JOURNAL OF LEARNING ANALYTICS

Volume 9(1), 13–31. https://doi.org/10.18608/jla.2022.7415

The Curious Case of Centrality Measures: A Large-Scale Empirical Investigation

Mohammed Saqr[1], Sonsoles López-Pernas[2]

Abstract

There has been extensive research using centrality measures in educational settings. One of the most common lines of such research has tested network centrality measures as indicators of success. The increasing interest in centrality measures has been kindled by the proliferation of learning analytics. Previous works have been dominated by single-course case studies that have yielded inconclusive results regarding the consistency and suitability of centrality measures as indicators of academic achievement. Therefore, large-scale studies are needed to overcome the multiple limitations of existing research (limited datasets, selective and reporting bias, as well as limited statistical power). This study aims to empirically test and verify the role of centrality measures as indicators of success in collaborative learning. For this purpose, we attempted to reproduce the most commonly used centrality measures in the literature in all the courses of an institution over five years of education. The study included a large dataset (n=3,277) consisting of 69 course offerings, with similar pedagogical underpinnings, using meta-analysis as a method to pool the results of different courses. Our results show that degree and eigenvector centrality measures can be a consistent indicator of performance in collaborative settings. Betweenness and closeness centralities yielded uncertain predictive intervals and were less likely to replicate. Our results have shown moderate levels of heterogeneity, indicating some diversity of the results comparable to single laboratory replication studies.

Notes for Practice

- Degree and Eigenvector centrality measures can be a consistent indicator of performance in settings where course design emphasizes collaboration.
- The correlation between degree and eigenvector centrality measures and academic achievement was reproducible regardless of the number of students, number of interactions, year of study, or course subject.
- Closeness and betweenness centralities showed inconsistent correlation with performance.
- Although our context was homogenous, there was moderate heterogeneity in the pooled effect sizes indicating the diversity of CSCL as a medium.

Keywords

Centrality measures, learning analytics, meta-analysis, reproducibility, replicability, social network analysis

Submitted: 01/02/21 — **Accepted:** 20/10/21 — **Published:** 11/03/22

Corresponding author [1]Email: mohammed.saqr@uef.fi Address: School of Computing, University of Eastern Finland, Joensuu Campus, Yliopistokatu 2, P.O. Box 111, fi-80100, Joensuu, Finland. ORCID ID: https://orcid.org/0000-0001-5881-3109
[2]Email: sonsoles.lopez.pernas@upm.es Address: Departamento de Sistemas Informáticos, ETSI Sistemas Informáticos, Universidad Politécnica de Madrid, c/ Alan Turing, s/n, 28031 Madrid, Spain. ORCID ID: https://orcid.org/0000-0002-9621-1392

1. Introduction

There has been extensive debate about the suitability of computational methods (e.g., social network analysis) to understand, optimize, and support collaborative learning (Wise & Schwarz, 2017). This ongoing debate has been fuelled by the lack of norms or consistent practices that embolden our understanding of collaborative learning. The absence of norms may have been a consequence of reliance on small case studies. As Wise and Schwarz (2017) posit, "even when appropriate analytic methods are applied, it is questionable whether the accumulation of case studies leads to clear progress in our field." Furthermore, Andres et al. (2015) cautioned against the lack of replication, which may result in perceiving exploratory studies as facts, which could have "dangerous" effects depending on the findings and the affected populations. Thus, to foster our confidence in the analytics methods, large-scale empirical or replication studies are needed. Such studies could offer robust empirical evidence beyond selective reporting and bias as well as overcome the limitations of evidence offered by small case studies.

A fundamental principle of science is the possibility to test, verify, or refute the claims or conclusions reported by other researchers. Therefore, reproducibility (obtaining similar results when repeating an experiment with a different team and a different experimental setup) and replicability (obtaining similar results when repeating an experiment with a different team and the same experimental setup) have become key drivers of scientific progress (Plesser, 2018).1 Reproducible research is more amenable to generalization, emboldens the credibility of research findings, and can translate into real-life impact (Aarts et al., 2015; Baker & Penny, 2016; Dawson et al., 2019; Hagger et al., 2016). The scientific community has taken serious steps to encourage reproducibility/replicability in research after what has been dubbed "the reproducibility crisis" (Baker & Penny, 2016; Hagger et al., 2016; Klein, 2019). A Nature survey has shown that 70% of scientists have failed to reproduce the results reported by other scientists, and more than half failed to reproduce the results of their own studies (Baker & Penny, 2016). Although replication studies are increasingly common, they remain relatively rare in education research. Makel and Plucker (2014) estimated that only 0.13% of educational research articles were replications of other studies, concluding that "despite increased attention to methodological rigor in education research, the field has focused heavily on experimental design and not on the merit of replicating important results." Replication is gaining ground in the learning analytics community. Several studies have emerged to replicate/reproduce findings of other studies (Andres et al., 2015, 2018; Li et al., 2017). Recently, Dawson et al. (2019) called for supporting a culture of replication and supporting replication studies.

Another threat to the credibility and integrity of knowledge is publication bias (Alexander et al., 2015; Joober et al., 2012). Publication bias occurs when positive results are more likely to be reported or published than negative results. This problem has a multifaceted etiology: publications with positive findings are more likely to be evaluated favourably by reviewers, considered for publication by editors, and later cited. Such an environment makes researchers more likely to submit positive findings knowing that they may receive preferential treatment (Dwan et al., 2013; Joober et al., 2012). Researchers may also choose a "convenience sample" with an interesting practice or findings to share. Therefore, publication bias leads to the amplification of the magnitude of findings or inflates the value of ineffective practices (Joober et al., 2012). The widespread recognition of publication bias has led to an increasing interest in replication as a possible way to validate the veracity of our knowledge or lack thereof (Joober et al., 2012).

There is a wealth of methods for the study of collaborative interactions at the hands of the researcher. Hoppe (2017) refers to the main approaches as "the trinity of methods," which include: 1) actor-to-actor networks, 2) actor-to-artefact networks, and 3) content analysis. The analysis of actor-to-actor networks using Social Network Analysis (SNA) measures (a mathematical quantification of students' relations and connectedness) has increasingly received attention from scholars (Borgatti, 2005; Borgatti & Brass, 2019; Freeman, 1978), especially after the rise in adoption of learning analytics (Cela et al., 2015; Saqr, Viberg et al., 2020). Common applications include prediction of student success, monitoring student interactions, classifying student roles (e.g., leaders, collaborators, inactive, and isolated), as well as quantifying peer influence (Cela et al., 2015; Dado & Bodemer, 2017). Results from different studies have been variable, i.e., studies have reported dissimilar results regarding the correlation of SNA measures with academic achievement (Cela et al., 2015; Saqr, Viberg et al., 2020). The variability extends to the type of centrality: whereas some studies have highlighted the value of, e.g., the closeness centrality as a predictor of productive interactions (Cho et al., 2007), others have highlighted the predictive value of, e.g., in-degree centrality (Hernández-García et al., 2015). Researchers have pointed out the role of instructional conditions, course design, and variability in methods as possible causes of variability (Dawson et al., 2019).

We argue that a large-scale empirical study that attempts to reproduce the literature findings of centrality measures as indicators of success has the potential to embolden our confidence in their value as well as to improve our understanding of which centralities to choose as predictors of academic achievement. In this study, we attempt to reproduce the previously reported findings on centrality measures as indicators of success. To control for course/context variability, we use a homogenous dataset of courses with a similar pedagogical underpinning (problem-based learning), from the same institution, as well as with a similar course design. To avoid bias, we report our findings for *all courses* that have a collaborative module. In doing so, we overcome possible selection (reporting on some interesting courses) or reporting bias (reporting only on positive findings). We report our results using traditional statistical methods as well as correlation meta-analysis. Meta-analysis enables the possibility of assessing the consistency of centrality measures as indicators of student success across all courses, the variability of results among courses using *heterogeneity* statistics, as well as the future predictability (*predictive interval)* or to what extent we can be confident that centrality measures will perform in the future as predictors of success. Our aim is to answer the following research questions:

RQ 1 Can we rely on centrality measures as indicators of success?

RQ 2 Which centrality measures are the most consistent indicators of success?

RQ 3 How certain can we be in future applications of centrality measures as indicators of success?

[1] The terms reproducibility and replicability are often used interchangeably with no consensus on the accurate definition. For example, the *Nature* article cited above uses reproducibility for reproducing self and others' results, which contradicts the ACM definition cited above.

ISSN 1929-7750 (online).

2. Background

The background section begins with an introduction about collaborative learning, SNA, and centrality measures, and how they have been used in education. We then describe each centrality measure and review how they were operationalized in prior works. We later review the results of computer-supported collaborative learning (CSCL) research investigating centrality measures and their correlation with performance.

2.1. Collaborative Learning

Collaborative learning has occupied a central position in pedagogy for more than four decades. The growing adoption has been kindled by a large corpus of empirical evidence confirming the worth of collaborative learning in, e.g., knowledge and skill acquisition, task performance, and student motivation (Bernard et al., 2009; Borokhovski et al., 2016; Chen et al., 2018; Wecker & Fischer, 2014). Collaborative learning is supported by several learning theories. In particular, the theory of social constructivism. Social constructivists view learning as a process of active construction of meaning that occurs through social interaction and dialogue with peers (Liu & Matthews, 2005; Martin & Dowson, 2009). Consequently, modern pedagogies have embraced collaborative learning strategies that encourage interactions among learners and promote social and co-operation skills. However, collaborative learning is challenging to implement and maintain successfully, thus requiring monitoring and support as well as regulation of cognitive and social processes (Jeong & Hmelo-Silver, 2016; Weinberger & Fischer, 2006). Technological and computational methods have been proposed as possible tools that could help teachers monitor and support collaboration (Dado & Bodemer, 2017). One such tool is SNA, which has been employed to visualize the network of collaborators, map collaborators' roles, identify gaps, and devise a possible intervention (Dado & Bodemer, 2017; Saqr, Fors, Tedre et al., 2018). What is more, SNA centrality measures have been frequently used as indicators of productive interactions, social positioning, and participation, as well as predictors of success (Cadima et al., 2012; de-Marcos et al., 2016; Joksimović et al., 2016; Osatuyi & Passerini, 2016; Putnik et al., 2016; Reychav et al., 2018; Saqr, Fors, & Tedre, 2018; Wise & Cui, 2018). We review these applications in detail in the next section, as they are the focus of our study.

2.2. Social Network Analysis

Interaction or communication among learners and/or teachers can be represented as a social network in which the actors are commonly referred to as nodes and the relationships or interactions among them are referred to as ties or edges (Borgatti et al., 2009; Shafie, 2019). SNA offers methods for the study of social networks in the form of visualization and mathematical analysis (Borgatti et al., 2009). SNA affords researchers a plethora of visualization methods that have a summarizing power to plot a whole class of students. Visualization has been used frequently in educational research to map the patterns of interactions, evaluate the interactivity of a group of learners, and identify active and inactive students. Other uses include raising awareness of learners, promoting collaboration, and supporting intervention (Dado & Bodemer, 2017). Quantitative SNA methods provide a quantification of node connectedness, interactions, position, or importance in the network, often referred to as centrality measures (Borgatti et al., 2009; Gašević et al., 2013; Joksimović et al., 2016; Saqr, Fors, & Nouri, 2018). Using centrality measures as indicators of students' online learning dates back over three decades. Early examples include work by Baldwin et al. (1997), a study of the social structure of different types of networks (advice, communication, and adversarial) in an MBA program. In this study, the authors concluded that relationships among teams had significant effects on team members' perceptions of team performance and effectiveness. Several other studies followed using centrality measures to predict performance (Cela et al., 2015; Dado & Bodemer, 2017). In the next section, we offer a detailed review of each commonly used centrality measure and how prior research has operationalized it in CSCL, as well as a review of the literature in which centrality measures were used as indicators of success.

2.3. Centrality Measures

While the literature cites more than two hundred centrality measures, very few centralities are used in education. Our literature review found only six centrality measures that have been used more than four times in CSCL as indicators of student success. These centralities lie within three categories: degree centralities, short-path based centralities, and eigenvector-based centralities.

2.4. Degree Centralities

Degree centrality refers to the total sum of posts, messages, or interactions sent and received by an actor (Cadima et al., 2012; Cho et al., 2007; de-Marcos et al., 2016; Joksimović et al., 2016; Liu et al., 2019; Liu, Kang, Su et al., 2018; Liu, Kang, Domanska et al., 2018; Osatuyi & Passerini, 2016; Putnik et al., 2016; Reychav et al., 2018; Saqr, Fors, & Nouri, 2018; Saqr, Fors, & Tedre, 2018; Saqr, Viberg et al., 2020; Wise & Cui, 2018). The interpretation of degree centrality varies by context and task: in collaborative learning, student posts are expected to establish social presence, engage in knowledge co-construction, or share resources. As such, degree centrality is operationalized as the degree of contribution to discourse: degree of influence (Hernández-García et al., 2015; Reychav et al., 2018), prominence and importance in knowledge construction

(Cadima et al., 2012; Liu, Kang, Domanska et al., 2018), popularity and communicative activity (Joksimović et al., 2016), social activity (Cadima et al., 2012), and embeddedness in the network (Cho et al., 2007). In directed networks, where the edges have an associated direction, degree centrality has two variants: out-degree and in-degree centralities. Out-degree centrality represents the number of outgoing posts, messages, or interactions contributed by an actor. Out-degree centrality is commonly operationalized in a similar way to degree centrality: as quantification of participation, social positioning, and communicative activity (Jo et al., 2017; Liu, Kang, Su et al., 2018; Reychav et al., 2018; Romero et al., 2013; Saqr, Fors, & Nouri, 2018; Saqr, Fors, & Tedre, 2018; Saqr & Alamro, 2019). In-degree centrality reflects the importance of the user, the worthiness of his/her contributions to receive a reply, the popularity, and authority (Jo et al., 2017; Liu, Kang, Su et al., 2018; Reychav et al., 2018; Romero et al., 2013; Saqr, Fors, & Nouri, 2018; Saqr, Fors, & Tedre, 2018; Saqr & Alamro, 2019).

2.5. Short-Path Centralities

Closeness centrality measures how "close" an actor is to all others in the network and, consequently, how quickly they can reach others (Cadima et al., 2012; Cho et al., 2007; Liu, Kang, Domanska et al., 2018; Osatuyi & Passerini, 2016; Putnik et al., 2016; Saqr, Fors, & Tedre, 2018). Closeness centrality is calculated as the inverse farness (distance to all other nodes). It is commonly operationalized as proximity, reachability, awareness of opportunities, access to information, diversity of resources and independence (the more closeness, the less reliance on limited sources), and control of information exchange (Cadima et al., 2012; de-Marcos et al., 2016; Gašević et al., 2019; Joksimović et al., 2016; Liu, Kang, Su et al., 2018; Osatuyi & Passerini, 2016; Reychav et al., 2018; Saqr, Fors, & Nouri, 2018; Saqr, Fors, & Tedre, 2018; Saqr, Viberg et al., 2020; Saqr & Alamro, 2019; Wise & Cui, 2018).

Betweenness centrality is the frequency a node has connected two other unconnected nodes (i.e., the shortest path between them). Betweenness centrality reflects the frequency of mediation — or control over — information exchange in a network. Betweenness centrality is always operationalized as a bridging capital and access to diverse resources. By connecting diverse communities, an actor gains access to resources from both communities. Betweenness centrality has also been operationalized as access to novel information and opportunities (Cadima et al., 2012; Cho et al., 2007; de-Marcos et al., 2016; Hernández-García et al., 2015; Jo et al., 2017; Joksimović et al., 2016; Liu, Kang, Domanska et al., 2018; Liu, Kang, Su et al., 2018; Osatuyi & Passerini, 2016; Putnik et al., 2016; Reychav et al., 2018; Saqr, Fors, & Tedre, 2018; Wise & Cui, 2018).

2.6. Eigenvector-Based Centralities

Eigenvector centrality tries to surmount the shortcomings of degree centrality, which treats all connections as equal, by calculating actors' centrality based on their neighbours' centralities, i.e., connectedness to important (influential) nodes in the network translates to higher Eigenvector centrality. The principle behind this centrality is that fewer connections to important actors may be more valuable than many connections to isolated actors. Eigenvector centrality has been operationalized as the strength of social capital, connectedness, and social presence (de-Marcos et al., 2016; Hernández-García et al., 2015; Liu, Kang, Su et al., 2018; Putnik et al., 2016; Saqr, Fors, & Nouri, 2018; Saqr, Fors, & Tedre, 2018; Saqr, Viberg et al., 2020; Traxler et al., 2016; Wise & Cui, 2018).

2.7. Predicting Student Success

One of the main threads of research in learning analytics has been focused on predicting student success. Predicting students who may fail or underachieve may pave the way for the provision of appropriate support and proactive intervention (Conijn et al., 2017; Gašević et al., 2016; Ifenthaler & Yau, 2020). Three main themes of this research can be observed: 1) studies performed in limited settings (e.g., a single course); 2) studies performed in multiple courses, and 3) studies replicating other findings of similar studies (Ifenthaler & Yau, 2020; Li et al., 2017).

The first type (single courses), "convenient sample" studies, has prevailed in the early research of learning analytics, giving rise to many exploratory studies aimed at testing a method or an algorithm in a single course or within a limited setting (see Ifenthaler & Yau, 2020 for a recent review). However, the results obtained from such limited settings have a low potential for generalization and reproducibility. Another potential problem is the issue of publication bias (Rienties et al., 2017). Notwithstanding the value of such examples, they are more likely to be "special cases" that are less subject to generalization (Chan et al., 2004; Rienties et al., 2017).

The second type (multiple courses) is becoming increasingly common in learning analytics. Results from large-scale studies have reported noticeable variability in indicators of student success, as well as in the precision or portability of predictive models (Conijn et al., 2017; Gašević et al., 2016). This variability was reported across different institutions as well as within the same institution. The conclusions drawn from these studies indicate that *course-agnostic* models may be far from obtainable. Therefore, research in learning analytics must take into account instructional conditions, course design, and teaching practices (Gašević et al., 2016).

The third type (replication studies) has gained recent attention within the community of learning analytics. In a series of studies, Andres et al. (2015, 2018) investigated the degree to which previously published findings on MOOC completion could be replicated. They showed evidence of considerable commonality regarding which behaviours are associated with success in MOOCs on a heterogeneous range of topics. Nonetheless, they found that specific behaviours, such as the linguistic features derived from student discussion forum posts, fail to replicate with statistical significance or even replicate in the opposite direction. In turn, Li et al. (2017) investigated whether and how the learning satisfaction experiences of new vs. continuing students were different. Their findings indicated that while most key drivers for learning satisfaction across two consecutive academic years were similar, new learners differed subtly in their learning experiences. The differences between both cohorts indicate that institutions need to continuously monitor and act upon changing learning needs.

2.8. Centrality Measures as Indicators of Success

Several learning theories (e.g., social constructivism) support the belief that successful interactions in a collaborative setting catalyze learning. Thus, the association between interactions in CSCL and improved performance is widely viewed as "prevailing, and largely unchallenged" (Joksimović et al., 2016), or that "agreement exists" (Romero et al., 2013) on the positive value of participation in CSCL to help students learn (Agudo-Peregrina et al., 2014). The value of collaborative learning is further supported by a large corpus of empirical evidence and several large-scale reviews and meta-analyses which established a positive association between collaborative learning and cognitive gain, knowledge, and skill acquisition (Bernard et al., 2009; Borokhovski et al., 2016; Chen et al., 2018).

To capitalize on the potential of SNA measures as possible indicators of collaborative learning, several studies have explored the potential of centrality measures in translating these interactions into indicators of success. Table 1 offers a detailed review of such studies in CSCL settings. The table reveals several important observations:

1) Most studies have investigated a single course (14 out of 19) or two courses (4 out of the remaining 5 studies).
2) The reviewed studies had no agreement on which centrality measures to use, i.e., each study used a different set of centralities.
3) Three studies with multiple courses have shown mixed results (Jiang et al., 2014; Joksimović et al., 2016; Saqr, Fors, & Nouri, 2018); in other words, the association between centrality measures and performance was positively correlated in some courses and negatively in others.

The results on individual courses show variability: while some have reported a statistical and positive correlation with centrality measures, others have not replicated such findings. For instance, Cho et al. (2007) reported that closeness centrality was the only centrality associated with performance, whereas Cadima et al. (2012) showed opposite findings, i.e., a negative association.

Furthermore, Hernández-García et al. (2015) found a positive and statistically significant correlation with degree, closeness, and betweenness centralities, whereas Saqr and Alamro (2019) found no statistical significance for the exact same centralities. The variations are also prevalent within studies reported by the same researcher in the same setting (Saqr, Fors, & Nouri, 2018; Saqr, Fors, & Tedre, 2018; Saqr & Alamro, 2019). As such, results have been inconclusive on which set of centrality measures are consistent indicators of success in CSCL.

In this study, we use meta-analysis as a method to pool different results of different courses. Meta-analysis represents the gold standard method for synthesizing research results due to the maturity of the methods and the rich repertoire of statistical tools. Meta-analysis affords researchers a rigorous statistical method to pool different effect sizes of different studies while taking into account the different weight and sample size of each study. Additionally, meta-analysis can help estimate the heterogeneity of results (how far the reported results vary) and how far we can be confident that the future application of the same measure can result in similar findings by estimating the predictive interval.

Such methods represent an improvement over the commonly used methods that report data from different courses, e.g., pooling all courses in the same pool or listing results side by side and counting the "votes" of each study, i.e., counting how many studies have reported similar findings.

To prepare Table 1, we searched Scopus, Web of Science, and ERIC databases on July 26, 2019, using the following search query: "social network analysis" AND "learning" AND "centrality." The search yielded 429 articles from the Scopus database, 151 articles from the Web of Science database, and 30 from the ERIC database. The total number of articles was 610, dropping to 474 after removing duplicates. The search was limited to peer-reviewed articles published from 2000 through 2019 and written in English. We further included 241 articles referenced in the previously published systematic reviews of SNA (Cela et al., 2015; Dado & Bodemer, 2017; Sie et al., 2012). The abstracts, titles, and keywords of the 581 articles remaining after removing duplicates were independently reviewed by two researchers. The agreement between the two researchers was 0.88. Articles with disagreement were discussed until satisfactory agreement was reached, which resulted in 246 articles. Then, one of the researchers proceeded to review the full text of the resulting 246 articles and met with the second researcher to discuss and resolve uncertainties. This step resulted in 61 articles reporting on centrality measures in a learning setting, of which only 19 were about CSCL and matched our inclusion and exclusion criteria: 1) empirical research, 2) reporting a quantitative measure of achievement, 3) using data sources that fit the definition of CSCL, 4) reporting enough statistics for calculation of a combined correlation coefficient, and 5) not about special education. As several studies included more than one course, we synthesized the results considering each course as a cohort of students.

Table 1. Sometimes Inconclusive or Contradictory Reports from Different Studies

Authors, date	N	Degree	In-degree	Out-degree	Closeness	Betweenness	Eigenvector
Cho et al., 2007	2	Positive Non-significant			Positive Significant	Positive Non-significant	
Cadima et al., 2012	2	Positive Significant			Negative Significant	Positive Non-significant	
Romero et al., 2013	1		Positive Significant	Positive Significant			
Hernández-García et al., 2015	1	Positive Significant	Positive Significant	Positive Significant	Positive Significant	Positive Significant	Positive Significant
Putnik et al., 2016	1	Positive Significant			Positive Significant	Positive Significant	Positive Significant
de-Marcos et al., 2016	1	Positive Significant			Positive Non-significant	Positive Significant	Positive Significant
Jiang et al., 2014	2	Contradictory			Contradictory	Contradictory	
Joksimović et al., 2016	2	Positive Significant			Contradictory	Contradictory	
Osatuyi & Passerini, 2016	1	Negative Significant			Positive Significant	Positive Significant	
Jo et al., 2017	1		Positive Significant	Positive Significant			
Wise & Cui, 2018	1	Positive Significant			Positive Significant	Positive Non-significant	Positive Significant
Saqr, Fors, & Tedre, 2018	1	Positive Non-significant	Positive Significant	Positive Non-significant	Positive Significant	Positive Significant	Positive Significant
Saqr, Fors, & Nouri, 2018	4	Positive Significant	Positive Significant	Positive Significant	Positive Significant	Contradictory	Positive Significant
Reychav et al., 2018	1		Positive Non-significant	Positive Non-significant	Positive Non-significant	Positive Non-significant	
Liu, Kang, Su et al., 2018	1	Positive Significant	Positive Significant	Positive Non-significant	Positive Significant	Positive Significant	Positive Significant
Liu, Kang, Domanska et al., 2018	1	Positive Significant	Positive Significant	Positive Significant	Positive Significant	Positive Significant	
Saqr & Alamro, 2019	1	Positive Non-significant	Positive Non-significant	Positive Significant	Positive Non-significant	Positive Non-significant	
Liu et al., 2019	1	Positive Significant			Negative Non-significant	Positive Non-significant	
Gašević et al., 2019	1	Negative Non-significant			Negative Significant	Positive Non-significant	

Positive Significant · Positive Non-significant · Negative Significant · Negative Non-significant · Contradictory

3. Methods

The methods implemented in this study are described in this section, in which we explain the context, the data collection methods, and the data analysis. A flowchart of the process is shown in Figure 1.

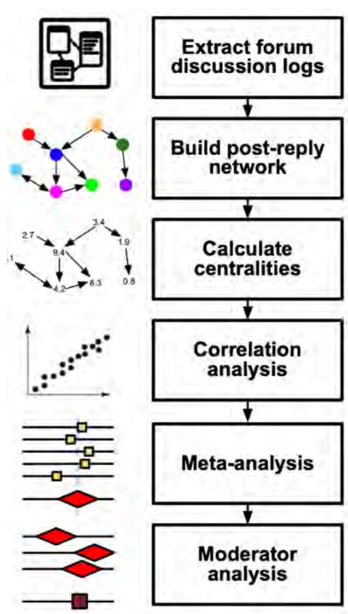

Figure 1. Methods implemented in the study.

3.1. Context

The present study is based on a full dataset of ***all*** the courses offered in a healthcare college of Qassim University between 2013 and 2019. To avoid selection or publication bias, *all courses were included* as long as they fulfilled the following criteria: 1) based on Problem-Based Learning (PBL) curriculum, 2) included online collaborative PBL discussions, 3) had 30 students or more (central limit theorem), and 4) had at least an average of two posts per students per week (to exclude courses with rare online interactions). The study included 69 course offerings (15 different courses) that matched the selection criteria, with a total of 3,747 students (3,277 completed the courses). Course duration ranged from four to eight weeks. Although each course covered different healthcare-related topics, they all had a common pedagogical underpinning based on PBL and the same assessment methods. Students were divided into small groups (Saqr et al., 2019) and were assigned an open-ended problem on a weekly basis. At the beginning of the week, students met face-to-face in their small groups to discuss and define the learning objectives with the help of a tutor. After the session and throughout the week, students engaged in online group discussions in a dedicated forum. The online forum facilitated the co-construction of knowledge, the exchange of perspectives, as well as discussions about possible ways to understand, solve, or study the problem. In other words, the online forum allowed the PBL process to be continuous throughout the week. By the end of the week, a closing face-to-face session took place in which students wrapped up what they had learned throughout the week. The course lectures, practical sessions, and seminars were well aligned with the weekly problems. The performance was measured by the course grades.

The course grades are the total of two grades. Continuous assessment, comprising 20% of the grade, is the total of grades obtained from assessment during the course. The remaining 80% of the grade comes from written exams (multiple-choice and short essay questions) which tests the knowledge acquisition of the PBL objectives in the forum interactions. The written exams come from an exam pool composed by the subject teachers (20 to 30 teachers). An exam committee, composed of the assessment, education, and quality assurance units (comprising domain and education expertise), holds several meetings and revises each item for language, content, and conformity with intended course objectives. After the exams, the assessment unit revises each item according to the standards adopted for psychometrics quality. Questions deemed problematic are excluded and the exam grade is calculated. The assessment methods in all courses were the same regarding question type, distribution of grades, and assessment methods and were all standardized throughout the program. Since the exams were set by the three committees described earlier, they were also less subject to teacher variations.

3.2. The Theoretical Analytical Framework

PBL is a student-centric approach in which learners engage in goal-directed inquiry. Students work in small groups with a facilitator (tutor) to discuss an assigned problem. As a collaborative learning approach, PBL emphasizes accountability, interdependence, and responsibility. Furthermore, the PBL process is well structured and scripted with pre-assigned roles. Students begin by discussing the problem, connecting it to their previous knowledge, defining their learning objectives, using their new knowledge to solve the problem, and then reflecting on their performance and group dynamics (Davidson & Major, 2014; Saqr, Nouri et al., 2020).

We adopted the methods of Enyedy and Stevens (2016) for analyzing collaboration. The authors state that, when analyzing collaborative learning, the unit of analysis can be either the individual learner or the collaborating group. Similarly, the interaction process analyzed can be at the individual learner level or the group interaction level. The outcome can be proximal (i.e., within the collaborative process) such as the very process of the collaboration or distal (i.e., outside the collaborative process) explained in terms of student performance in class or grades. This study addresses how the interactions and relations within the collaborative (proximal) process correlate with learning outcomes and performance operationalized as course grades of individual learners (distal outcome; Chen et al., 2018; Dado & Bodemer, 2017; Enyedy & Stevens, 2016). Since CSCL is a multifaceted process, we operationalize centrality measures to capture the participatory and social aspects of knowledge constructions (see Kreijns et al., 2013; Weinberger & Fischer, 2006). The operationalization of each centrality measure and review is detailed in the methods section.

3.3. Data Collection

Course data were extracted from the learning management system logs. The data collected included all forum posts related to collaboration tasks (PBL) in all courses. We retrieved the metadata for each post (ID, time, subject, title, content, post writer, and replies) and post writer (user ID, group ID, course ID, and grades). Non-collaborative forum posts (i.e., news, announcements, social interactions) were excluded. The post data were used to construct a post-reply network by considering the post writer as the *source* and the replied-to as the *target*, which is the most common configuration in the existing literature (Table 1). An example discussion is shown in Figure 2 along with the post-reply network that represents it.

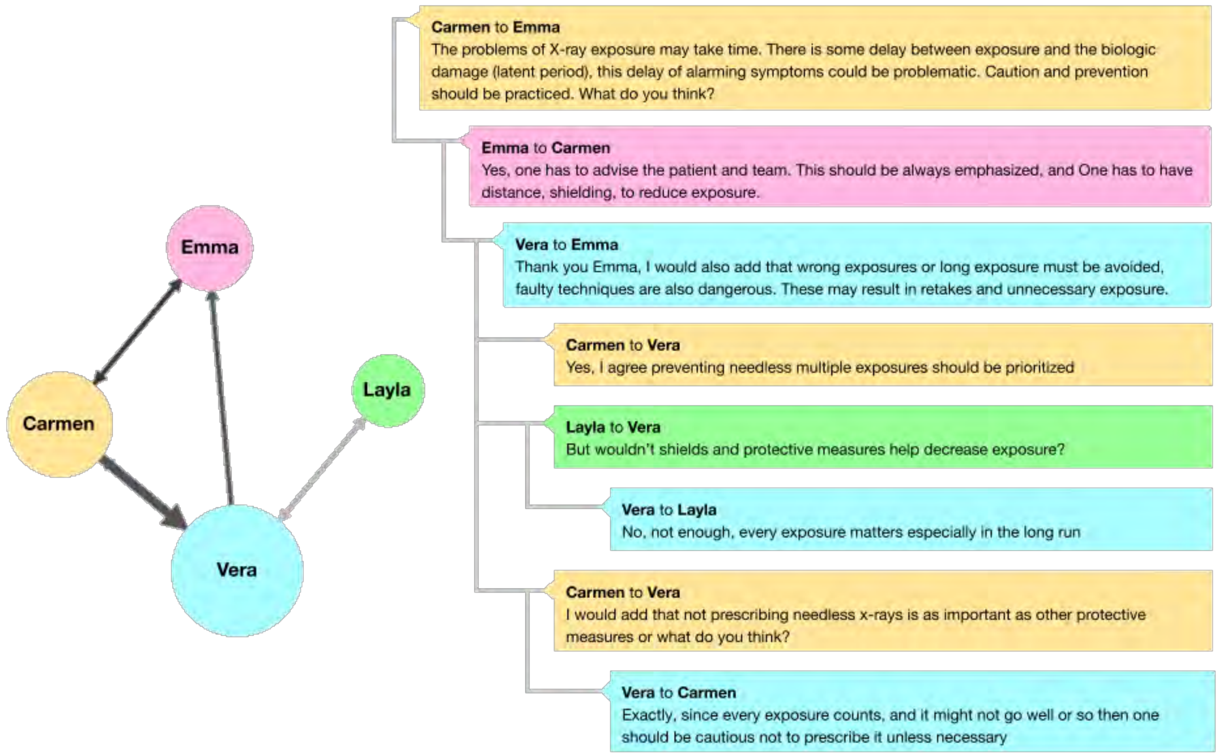

Figure 2. Sample discussion and post-reply network.

3.4. Network Analysis

All networks were prepared in an appropriate format and prepared for analysis with the Igraph R package (Csardi & Nepusz, 2006) implemented in the R programming language version 3.62 (R Core Team, 2018). For each student in each course, the six most used centrality measures were calculated:

Out-degree centrality: Number of posts contributed by a student. Out degree was operationalized as the effort and participation of the learner in the forums (Hernández-García et al., 2015; Saqr, Fors, & Nouri, 2018; Saqr, Viberg et al., 2020; Saqr & Alamro, 2019).
In-degree centrality: Number of replies a student received. In-degree was operationalized as the worthiness of the argument contributed by a student to stimulate discussion or debate (Hernández-García et al., 2015; Saqr, Fors, & Nouri, 2018; Saqr, Viberg et al., 2020; Saqr & Alamro, 2019).
Degree centrality: Sum of out-degree and in-degree centralities.
Betweenness centrality: Number of times a student bridged other unconnected students (the shortest path between them). Students with high betweenness centralities mediate interactions, control the flow of information, and have access to diverse perspectives and resources (Lü et al., 2016; Saqr, Viberg et al., 2020; Stephenson & Zelen, 1989).
Closeness centrality: The inverse farness from all students in the network. Closeness centrality was operationalized as accessibility, reachability, and ease of communication by all others (Lü et al., 2016; Saqr, Viberg et al., 2020; Stephenson & Zelen, 1989).
Eigenvector centrality: Reflects student positioning, selection of peers, and strength of all the relationships a student has (Hernández-García et al., 2015; Saqr, Fors, & Nouri, 2018; Saqr, Viberg et al., 2020; Saqr & Alamro, 2019).

3.5. Statistical Analysis

To enable comparison with previous studies, we separately computed Pearson's correlation coefficient and Spearman's rank correlation coefficient between grades and the centrality measures for each course offering. To prepare the variables for Pearson's correlation, we applied Box-Cox transformation to all variables so that they were closer to the normal distribution. We also calculated the Spearman correlation coefficients for comparison.

A multi-level correlation meta-analysis (using course and year level as subgroups) was performed to pool Pearson's correlation coefficients between grades and each centrality measure in all course offerings (each course offering was considered a separate study), i.e., we performed six multi-level meta-analyses for the six centralities to pool the correlations in the 69 course offerings. Meta-analysis offers a robust way to pool results while taking into account heterogeneity as well as sample sizes. The combined correlation coefficient was a weighted average. To obtain an accurate weight for each course offering, we performed inverse-variance pooling of Fisher's z transformed correlations (Marín-Martínez & Sánchez-Meca, 2010). A random-effects model was selected to report results since we expected the course offerings to be heterogeneous (which was confirmed by the moderate levels of heterogeneity indicators). A random-effects model assumes that the effect sizes have more variance than when drawn from a single population (Marín-Martínez & Sánchez-Meca, 2010). A moderator analysis was performed using the course subject and year levels as grouping variables to test if they influenced the results.

To evaluate the consistency of the correlation coefficients, we estimated the heterogeneity or between-study variance (i.e., the extent to which effect sizes vary within a meta-analysis (Higgins & Thompson, 2002; Schwarzer et al., 2015). Heterogeneity was determined using the Sidik-Jonkman estimator. The higher the levels of heterogeneity, the less consistent the correlation coefficients among the studies, which reflects on the low levels of confidence intervals. In other words, low levels of heterogeneity are a sign of consistent findings that increase the certainty that future applications of a given centrality measure would produce similar results to the ones obtained. I^2 — a measure of heterogeneity — was selected because it is not sensitive to changes in the number of studies and it is easy to interpret. An I^2 of 25% or lower indicates very low heterogeneity, an I^2 of 25–50% indicates low heterogeneity, an I^2 of 50–75% indicates moderate heterogeneity, and an I^2 greater than 75% indicates substantial heterogeneity (Hardy & Thompson, 1998; Higgins & Thompson, 2002).

The prediction interval is a measure of heterogeneity, which has been recently recommended to be reported in all meta-analyses as a robust and rigorous measure (IntHout et al., 2016). Prediction interval estimates our certainty of the future application of a centrality measure, i.e., the expected range of values within which the future correlation would probably lie. The prediction interval can be interpreted in a similar way to confidence intervals. That is, if the lower and upper bounds were on the positive side or both on the negative side, we would expect that future applications within similar contexts would have comparable results within the bounds of the predictive interval (IntHout et al., 2016).

Forest plots were used to illustrate the results of the meta-analysis graphically. They offer a summary of findings and statistical significance in an easy-to-read standardized manner. The forest plots layout for this study was selected to follow Cochrane Revman5 style, which is the most common (Schriger et al., 2010). The vertical line in the centre of the forest plot represents a correlation value of 0, whereas the horizontal lines represent the 95% confidence interval of the correlations for the corresponding course. The box in the middle represents the weight of each study (course offering in our case). The point inside the box represents the effect size. Studies with confidence intervals crossing the 0 line on either side are considered statistically insignificant. Studies with both confidence interval bounds on the right side of the 0 line are considered in favour of a statistically positive and significant correlation. Lastly, studies with both bounds of the confidence interval on the left side of the line are in favour of a statistically significant negative correlation.

4. Results

Table 2 shows the summary statistics by quartile of the courses included in the study. The median number of students enrolled in a course was 54 (a median of 48 completed the course) whereas the median number of teachers was 5 (i.e., there were 5 different small groups). The median frequency of interactions in a course was 1,210; the median degree of a student was 9, while the median degree of a teacher was 14. As such, the dataset had small-sized courses with relatively interactive students. The descriptive statistics of the centrality measures are detailed in Appendix A.

Table 2. Summary Statistics by Quartile of Courses

	Teachers			Students		
	25%	**Median**	**75%**	**25%**	**Median**	**75%**
Nodes	4	5	5	47	54	62
Edges	909	1,210	1,737	909	1,210	1,737
Degree	10.00	14.00	16.00	4.00	9.00	14.00
Closeness	0.07	0.09	0.11	0.05	0.07	0.09
Betweenness*	0.01	0.20	0.69	0.00	0.02	0.09
Eigenvector	0.15	0.33	0.66	0.12	0.32	0.65

* Betweenness is normalized.

4.1. Correlation Statistics

Figure 3 shows the frequency with which each centrality measure was correlated, whether positively or negatively, with academic achievement in each of the 69 course offerings.

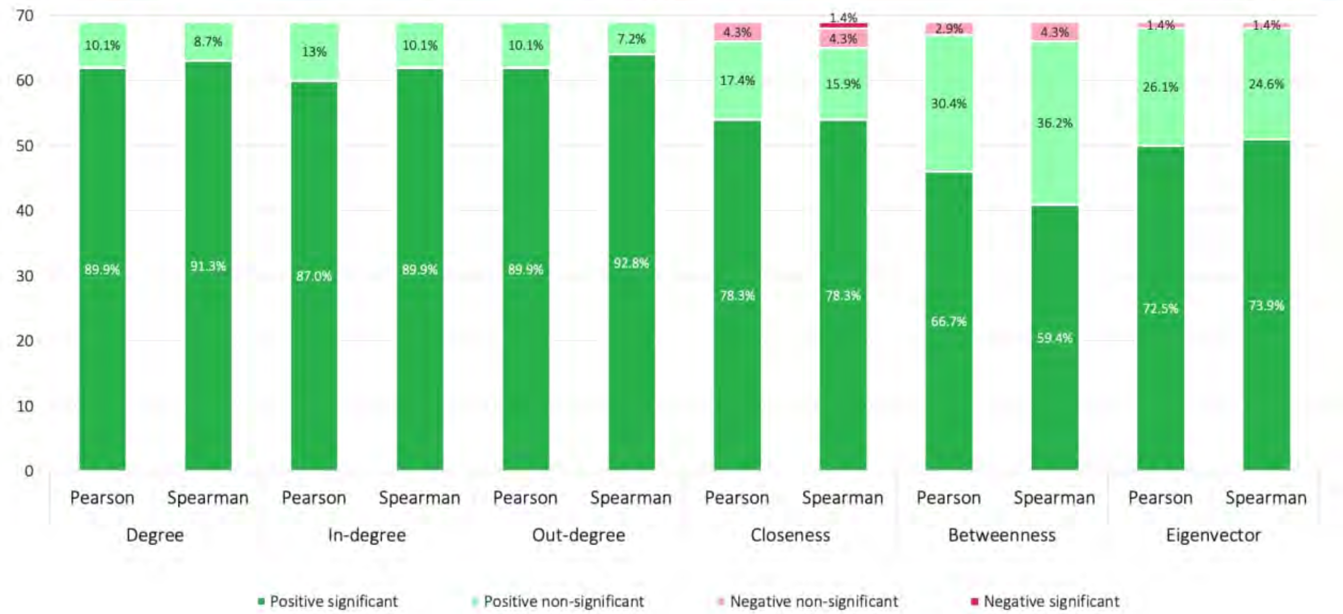

Figure 3. Summary of the frequency of correlations between grades and centrality measures.

Degree centralities were positively correlated with grades in all courses and more frequently so than in all other centralities. Pearson's correlation between grades and degree centrality was positive in all the studied courses. Closeness centrality followed in frequency, being positively and statistically significantly correlated with grades in 78.3% of courses (Pearson's correlation), while negatively but non-significantly correlated with grades in three (4.3%) courses, and negatively and statistically significantly correlated in a single course (only Spearman's correlation). Eigenvector centrality followed with 72.5% positively and statistically significantly correlated courses (73.9% Spearman's correlation); only in one course was eigenvector centrality negatively correlated with grades, although the resulting correlation was not statistically significant. Lastly, betweenness centrality was positively and statistically significantly correlated with grades in 66.7% of the courses (59.4% Spearman's correlation). In two of the courses, betweenness was negatively (although non-significantly) correlated with grades (three for Spearman's correlation). In summary, closeness and betweenness showed negative correlation with performance in around 4% of courses, while eigenvector centrality in 1%, and degree centralities showed no negative correlations.

4.2. Meta-Analysis

Since the simple frequency of correlations is far from optimal for estimating our confidence and certainty in these measures, we proceed to report the results of the meta-analysis for more detailed and in-depth analysis (Figure 4 and Table 3). The random-effects model combined correlation coefficients for the degree and out-degree centralities were proximate. So were the confidence intervals (r=0.56 [CI 0.52:0.60] for degree; r=0.55 [CI 0.51:0.59] for out-degree) and the predictive intervals ([0.25;0.77] for degree; [0.26;0.75] for out-degree). Both centralities had medium heterogeneity: I^2 was 55% for degree and 50% for out-degree. The in-degree centrality had a combined correlation coefficient that was close to degree and out-degree (r=0.54 [CI 0.49:0.58]). However, the lower limit of the predictive interval was relatively lower than the rest of degree centralities [0.17;0.77], and heterogeneity was relatively higher (I^2=63%). In summary, degree centralities have a combined correlation coefficient of medium strength, a reasonable predictive interval, and low heterogeneity. Such results highlight the value of these centrality measures as reliable indicators of achievement.

The combined correlation coefficient for betweenness centrality was markedly lower than that of the degree centralities (r=0.38 [CI 0.33:0.42]), with a lower predictive interval [0.07;0.62] and lower heterogeneity (I^2=45%). The low heterogeneity points to a consistent indicator of success. Although the combined correlation coefficient of closeness centrality (r=0.49 [CI 0.45:0.55]) was moderate, the lower bound of the predictive interval was negative [-0.06;0.81] and there was a high level of heterogeneity (I^2=79%). Taken together, the values of closeness centrality indicate that this measure is far from reliable as an indicator of student success.

The eigenvector centrality had a combined correlation coefficient slightly higher than that of betweenness centrality (r=0.41 [CI 0.36:0.45]), with a comparable predictive interval [0.07;0.54] and a slightly higher heterogeneity (I^2=54%). Similar to betweenness centrality, the eigenvector centrality is a consistent indicator of student success.

To investigate whether different courses or year levels influenced the magnitude or direction of correlations, we performed a test for subgroup differences (with the course and the year levels as grouping variables) using the random-effects-model. The

ISSN 1929-7750 (online).

test was statistically insignificant in all reported centralities, indicating that neither course type nor course level were significant moderators (Table 3).

Table 3. Summary of Multi-Level Meta-Analysis Results

	Correlation		Prediction interval		Heterogeneity	Test for subgroup differences	
Centrality	**r [CI]**	**p**	**Low**	**High**	**I^2 [CI]**	***Q***	**p**
Degree	0.56 [0.52;0.60]	< 0.0001	0.25	0.77	0.55 [0.41;0.66]	7.97	0.89
Out-degree	0.55 [0.51;0.59]	< 0.0001	0.26	0.75	0.50 [0.34;0.62]	9.44	0.80
In-degree	0.54 [0.49;0.58]	< 0.0001	0.17	0.77	0.63 [0.52;0.72]	8.86	0.84
Betweenness	0.38 [0.33;0.42]	< 0.0001	0.07	0.62	0.45 [0.27;0.59]	17.42	0.23
Closeness	0.49 [0.43;0.55]	< 0.0001	-0.06	0.81	0.79 [0.74;0.83]	13.77	0.47
Eigenvector	0.41 [0.36;0.45]	< 0.0001	0.07	0.66	0.54 [0.39;0.65]	8.29	0.87

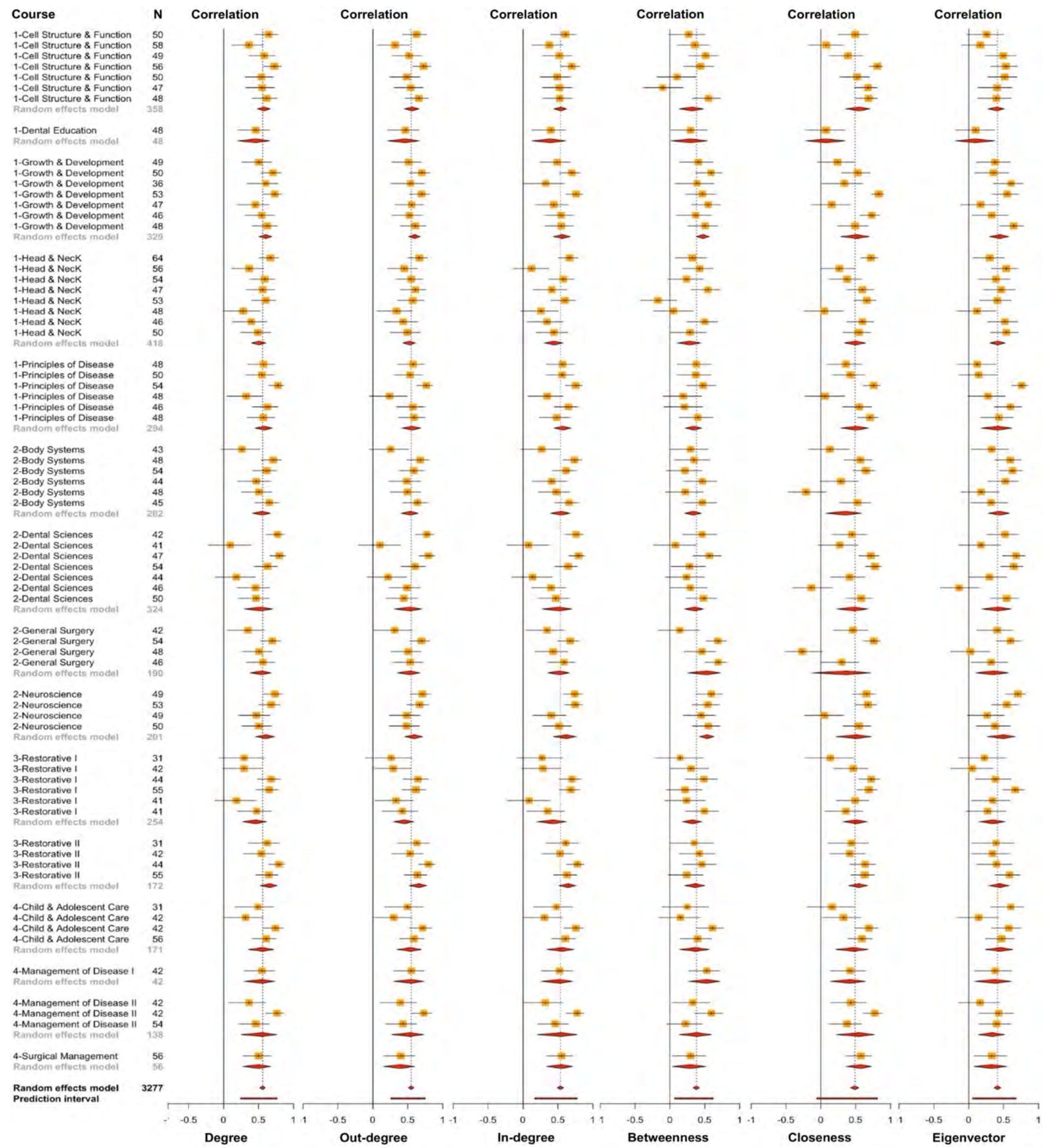

Figure 4. Forest plots of multi-level meta-analysis of correlation of centrality measures with grades.

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

5. Discussion

Which centrality measures are the most consistent indicators of success?

Among all centralities investigated in this study, degree centralities were the most positively and statistically significantly correlated with academic performance, which corroborates most of the previous literature. Three studies out of the reviewed 19 had non-statistically significant positive correlations (Cho et al., 2007; Saqr, Fors, & Tedre, 2018; Saqr & Alamro, 2019). In the case of Cho et al. (2007), the study included a small number of students (31) in an engineering course. While in Saqr, Fors, and Tedre (2018), the network was highly centralized with few collaborative interactions among students. Students who received more replies (higher in-degree) were more likely to score higher. Similarly, the study by Saqr and Alamro (2019) in which problem-based learning was used, showed positive correlation between grades and in-degree (although with p=0.06), and statistically significant correlation with out-degree centrality. The two studies that reported a negative association used regression models with centralities combined (Gašević et al., 2019; Osatuyi & Passerini, 2016). Our results have shown more consistent correlations of moderate to high strength, narrow confidence intervals, and moderate to high prediction intervals. This may be explained by the fact that our study was conducted in a context where the collaborative module was central to the course design. Degree centralities are always operationalized as indicators of participation in the discourse, effort, contribution to the collaborative task (Hernández-García et al., 2015; Reychav et al., 2018; Saqr, Viberg et al., 2020), and social positioning (Poquet & Jovanovic, 2020), which are important determinants of student learning (Weinberger & Fischer, 2006). Taken together, these findings point to the portability of degree centrality as a reliable indicator of success regardless of the course subject. That is, in courses that emphasize the collaborative module as an essential part of course design, it is expected that degree centralities represent the highest, most consistent indicators of success among all centrality measures. This study offers evidence of the utility of said measure in relevant settings where such participatory dimensions are emphasized.

The results of betweenness centrality have shown a relatively low correlation coefficient (compared to degree centrality) as well as a small lower bound for the predictive interval (0.07). Among the reviewed studies, only seven have reported statistically positive, significant correlations (de-Marcos et al., 2016; Hernández-García et al., 2015; Liu, Kang, Domanska et al., 2018; Liu, Kang, Su et al., 2018; Osatuyi & Passerini, 2016; Putnik et al., 2016; Saqr, Fors, & Tedre, 2018). Three studies have reported inconsistent results among the offered courses. Two studies were conducted in MOOCs (Jiang et al., 2014; Joksimović et al., 2016) where participation may not be as intense or emphasized as in traditional higher education, and participants' pre-knowledge is more heterogenous. These findings corroborate the pattern observed with degree centralities, where centrality measures do not show strong correlations. In another study by (Saqr, Fors, & Nouri, 2018), CSCL was offered to small groups where everyone was connected; therefore, betweenness may not be a differentiating factor.

Regarding closeness centrality, our findings showed a moderate combined correlation coefficient but a non-statistically significant predictive interval, pointing to its poor reliability as an indicator of student success. These results are supported by the existing literature, which showed a negative statistically significant correlation in one study (Cadima et al., 2012), as well as contradictory results within the same study among different courses (Gašević et al., 2019; Jiang et al., 2014; Joksimović et al., 2016; Liu et al., 2019). Most of the contradictory results have been reported in MOOCs where that pattern has been observed before.

Although closeness and betweenness centralities were the most commonly used in our reviewed literature, their performance — as indicators of success — was not as initially expected in the reviewed works nor in our study (Jiang et al., 2014; Joksimović et al., 2016). An explanation may be that closeness and betweenness centralities are calculated based on shortest paths, which raises questions about their relevance to some digital interaction settings. For instance, betweenness centrality counts the times a post has bridged two others (Borgatti, 2005). In some tasks, this might be relevant (e.g., when students have to argue or debate), while in other configurations and network topologies, this might not be appropriate (in small groups where everybody is connected; e.g., Saqr, Fors, & Nouri, 2018). What is more, betweenness centrality might be challenging to interpret and operationalize in settings where students are well connected. The same issue applies to closeness centrality: the measure captures the distance to "all" others. However, it is not necessary for each student to connect or interact with "all" others. The "all" others are particularly difficult in MOOC situations with numerous students, many of whom barely participate. Furthermore, it is meaningless to calculate closeness centrality when some students are disconnected (due to the absence of a path to "all"), creating difficulty in interpretation and operationalization (Borgatti & Brass, 2019; Borgatti & Everett, 2006; Freeman, 1978). The fact that closeness and betweenness centralities showed negative correlations with performance in our study in some courses and showed narrow or insignificant predictive intervals in our large and homogenous sample raises serious concerns about their utility at least in similar contexts. Such concerns are not new, perhaps requiring the educational community to examine new centrality measures more relevant to the studied tasks. While this is far from a trivial undertaking, it remains long overdue. On the other hand, our results corroborate those of the existing literature on the consistency of eigenvector centrality, which was statistically significant in all the reviewed studies (de-Marcos et al., 2016; Hernández-García et al., 2015; Liu, Kang, Su et al., 2018; Putnik et al., 2016; Saqr, Fors, & Nouri, 2018; Saqr, Fors, & Tedre, 2018; Saqr, Viberg et al., 2020; Traxler et al., 2016; Wise & Cui, 2018). Therefore, based on the findings of this study, we may suggest that researchers with similar contexts (e.g., CSCL where the collaborative module is central to the course design) rely on degree and eigenvector centralities as indicators for success while exercising caution when using short-path centralities (closeness and betweenness).

How certain can we be in future applications of centrality measures as indicators of success?

Our third research question aimed to investigate the consistency of results obtained from a large homogenous sample. That is, how homogenous or heterogenous our results are and, therefore, how far we can be confident that future applications of the same measure can result in similar findings. Heterogeneity, a method commonly implemented in meta-analysis, allows for such estimation. The heterogeneity reported for most centrality measures ranged from low to moderate. Since courses can rarely be an exact copy of each other, there may be changes in, for example, teachers, schedules, and load from other parallel courses. Thus, a moderate level of heterogeneity is expected. The presence of low to moderate levels of heterogeneity is in line with the reported large-scale replications that previously attempted to replicate psychological studies (e.g., Aarts et al., 2015; Hagger et al., 2016; Klein, 2019). In fact, previous meta-analyses of very close replications (studies with similar conditions) have reported similar low to moderate levels of heterogeneity (Aarts et al., 2015; Hagger et al., 2016; Klein, 2019). For instance, the "Many Labs" study replicated behavioural experiments in 36 labs using identical conditions, and reported heterogeneity in all fourteen studied effects (Aarts et al., 2015). In the present study, we found no evidence that course or year level contributed to the heterogeneity reported in our study.

In the case of closeness centrality, we found substantial heterogeneity. This can be explained by the way closeness centrality is estimated based on the distance to "every" other student in the network. The presence of inactive or isolated students to whom the distance will be very high would skew the measure. Thus, interactive students in a network with isolates will get low closeness centrality scores. These findings highlight the needed caution when applying the measure as an indicator for success in disconnected networks or groups with inactive students.

The choice of meta-analysis was based on the seminal examples in the literature that assessed the replicability and consistency of previously reported results (e.g., Hagger et al., 2016; Klein, 2019). In view of the results obtained, the meta-analysis has been useful compared to traditional methods, which often list the number of studies side-by-side or use all data together as a single dataset. The benefits of the meta-analysis include estimation of the heterogeneity, accurate estimation of the combined correlation coefficient (taking into account sample size), as well as estimation of the predictive interval (taking into account heterogeneity; Aarts et al., 2015; Hagger et al., 2016; Klein, 2019). For instance, relying on traditional methods, one would assume that closeness centrality would be positively correlated with performance in more than 80% of cases. However, the predictive interval shows that future results of closeness centralities are far from certain (being statistically insignificant). In our study, the large sample size and number of courses, and the choice of methods makes our results more likely to generalize in similar contexts, in particular the predictive intervals. However, it remains to be seen whether future studies can prove or refute our findings.

Centrality measures represent a single aspect of CSCL (Hoppe, 2017), and researchers need to thoughtfully use different computational methods at hand to gain insights about different determinants of learning. For instance, content analysis would allow a more nuanced idea about how knowledge is constructed. Automated approaches and computational methods can serve an important role in the automation of content analysis (e.g., Erkens et al., 2016). Analysis of the temporal aspects of the interactions using methods that capture the rhythm and sequence of the process gives an overarching idea about the collaborative process (Boroujeni et al., 2017; Skrypnyk et al., 2015). Triangulation of data from different sources would help us better understand collaborative learning and draw the correct conclusions (Hoppe, 2017). Another limitation of the commonly used *traditional* centrality measures reviewed here is that they fail to capture the branching and the nestedness of discussions. Therefore, educational researchers interested in capturing how a post is nested, stimulating or helping other students to engage and build-upon, use diffusion (Hoppe, 2017; Poquet & Jovanovic, 2020; Saqr & Viberg, 2020; Suthers, 2015; Suthers & Desiato, 2012).

Centrality measures are used to capture students' social positioning where ties can indicate trust, friendship, and access to knowledge. Therefore, operationalization of network measures must consider not just the structural position but also the context, the interaction type, and, most importantly, the benefits received by students (Poquet & Jovanovic, 2020). As pointed by Poquet and Jovanovic (2020), closeness centrality is always interpreted as being close to or reachable by all students. However, this is far from realistic since forum networks are temporary events with transitory edges. Betweenness centrality is also commonly interpreted as "bridging" or linking isolated communities. This view was borrowed from organizational network analysis and does not reflect the context or structure of learning networks. In our study, both closeness and betweenness centralities have been poorly linked to learning outcomes, indicating the need to consider their usage, interpretation, and operationalization.

The increasing availability of interaction data has allowed several possibilities for constructing leaner networks (Poquet & Jovanovic, 2020). While our study focused on the often-used post-reply networks, several other possibilities of constructing networks warrant further investigation, e.g., co-participation networks (Fincham et al., 2018; Poquet et al., 2020; Poquet & Jovanovic, 2020; Wise et al., 2017). Since centrality measures may vary depending on how the network is constructed, it is expected that the results obtained in this study could vary if another network configuration or edge weight were used. Combining network configuration methods is also an interesting, rarely explored area. Poquet and Jovanovic (2020) demonstrated that building both post-reply and co-participation networks can provide a nuanced interpretation of centrality measures. While these approaches are commonly seen as exclusive, the authors demonstrate how they can be complementary. (For a discussion of network configuration methods, interested readers are encouraged to consult Poquet & Jovanovic, 2020).

Centrality measures are valuable tools which have proven useful across many fields, including education. Nonetheless, there are still many unanswered questions regarding the measurement, operationalization, and use of centrality measures. Further research could investigate how different network configuration methods — e.g., co-participation networks or combinations of network configurations — could contribute to our understanding of social learning and how such configuration can help obtain reliable indicators of student success. Future research could also contribute to a more nuanced alignment between learning contexts, social positioning, and operationalization of centrality measures (Poquet et al., 2021; Poquet & Jovanovic, 2020). Another area of further research would be improving measurement, reporting, and computation of novel centrality measures (Saqr & López-Pernas, 2021a, 2021b).

The present study has some limitations. First, replication of our results beyond the healthcare context may be needed to verify or refute our findings. Moreover, in our study, we have not combined the centrality measures, reporting instead on each centrality individually. We have resorted to this option since we found significant multi-collinearity in most courses between several centrality measures that made any statistically robust predictive model containing any combination difficult. Therefore, understanding the correlation of combined centrality measures with academic achievement is an important issue for future research.

6. Conclusions

In this study, we empirically tested and verified the role of centrality measures as indicators of success in collaborative learning. For this purpose, we attempted to reproduce the most commonly used centrality measures in the literature in all the courses of a college over five years of education. Using a large dataset in a meta-analysis enabled us to make important conclusions. The first and most important conclusion is that degree and eigenvector centrality measures can be consistent indicators of performance in settings where course design emphasizes collaboration. The correlation between centrality measures and academic achievement was reproducible regardless of the number of students, number of interactions, year of study, or course subject. Thus, our results provide evidence of the validity and suitability of centralities as indicators of success and, possibly, of productive interactions. Our results also shed light on novel aspects of the consistency of centrality measures that have not been studied before, namely, heterogeneity and pooled effect sizes, as well as the predictive interval. Our findings raise questions about the suitability of closeness (e.g., in disconnected groups) and betweenness centrality (e.g., in connected groups) measures as indicators of success in collaborative settings.

Declaration of Conflicting Interest

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors declare no financial support for the research, authorship, and/or publication of this article.

References

Aarts, A., Anderson, J., Anderson, C., Attridge, P., Attwood, A., Axt, J., Babel, M., Bahník, Š., Baranski, E., Barnett-Cowan, M., Bartmess, E., Beer, J., Bell, R., Bentley, H., Beyan, L., Binion, G., Borsboom, D., Bosch, A., Bosco, F., & Penuliar, M. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). https://doi.org/10.1126/science.aac4716

Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M. Á., & Hernández-García, Á. (2014). Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning. *Computers in Human Behavior*, *31*(1), 542–550. https://doi.org/10.1016/j.chb.2013.05.031

Alexander, J. L., Smith, K. A., Mataras, T., Shepley, S. B., & Ayres, K. M. (2015). A meta-analysis and systematic review of the literature to evaluate potential threats to internal validity in probe procedures for chained tasks. *The Journal of Special Education*, *49*(3), 135–145. https://doi.org/10.1177/0022466914550096

Andres, J. M. L., Baker, R. S., Gašević, D., Siemens, G., Crossley, S. A., & Joksimović, S. (2018). Studying MOOC completion at scale using the MOOC replication framework. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (LAK '18), 5–9 March 2018, Sydney, NSW, Australia (pp. 71–78). ACM. https://doi.org/10.1145/3170358.3170369

Andres, J. M. L., Baker, R. S., Siemens, G., Gašević, D., & Spann, C. A. (2015). Replicating 21 findings on student success in online learning. *Technology, Instruction, Cognition, and Learning*, *10*, 313–333.

Baker, M., & Penny, D. (2016). Is there a reproducibility crisis? *Nature*, *533*(7604), 452–454. https://doi.org/10.1038/533452A

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

Bernard, R. M., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A., & Bethel, E. C. (2009). A meta-analysis of three types of interaction treatments in distance education. *Review of Educational Research*, *79*(3), 1243–1289. https://doi.org/10.3102/0034654309333844

Borgatti, S. P. (2005). Centrality and network flow. *Social Networks*, *27*(1), 55–71. https://doi.org/10.1016/j.socnet.2004.11.008

Borgatti, S. P., & Brass, D. J. (2019). Centrality: Concepts and measures. In D. J. Brass & S. P. Borgatti (Eds.), *Social Networks at Work* (pp. 9–22). Routledge.

Borgatti, S. P., & Everett, M. G. (2006). A graph-theoretic perspective on centrality. *Social Networks*, *28*(4), 466–484. https://doi.org/10.1016/j.socnet.2005.11.005

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science*, *323*(5916), 892–895. https://doi.org/10.1126/science.1165821

Borokhovski, E., Bernard, R. M., Tamim, R. M., Schmid, R. F., & Sokolovskaya, A. (2016). Technology-supported student interaction in post-secondary education: A meta-analysis of designed versus contextual treatments. *Computers & Education*, *96*, 15–28. https://doi.org/10.1016/j.compedu.2015.11.004

Boroujeni, M. S., Hoppe, H. U., Hecking, T., & Dillenbourg, P. (2017). Dynamics of MOOC discussion forums. *Proceedings of the 7th International Conference on Learning Analytics and Knowledge* (LAK '17), 13–17 March 2017, Vancouver, BC, Canada (pp. 128–137). ACM. https://doi.org/10.1145/3027385.3027391

Cadima, R., Ojeda, J., & Monguet, J. M. (2012). Social networks and performance in distributed learning communities. *Educational Technology & Society*, *15*(4), 296–304.

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

Chan, A.-W., Hróbjartsson, A., Haahr, M. T., Gøtzsche, P. C., & Altman, D. G. (2004). Empirical evidence for selective reporting of outcomes in randomized trials. *JAMA*, *291*(20), 2457. https://doi.org/10.1001/jama.291.20.2457

Chen, J., Wang, M., Kirschner, P. A., & Tsai, C.-C. (2018). The role of collaboration, computer use, learning environments, and supporting strategies in CSCL: A meta-analysis. *Review of Educational Research*, *88*(6), 799–843. https://doi.org/10.3102/0034654318791584

Cho, H., Gay, G., Davidson, B., & Ingraffea, A. (2007). Social networks, communication styles, and learning performance in a CSCL community. *Computers & Education*, *49*(2), 309–329. https://doi.org/10.1016/j.compedu.2005.07.003

Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2017). Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle LMS. *IEEE Transactions on Learning Technologies*, *10*(1), 17–29. https://doi.org/10.1109/TLT.2016.2616312

Csardi, G., & Nepusz, T. (2006). The Igraph software package for complex network research. *InterJournal*, *Complex Sy*, 1695.

Dado, M., & Bodemer, D. (2017). A review of methodological applications of social network analysis in computer-supported collaborative learning. *Educational Research Review*, *22*, 159–180. https://doi.org/10.1016/j.edurev.2017.08.005

Davidson, N., & Major, C. H. (2014). Boundary crossings: Cooperative learning, collaborative learning, and problem-based learning. *Journal on Excellence in College Teaching*, *25*(3&4), 7–55.

Dawson, S., Joksimovic, S., Poquet, O., & Siemens, G. (2019). Increasing the impact of learning analytics. *Proceedings of the 9th International Conference on Learning Analytics and Knowledge* (LAK '19), 4–8 March 2019, Tempe, AZ, USA (pp. 446–455). ACM. https://doi.org/10.1145/3303772.3303784 de-Marcos, L., García-López, E., García-Cabot, A., Medina-Merodio, J.-A., Domínguez, A., Martínez-Herráiz, J.-J., & Diez-Folledo, T. (2016). Social network analysis of a gamified e-learning course: Small-world phenomenon and network metrics as predictors of academic performance. *Computers in Human Behavior*, *60*, 312–321. https://doi.org/10.1016/j.chb.2016.02.052

Dwan, K., Gamble, C., Williamson, P. R., & Kirkham, J. J. (2013). Systematic review of the empirical evidence of study publication bias and outcome reporting bias: An updated review. *PLOS ONE*, *8*(7). https://doi.org/10.1371/journal.pone.0066844

Enyedy, N., & Stevens, R. (2016). Analyzing collaboration. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 191–212). Cambridge University Press. https://doi.org/10.1017/CBO9781139519526.013

Erkens, M., Bodemer, D., & Hoppe, H. U. (2016). Improving collaborative learning in the classroom: Text mining based grouping and representing. *International Journal of Computer-Supported Collaborative Learning*, *11*(4), 387–415. https://doi.org/10.1007/s11412-016-9243-5

Fincham, E., Gašević, D., & Pardo, A. (2018). From social ties to network processes: Do tie definitions matter? *Journal of Learning Analytics*, *5*(2), 9–28. https://doi.org/10.18608/jla.2018.52.2

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239. https://doi.org/10.1016/0378-8733(78)90021-7

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, *28*, 68–84. https://doi.org/10.1016/j.iheduc.2015.10.002

Gašević, D., Joksimović, S., Eagan, B. R., & Shaffer, D. W. (2019). SENS: Network analytics to combine social and cognitive perspectives of collaborative learning. *Computers in Human Behavior*, *92*, 562–577. https://doi.org/10.1016/j.chb.2018.07.003

Gašević, D., Zouaq, A., & Janzen, R. (2013). "Choose your classmates, your GPA is at stake!" *American Behavioral Scientist*, *57*(10), 1460–1479. https://doi.org/10.1177/0002764213479362

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., Brand, R., Brandt, M. J., Brewer, G., Bruyneel, S., Calvillo, D. P., Campbell, W. K., Cannon, P. R., Carlucci, M., Carruth, N. P., Cheung, T., Crowell, A., De Ridder, D. T. D., Dewitte, S., . . . Zwienenberg, M. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science*, *11*(4), 546–573. https://doi.org/10.1177/1745691616652873

Hardy, R. J., & Thompson, S. G. (1998). Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine*, *17*(8), 841–856. https://doi.org/10.1002/(SICI)1097-0258(19980430)17:8<841::AID-SIM781>3.0.CO;2-D

Hernández-García, Á., González-González, I., Jiménez-Zarco, A. I., & Chaparro-Peláez, J. (2015). Applying social learning analytics to message boards in online distance learning: A case study. *Computers in Human Behavior*, *47*, 68–80. https://doi.org/10.1016/j.chb.2014.10.038

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

Hoppe, H. U. (2017). Computational methods for the analysis of learning and knowledge building communities. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *Handbook of learning analytics* (pp. 23–33). Society for Learning Analytics Research (SoLAR). https://doi.org/10.18608/hla17.002

Ifenthaler, D., & Yau, J. Y.-K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research and Development*, *68*(4), 1961–1990. https://doi.org/10.1007/s11423-020-09788-z

IntHout, J., Ioannidis, J. P. A., Rovers, M. M., & Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*, *6*(7). https://doi.org/10.1136/bmjopen-2015-010247

Jeong, H., & Hmelo-Silver, C. E. (2016). Seven affordances of computer-supported collaborative learning: How to support collaborative learning? How can technologies help? *Educational Psychologist*, *51*(2), 247–265. https://doi.org/10.1080/00461520.2016.1158654

Jiang, S., Fitzhugh, S. M., & Warschauer, M. (2014). Social positioning and performance in MOOCs. *CEUR Workshop Proceedings*, *1183*, 55–58.

Jo, I., Park, Y., & Lee, H. (2017). Three interaction patterns on asynchronous online discussion behaviours: A methodological comparison. *Journal of Computer Assisted Learning*, *33*(2), 106–122. https://doi.org/10.1111/jcal.12168

Joksimović, S., Manataki, A., Gašević, D., Dawson, S., Kovanović, V., & de Kereki, I. F. (2016). Translating network position into performance. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 314–323). ACM. https://doi.org/10.1145/2883851.2883928

Joober, R., Schmitz, N., Annable, L., & Boksa, P. (2012). Publication bias: What are the challenges and can they be overcome? *Journal of Psychiatry & Neuroscience*, *37*(3), 149–152. https://doi.org/10.1503/jpn.120065

Klein, R. A. (2019). Correction to Klein et al. (2014). *Social Psychology*, *50*(3), 211–213. https://doi.org/10.1027/1864-9335/a000373

Kreijns, K., Kirschner, P. A., & Vermeulen, M. (2013). Social aspects of CSCL environments: A research framework. *Educational Psychologist*, *48*(4), 229–242. https://doi.org/10.1080/00461520.2012.750225

Li, N., Marsh, V., Rienties, B., & Whitelock, D. (2017). Online learning experiences of new versus continuing learners: A large-scale replication study. *Assessment & Evaluation in Higher Education*, *42*(4), 657–672. https://doi.org/10.1080/02602938.2016.1176989

Liu, C. H., & Matthews, R. (2005). Vygotsky's philosophy: Constructivism and its criticisms examined. *International Education Journal*, *6*(3), 386–399.

Liu, S., Chai, H., Liu, Z., Pinkwart, N., Han, X., & Hu, T. (2019). Effects of proactive personality and social centrality on learning performance in SPOCs. *Proceedings of the 11th International Conference on Computer Supported Education* (CSEDU 2019) 2–4 May 2019, Heraklion, Crete, Greece (Vol. 2, pp. 481–487). ScitePress. https://doi.org/10.5220/0007756604810487

Liu, Z., Kang, L., Domanska, M., Liu, S., Sun, J., & Fang, C. (2018). Social network characteristics of learners in a course forum and their relationship to learning outcomes. *Proceedings of the 10th International Conference on Computer Supported Education* (CSEDU 2018) 15–17 March 2018, Funchal, Madeira, Portugal (Vol. 1, pp. 15–21). ScitePress. https://doi.org/10.5220/0006647600150021

Liu, Z., Kang, L., Su, Z., Liu, S., & Sun, J. (2018). Investigate the relationship between learners' social characteristics and academic achievements. *Journal of Physics: Conference Series*, *1113*(1), 012021. https://doi.org/10.1088/1742-6596/1113/1/012021

Lü, L., Chen, D., Ren, X.-L., Zhang, Q.-M., Zhang, Y.-C., & Zhou, T. (2016). Vital nodes identification in complex networks. *Physics Reports*, *650*, 1–63. https://doi.org/10.1016/j.physrep.2016.06.007

Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty. *Educational Researcher*, *43*(6), 304–316. https://doi.org/10.3102/0013189X14545513

Marín-Martínez, F., & Sánchez-Meca, J. (2010). Weighting by inverse variance or by sample size in random-effects meta-analysis. *Educational and Psychological Measurement*, *70*(1), 56–73. https://doi.org/10.1177/0013164409344534

Martin, A. J., & Dowson, M. (2009). Interpersonal relationships, motivation, engagement, and achievement: Yields for theory, current issues, and educational practice. *Review of Educational Research*, *79*(1), 327–365. https://doi.org/10.3102/0034654308325583

Osatuyi, B., & Passerini, K. (2016). Twittermania: Understanding how social media technologies impact engagement and academic performance of a new generation of learners. *Communications of the Association for Information Systems*, *39*(1), 509–528. https://doi.org/10.17705/1CAIS.03923

Plesser, H. E. (2018). Reproducibility vs. replicability: A brief history of a confused terminology. *Frontiers in Neuroinformatics*, *11*(January), 1–4. https://doi.org/10.3389/fninf.2017.00076

Poquet, O., & Jovanovic, J. (2020). Intergroup and interpersonal forum positioning in shared-thread and post-reply networks. *Proceedings of the 10th International Conference on Learning Analytics and Knowledge* (LAK '20), 23–27 March 2020, Frankfurt, Germany (pp. 187–196). ACM. https://doi.org/10.1145/3375462.3375533

Poquet, O., Saqr, M., & Chen, B. (2021). Recommendations for network research in learning analytics: To open a conversation. In O. Poquet, B. Chen, M. Saqr, & T. Hecking (Eds.), *Proceedings of the NetSciLA2021 Workshop "Using Network Science in Learning Analytics: Building Bridges towards a Common Agenda" (NetSciLA2021)* (Issue 2868, pp. 34–41). http://ceur-ws.org/Vol-2868/

Poquet, O., Tupikina, L., & Santolini, M. (2020). Are forum networks social networks? *Proceedings of the 10th International Conference on Learning Analytics and Knowledge* (LAK '20), 23–27 March 2020, Frankfurt, Germany (pp. 366–375). ACM. https://doi.org/10.1145/3375462.3375531

Putnik, G., Costa, E., Alves, C., Castro, H., Varela, L., & Shah, V. (2016). Analysing the correlation between social network analysis measures and performance of students in social network-based engineering education. *International Journal of Technology and Design Education*, *26*(3), 413–437. https://doi.org/10.1007/s10798-015-9318-z

R Core Team. (2018). *R: A language and environment for statistical computing*.

Reychav, I., Raban, D. R., & McHaney, R. (2018). Centrality measures and academic achievement in computerized classroom social networks. *Journal of Educational Computing Research*, *56*(4), 589–618. https://doi.org/10.1177/0735633117715749

Rienties, B., Cross, S., & Zdrahal, Z. (2017, February). Implementing a learning analytics intervention and evaluation framework: What works? In B. Kei Daniel (Ed.), *Big data and learning analytics in higher education* (pp. 147–166). Springer International Publishing. https://doi.org/10.1007/978-3-319-06520-5_10

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. https://doi.org/10.1016/j.compedu.2013.06.009

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*(1), 160. https://doi.org/10.1186/s12909-019-1599-6

Saqr, M., Fors, U., & Nouri, J. (2018). Using social network analysis to understand online problem-based learning and predict performance. *PLOS ONE*, *13*(9). https://doi.org/10.1371/journal.pone.0203590

Saqr, M., Fors, U., & Tedre, M. (2018). How the study of online collaborative learning can guide teachers and predict students' performance in a medical course. *BMC Medical Education*, *18*(1), 24. https://doi.org/10.1186/s12909-018-1126-1

Saqr, M., Fors, U., Tedre, M., & Nouri, J. (2018). How social network analysis can be used to monitor online collaborative learning and guide an informed intervention. *PLOS ONE*, *13*(3). https://doi.org/10.1371/journal.pone.0194777

Saqr, M., & López-Pernas, S. (2021a). Modelling diffusion in computer-supported collaborative learning: A large-scale learning analytics study. *International Journal of Computer-Supported Collaborative Learning*, in-press.

Saqr, M., & López-Pernas, S. (2021b). Idiographic learning analytics: A definition and a case study. *Proceedings of the 21st IEEE International Conference on Advanced Learning Technologies* (ICALT 2021), 12–15 July 2021, online (pp. 163–165). IEEE Computer Society. https://doi.org/10.1109/ICALT52272.2021.00056

Saqr, M., Nouri, J., & Jormanainen, I. (2019). A learning analytics study of the effect of group size on social dynamics and performance in online collaborative learning. In M. Scheffel, J. Broisin, V. Pammer-Schindler, A. Ioannou, & J. Schneider (Eds.), *Lecture Notes in Computer Science* (Vol. 11722, pp. 466–479). Springer. https://doi.org/10.1007/978-3-030-29736-7_35

Saqr, M., Nouri, J., Vartiainen, H., & Malmberg, J. (2020). What makes an online problem-based group successful? A learning analytics study using social network analysis. *BMC Medical Education*, *20*(1), 80. https://doi.org/10.1186/s12909-020-01997-7

Saqr, M., & Viberg, O. (2020). Using diffusion network analytics to examine and support knowledge construction in CSCL settings. In C. Alario-Hoyos, M. J. Rodríguez-Triana, M. Scheffel, I. Arnedillo-Sánchez, & S. M. Dennerlein (Eds.), *Proceedings of EC-TEL 2020: Addressing Global Challenges and Quality Education* (Vol. 12315, pp. 158–172). Springer International Publishing. https://doi.org/10.1007/978-3-030-57717-9_12

Saqr, M., Viberg, O., & Vartiainen, H. (2020). Capturing the participation and social dimensions of computer-supported collaborative learning through social network analysis: Which method and measures matter? *International Journal of Computer-Supported Collaborative Learning*, *15*(2), 227–248. https://doi.org/10.1007/s11412-020-09322-6

Schriger, D. L., Altman, D. G., Vetter, J. A., Heafner, T., & Moher, D. (2010). Forest plots in reports of systematic reviews: A cross-sectional study reviewing current practice. *International Journal of Epidemiology*, *39*(2), 421–429. https://doi.org/10.1093/ije/dyp370

Schwarzer, G., Carpenter, J. R., & Rücker, G. (2015). *Meta-analysis with R*. Springer International Publishing. https://doi.org/10.1007/978-3-319-21416-0

Shafie, T. (2019). A multigraph approach to social network analysis. *Journal of Social Structure*, *16*(1), 0–21. https://doi.org/10.21307/joss-2019-011

Sie, R. L. L., Ullmann, T. D., Rajagopal, K., Cela, K., Rijpkema, M. B., & Sloep, P. B. (2012). Social network analysis for technology-enhanced learning: Review and future directions. *International Journal of Technology Enhanced Learning*, *4*(3/4), 172. https://doi.org/10.1504/IJTEL.2012.051582

Skrypnyk, O., Joksimović, S., Kovanović, V., Gasšević, D., & Dawson, S. (2015). Roles of course facilitators, learners, and technology in the flow of information of a CMOOC. *International Review of Research in Open and Distance Learning*, *16*(3), 188–217.

Stephenson, K., & Zelen, M. (1989). Rethinking centrality: Methods and examples. *Social Networks*, *11*(1), 1–37. https://doi.org/10.1016/0378-8733(89)90016-6

Suthers, D. D. (2015). From contingencies to network-level phenomena. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 368–377). ACM. https://doi.org/10.1145/2723576.2723626

Suthers, D. D., & Desiato, C. (2012). Exposing chat features through analysis of uptake between contributions. *Proceedings of the 45th Hawaii International Conference on System Sciences* (HICSS-45), 4–7 January 2012, Maui, HI, USA (pp. 3368–3377). IEEE Computer Society. https://doi.org/10.1109/HICSS.2012.274

Traxler, A. L., Gavrin, A., & Lindell, R. S. (2016). CourseNetworking and community: Linking online discussion networks and course success. *Proceedings of the 2016 Physics Education Research Conference* (PERC 2016), 20–21 July 2016, Sacramento, CA, USA (pp. 352–355). https://doi.org/10.1119/perc.2016.pr.083

Wecker, C., & Fischer, F. (2014). Where is the evidence? A meta-analysis on the role of argumentation for the acquisition of domain-specific knowledge in computer-supported collaborative learning. *Computers & Education*, *75*, 218–228. https://doi.org/10.1016/j.compedu.2014.02.016

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*, *46*(1), 71–95. https://doi.org/10.1016/j.compedu.2005.04.003

Wise, A. F., & Cui, Y. (2018). Unpacking the relationship between discussion forum participation and learning in MOOCs. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (LAK '18), 5–9 March 2018, Sydney, NSW, Australia (pp. 330–339). ACM. https://doi.org/10.1145/3170358.3170403

Wise, A. F., Cui, Y., & Jin, W. Q. (2017). Honing in on social learning networks in MOOC forums. *Proceedings of the 7th International Conference on Learning Analytics and Knowledge* (LAK '17), 13–17 March 2017, Vancouver, BC, Canada (pp. 383–392). ACM. https://doi.org/10.1145/3027385.3027446

Wise, A. F., & Schwarz, B. B. (2017). Visions of CSCL: Eight provocations for the future of the field. *International Journal of Computer-Supported Collaborative Learning*, *12*(4), 423–467. https://doi.org/10.1007/s11412-017-9267-5

Appendix A

Table 4. Descriptive Statistics of Centrality Measures for Each Course Offering

Course Name	Students	Teachers	Edges	Degree	Betweenness	Closeness	Eigenvector
1-Cell Structure & Function	50	3	1,737	14	0.07	0.09	0.37
1-Cell Structure & Function	59	4	1,400	13	0.02	0.06	0.37
1-Cell Structure & Function	55	4	1,051	7	0.00	0.04	0.19
1-Cell Structure & Function	58	5	1,554	11	0.02	0.06	0.33
1-Cell Structure & Function	68	6	1,526	15	0.06	0.07	0.64
1-Cell Structure & Function	77	6	1,479	14	0.09	0.06	0.35
1-Cell Structure & Function	67	4	1,348	12	0.02	0.07	0.54
1-Dental Education	48	2	496	9	0.09	0.08	0.59
1-Growth & Development	49	5	1,572	10	0.03	0.06	0.04
1-Growth & Development	52	5	1,022	7	0.01	0.03	0.34
1-Growth & Development	42	4	1,115	6	0.00	0.06	0.19
1-Growth & Development	54	5	1,981	12	0.02	0.08	0.21
1-Growth & Development	61	4	722	10	0.11	0.08	0.57
1-Growth & Development	69	6	1,050	10	0.03	0.06	0.29
1-Growth & Development	66	6	1,116	9	0.03	0.06	0.42
1-Head & Neck	75	4	2,193	9	0.02	0.05	0.13
1-Head & Neck	56	6	906	6	0.02	0.07	0.16
1-Head & Neck	54	5	1,236	10	0.04	0.06	0.39
1-Head & Neck	47	4	1,102	5	0.00	0.05	0.16
1-Head & Neck	54	5	1,730	15	0.02	0.08	0.42
1-Head & Neck	64	5	1,149	13	0.14	0.08	0.51
1-Head & Neck	73	5	811	9	0.09	0.06	0.40
1-Head & Neck	70	5	745	6	0.01	0.05	0.27
1-Principles of Disease	48	4	1,767	9	0.03	0.08	0.06
1-Principles of Disease	52	5	903	5	0.00	0.05	0.45
1-Principles of Disease	54	3	1,370	11	0.01	0.07	0.26
1-Principles of Disease	62	6	784	10	0.07	0.07	0.56

Course Name	Students	Teachers	Edges	Degree	Betweenness	Closeness	Eigenvector
1-Principles of Disease	71	6	1,170	11	0.05	0.06	0.43
1-Principles of Disease	67	4	1,009	9	0.01	0.06	0.49
2-Body Systems	43	5	1,219	10	0.06	0.09	0.49
2-Body Systems	48	5	1,476	6	0.00	0.06	0.35
2-Body Systems	54	5	3,134	18	0.01	0.09	0.53
2-Body Systems	55	5	1,666	16	0.02	0.08	0.65
2-Body Systems	68	7	2,771	15	0.06	0.09	0.36
2-Body Systems	62	7	1,555	9	0.02	0.08	0.52
2-Dental Sciences	42	5	2,535	12	0.03	0.09	0.19
2-Dental Sciences	41	5	2,075	9	0.02	0.10	0.16
2-Dental Sciences	47	5	810	4	0.00	0.03	0.14
2-Dental Sciences	57	4	1,210	10	0.01	0.05	0.36
2-Dental Sciences	53	5	1,010	11	0.05	0.08	0.49
2-Dental Sciences	70	7	2,197	12	0.08	0.08	0.14
2-Dental Sciences	67	6	731	3	0.00	0.05	0.13
2-General Surgery	42	4	542	5	0.05	0.05	0.20
2-General Surgery	54	4	1,116	13	0.05	0.08	0.40
2-General Surgery	68	6	1,040	10	0.09	0.08	0.27
2-General Surgery	62	7	746	4	0.00	0.07	0.46
2-Neuroscience	49	5	696	2	0.00	0.05	0.33
2-Neuroscience	53	5	1,033	12	0.02	0.08	0.47
2-Neuroscience	69	5	1,051	10	0.09	0.08	0.22
2-Neuroscience	66	4	567	1	0.00	0.01	0.08
3-Restorative I	31	5	2,156	8	0.00	0.11	0.14
3-Restorative I	42	5	2,482	13	0.01	0.11	0.29
3-Restorative I	45	5	1,173	6	0.00	0.06	0.36
3-Restorative I	55	5	2,453	17	0.02	0.08	0.33
3-Restorative I	49	5	1,215	12	0.06	0.09	0.58

Course Name	Students	Teachers	Edges	Degree	Betweenness	Closeness	Eigenvector
3-Restorative I	49	5	909	9	0.03	0.08	0.58
3-Restorative II	31	5	3,377	6	0.00	0.10	0.08
3-Restorative II	42	5	2,566	11	0.00	0.10	0.16
3-Restorative II	44	5	1,384	5	0.00	0.06	0.22
3-Restorative II	55	5	2,976	16	0.01	0.08	0.24
4-Child & Adolescent Care	31	5	2,295	6	0.00	0.09	0.07
4-Child & Adolescent Care	43	5	1,839	10	0.01	0.10	0.24
4-Child & Adolescent Care	44	5	1,239	5	0.00	0.06	0.28
4-Child & Adolescent Care	56	5	1,805	13	0.00	0.07	0.31
4-Management of Disease I	42	5	840	7	0.05	0.08	0.17
4-Management of Disease II	43	4	685	6	0.06	0.08	0.14
4-Management of Disease II	43	4	643	3	0.00	0.05	0.16
4-Management of Disease II	54	5	1,103	8	0.01	0.08	0.23
4-Surgical Management	56	5	809	8	0.00	0.06	0.15