nuance to their finding, suggesting that the ability to become a "leader" may depend not just on initiating ideas, but on presenting them with a level of clarity that invites participation. The clear and college-appropriate style of Student 65's post likely facilitated the peer interaction that solidified their central role. This supports the idea that central figures are pivotal for enhancing dialogue and integrating less active participants into a vibrant and inclusive educational environment (Lee & Sharma, 2025; Saqr & Lopez-Pernas; 2022).

Communication barriers and the limits of in-depth inquiry. The most striking finding is the case of Student 39, whose academic but textually dense posts received zero replies. This provides a critical nuance to the findings of Ouyang and Chang (2018), who noted that "in-depth individual inquiry often sparked interaction from peers." Our results suggest interaction is conditional. Student 39's inquiry, though conceptually deep, failed to engage peers, likely because its high textual complexity acted as a communication barrier. This implies that for knowledge inquiry to foster the collective knowledge building described by Ouyang and Chang , it must first be communicatively accessible to its intended peer audience. This finding specifies excessive textual density as a key, and currently underexplored, barrier in peer-to-peer forums, reinforcing the need for instructors to be attentive to issues of message clarity that can hinder student interaction. The study offers a deep level explanation for the gap between social and conceptual engagement identified by Chen et al. (2018) and Ouyang and Chang (2018). The textual accessibility of a post may be one of the key mechanisms affecting this relationship. When posts are textually inaccessible (like student 39's), the conceptual and social dimensions become decoupled; the idea exists, but it fails to create a social tie. This reinforces the conclusion from both studies that fostering social interaction does not automatically lead to high-level cognitive engagement and points to the need for intentional instructional design.

The in-depth analysis of three illustrative ego networks (Students 65, 39, and 188) provides a fine-grained perspective on engagement types that complements the quantitative results. This approach follows established qualitative SNA methods, where "thick descriptions" of individual cases are used to illuminate patterns and mechanisms that group metrics may obscure (Hollstein, 2011, p. 413). As Hollstein emphasizes, such qualitative analysis "is the objective and, as such, an 'end in itself'," offering detailed understanding rather than statistical generalizability. In our study, the chosen cases exemplify high, low, and no engagement scenarios, demonstrating how structural position and communicative accessibility interact within the broader classroom network. This case-based interpretation is consistent with prior SNA research and is especially useful for uncovering relational mechanisms that affect student participation and knowledge construction (Nimmon & Atherley, 2021; Macfadyen & Dawson, 2010).

IMPLICATIONS

This study offers several implications for both online instructional practice and the design of learning analytics tools. For instructional practice, the findings suggest that instructors may need to move beyond basic participation requirements and instead foster both strategic relational engagement and effective communication practices. Encouraging students to write clear, accessible posts can enhance peer interaction, particularly when combined with efforts to build diverse and meaningful network connections. Assigning roles that promote bridging between groups or supporting less-connected students to engage more broadly may help mitigate the isolating effects of clustered interactions and peripheral network positions. Additionally, instructors might provide guidance on writing for peer audiences, helping students recognize how textual clarity influences the likelihood of receiving responses.

The results also point to potential enhancements for learning platforms. Incorporating real-time network analytics could help educators identify students who may be structurally or communicatively isolated and support timely interventions. Feedback tools that highlight both network positions and textual readability could help students reflect on and adjust their participation strategies. Such tools may also guide educators in targeting scaffolding efforts toward students whose contributions may be less likely to elicit peer interaction.

Methodologically, this study demonstrates the value of integrating social network metrics with textual analysis to capture a more complete picture of engagement. The combined approach offers a richer understanding than either type of analysis alone. Future research building on this integrated approach—especially through longitudinal and qualitative extensions—could further clarify how students' social positions and communication strategies co-evolve over time in online learning environments.

CONCLUSION

This study used social network analysis, statistical analyses, and textual readability assessments to examine student engagement in asynchronous online discussions. The quantitative analyses—including ANOVA and multiple regression—revealed meaningful relationships between network metrics and student engagement. Specifically, students occupying central positions, characterized by high indegree, eigenvector, and betweenness centralities, were more likely to receive responses. In contrast, frequent posting alone (high outdegree) or interactions within highly clustered groups did not significantly enhance engagement. These findings align with prior literature emphasizing the importance of strategic positioning within interaction networks rather than merely high posting frequency. Moreover, the integration of readability analysis provided deeper insights, highlighting that accessible and clearly written posts were positively aligned with network advantages, while complex or dense writing further isolated students already at network peripheries. This demonstrates an important interaction between textual clarity and structural network positioning, consistent with established principles of critical inquiry and effective dialogue in online learning environments (Garrison et al., 1999).

This study makes several contributions to our understanding of online learning engagement. Its primary contribution lies in illustrating the interaction between structural network position and communicative clarity. By combining network analytics and readability assessment, this research reveals how these dimensions influence student interactions together, thus advancing beyond separate analyses of network structures or textual content alone. Theoretically, the findings highlight how overly complex academic language can unintentionally act as a communication barrier, potentially explaining discrepancies between social and conceptual engagement previously identified in the literature. Practically, the results encourage educators to emphasize communication clarity and diverse connections rather than simply requiring frequent participation. Additionally, online platforms could utilize network analytics to proactively identify and support students needing improvements in textual clarity or relational connections.

This study has limitations. First, the analysis provided a static, cross-sectional view of interactions and excluded extreme interaction patterns. This approach limits insights into how network roles might evolve over time. Future research should adopt longitudinal designs to explore these dynamics more comprehensively. Also, engagement was primarily measured by indegree, accompanied by surface-level readability analysis. This methodology did not capture covert engagement or the deeper semantic and argumentative qualities of student posts. Hence, subsequent research could employ mixed-method approaches—including interviews, qualitative discourse analysis, or clickstream data—to offer richer insights into these less visible aspects of student participation.

Ultimately, by clarifying the synergy between social network positions and textual communication, this study contributes to developing more integrated strategies for supporting meaningful student engagement in online learning communities.

REFERENCES

Bakharia, A., & Dawson, S. (2011, February). SNAPP: a bird's-eye view of temporal participant interaction. In Proceedings of the 1st international conference on learning analytics and knowledge (pp. 168-173). https://doi.org/10.1145/2090116.2090144

Baldwin, T. T., Bedell, M. D., & Johnson, J. L. (1997). The social fabric of a team-based M.B.A. program: Network effects on student satisfaction and performance. Academy of Management Journal, 40(6), 1369–1397. https://doi.org/10.2307/257037

Boling, E. C., Hough, M., Krinsky, H., Saleem, H., & Stevens, M. (2012). Cutting the distance in distance education: Perspectives on what promotes positive, online learning experiences. The Internet and Higher Education, 15(2), 118-126. https://doi.org/10.1016/j.iheduc.2011.11.006

Borgatti, S. P., Jones, C., & Everett, M. G. (1998). Network measures of social capital. Connections, 21(2), 27-36.

Borgatti, S. P., Agneessens, F., Johnson, J. C., & Everett, M. G. (2024). Analyzing social networks (2nd ed.).

Borup, J., West, R. E., & Graham, C. R. (2012). Improving online social presence through asynchronous video. The Internet and Higher Education, 15(3), 195-203. https://doi.org/10.1016/j.iheduc.2011.11.001

Bryman, A. (2016). *Social research methods* . Oxford university press.

Carvalho, L., Goodyear, P., & de Laat, M. (Eds.). (2016). Place-based spaces for networked learning. Routledge. https://doi.org/10.4324/9781315694676

Chen, B., Chang, Y.-H., Ouyang, F., & Zhou, W. (2018). Fostering student engagement in online discussion through social learning analytics. The Internet and Higher Education, 37, 21–30. https://doi.org/10.1016/j.iheduc.2017.12.002

Crossley, S. A., Allen, D. B., & McNamara, D. S. (2011). Text readability and intuitive simplification: A comparison of readability formulas. Reading in a foreign language, 23(1), 84-101.

Dawson, S. (2010). 'Seeing' the learning community: An exploration of the development of a resource for monitoring online student networking. British Journal of Educational Technology, 41(5), 736-752. https://doi.org/10.1111/j.1467-8535.2009.00970.x

De Laat, M. (2012). Enabling professional development networks: How connected are you?. Open Universteit.

De Laat, M., & Prinsen, F. (2014). Social Learning Analytics: Navigating the Changing Settings of Higher Education. Journal of Research & Practice in Assessment, 9, 51-60. http://www.rpajournal.com/social-learning- analytics-navigating-the-changing-settings-of-higher-education/

Dennen, V. P. (2008). Looking for evidence of learning: Assessment and analysis methods for online discourse. Computers in Human Behavior, 24(2), 205-219. https://doi.org/10.1016/j.chb.2007.01.010

Dixson, M. D. (2015). Measuring student engagement in the online course: The Online Student Engagement scale (OSE). Online Learning, 19(4), 143–157. https://doi.org/10.24059/olj.v19i4.561

Dringus, L. P., & Ellis, T. (2010). Temporal transitions in participation flow in an asynchronous discussion forum. Computers & Education, 54(2), 340-349. https://doi.org/10.1016/j.compedu.2009.08.011

Edwards, G. (2010). Mixed-method approaches to social network analysis.

Galikyan, I., & Admiraal, W. (2019). Students' engagement in asynchronous online discussion: The relationship between cognitive presence, learner prominence, and academic performance. The Internet and Higher Education, 43, 100692.

Gao, F., Zhang, T., and Franklin, T. 2013. Designing asynchronous online discussion environments: Recent progress and possible future directions. British Journal of Educational Technology. 44, 3, 469–483. https://doi.org/10.1111/j.1467-8535.2012.01330.x

Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in higher education. The Internet and Higher Education, 2(2-3), 87-105. https://doi.org/10.1016/S1096-7516(00)00016-6

Gros, B., & García-Peñalvo, F. J. (2023). Future trends in the design strategies and technological affordances of e-learning. In Learning, design, and technology: An international compendium of theory, research, practice, and policy (pp. 345-367). Cham: Springer International Publishing.

Gunning, R. (1952). The technique of clear writing.

Haythornthwaite, C. (2000). Online personal networks: Size, composition and media use among distance learners. New Media & Society, 2(2), 195-226. https://doi.org/10.1177/14614440022225751

Haythornthwaite, C. (2005). Social networks and Internet connectivity effects. Information, Communication & Society, 8(2), 125-147. https://doi.org/10.1080/13691180500146185

Haythornthwaite, C., & De Laat, M. (2010, May). Social networks and learning networks: Using social network perspectives to understand social learning. In Proceedings of the 7th international conference on networked learning (pp. 183-190). Aalborg, Denmark: Lancaster University.

Hew, K. F., & Cheung, W. S. (2011). Higher-level knowledge construction in asynchronous online discussions: An analysis of group size, duration of online discussion, and student facilitation techniques. Instructional Science, 39, 303-319. https://doi.org/10.1007/s11251-010-9129-2

Hollstein, B. (2011). Qualitative approaches. The SAGE handbook of social network analysis, 404-416. https://doi.org/10.4135/9781446294413.n27

Hrastinski, S. (2008). Asynchronous and synchronous e-learning. Educause Quarterly, 31(4), 51-55.

Kaliisa, R., Rienties, B., Mørch, A. I., & Kluge, A. (2022). Social learning analytics in computer-supported collaborative learning environments: A systematic review of empirical studies. Computers and Education Open, 3, 100073. https://doi.org/10.1016/j.caeo.2022.100073

Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel.

Koh, J. H. L., Herring, S. C., & Hew, K. F. (2010). Project-based learning and student knowledge construction during asynchronous online discussion. The Internet and Higher Education, 13(4), 284-291. https://doi.org/10.1016/j.iheduc.2010.09.003

Kovanović, V., Gašević, D., Joksimović, S., Hatala, M., & Adesope, O. (2015). Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions. The Internet and Higher Education, 27, 74-89. https://doi.org/10.1016/j.iheduc.2015.06.002

Lee, M.K., Sharma, P. (2025). Applying a community detection algorithm to examine group formation in online discussions: Exploring network characteristics and dynamics, International Journal of Educational Technology in Higher Education. https://doi.org/10.1186/s41239-024-00495-w

Marbouti, F., & Wise, A. F. (2016). Starburst: A new graphical interface to support purposeful attention to others' posts in online discussions. Educational Technology Research and Development, 64(1), 89-111.

Macfadyen, L.P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education, 54*(2), 588–599. https://doi.org/10.1016/j.compedu.2009.09.008

Mc Laughlin, G. H. (1969). SMOG grading-a new readability formula. Journal of reading, 12(8), 639-646.

Ouyang, F., & Chang, Y. H. (2019). The relationships between social participatory roles and cognitive engagement levels in online discussions. British Journal of Educational Technology, 50(3), 1396-1414. **https://doi.org/10.1111/bjet.12647**

Palonen, T., & Hakkarainen, K. (2013, April). Patterns of interaction in computer-supported learning: A social network analysis. In International Conference of the Learning Sciences (pp. 334-339). Psychology Press.

Romero, C., López, M. I., Luna, J. M., & Ventura, S. (2013). Predicting students' final performance from participation in on-line discussion forums. Computers & Education, 68, 458-472.

Rosé, C. P., & Ferschke, O. (2016). Technology support for discussion based learning: From computer supported collaborative learning to the future of massive open online courses. International Journal of Artificial Intelligence in Education, 26, 660-678. https://doi.org/10.1007/s40593-016-0107-y

Russo, T. C., & Koesten, J. (2005). Prestige, centrality, and learning: A social network analysis of an online class. Communication Education, 54(3), 254-261.

Saqr, M., & Alamro, A. (2019). The role of social network analysis as a learning analytics tool in online problem-based learning. BMC Medical Education, 19, 160. https://doi.org/10.1186/s12909-019-1599-6

Saqr, M., Fors, U., & Tedre, M. (2020a). How learning analytics can early predict under-achieving students in a blended medical education course. BMC Medical Education, 20(1), 1-11. https://doi.org/10.1186/s12909-020-1972-9

Saqr, M., Nouri, J., Vartiainen, H., & Tedre, M. (2020b). 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

Saqr, M., & López-Pernas, S. (2021). The longitudinal trajectories of online engagement over a full program. Computers & Education, 175, Article 104325. https://doi.org/10.1016/j.compedu.2021.104325

Saqr, M., & López-Pernas, S. (2022). How CSCL roles emerge, persist, transition, and evolve over time: A four-year longitudinal study. Computers & education, 189, 104581.

Saqr, M., Elmoazen, R., Tedre, M., López-Pernas, S., & Hirsto, L. (2022). How well centrality measures capture student achievement in computer-supported collaborative learning?–A systematic review and meta-analysis. Educational Research Review, 35, 100437.

Shum, S. B., & Ferguson, R. (2012). Social learning analytics. Journal of educational technology & society, 15(3), 3-26.

Siemens, G. (2013). Learning analytics: The emergence of a discipline. American Behavioral Scientist, 57(10), 1380-1400. https://doi.org/10.1177/0002764213498851

Simpson, O. (2013). Supporting students for success in online and distance education. Routledge.

Tabachnick, B. G. (2007). Using multivariate statistics. *Alyn and Bacon*.

Tinto, V. (1998). Colleges as communities: Taking research on student persistence seriously. The Review of Higher Education, 21(2), 167–177. https://doi.org/10.1353/rhe.1998.0024

Vlachopoulos, P. (2012, November). The importance of power dynamics in the development of asynchronous online learning communities. In Proceedings of ASCILITE-Australian Society for Computers in Learning in Tertiary Education Annual Conference.

Wise, A. F., Perera, N., Hsiao, Y.-T., Speer, J., & Marbouti, F. (2012). Microanalytic case studies of individual participation patterns in an asynchronous online discussion in an undergraduate blended course. The Internet and Higher Education, 15(2), 108-117. https://doi.org/10.1016/j.iheduc.2011.11.007

Wise, A. F., Zhao, Y., & Hausknecht, S. (2014). Learning analytics for online discussions: Embedded and https://doi.org/10.18608/jla.2014.12.4

Wise, A. F., & Cui, Y. (2018). Unpacking the relationship between discussion forum participation and learning in MOOCs: Content is key. ACM International Conference Proceeding Series, 330–339. https://doi.org/10.1145/3170358.3170403

Yang, C. C., Flanagan, B., Akçapınar, G., and Ogata, H. 2019. Investigating subpopulation of students in digital textbook reading logs by clustering. In Companion Proceedings of the 9th International Conference on Learning Analytics and Knowledge (LAK'19). Society for Learning Analytics Research (SoLAR), 465-470. https://doi.org/10.13140/RG.2.2.17462.60483

Xie, K., & Ke, F. (2011). The role of students' motivation in peer-moderated asynchronous online discussions. British Journal of Educational Technology, 42(6), 916-930. **https://doi.org/10.1111/j.1467-8535.2010.01140.x**