Large-scale Analysis of Discussion Networks in College Courses

Kimberly Williamson
Cornell University
Ithaca, NY, USA
khw44@cornell.edu

René F. Kizilcec
Cornell University
Ithaca, NY, USA
kizilcec@cornell.edu

ABSTRACT

Online discussion boards serve an important role in college courses by facilitating social learning and student support. However, the student experience and learning outcomes are likely to depend on the structure of student engagement with one another and the teaching staff. Using social network analysis, we investigated the network structure of 616 course discussion boards at a selective research university. We first examine variation in discussion boards using a wide range of composite metrics from the social network analysis literature. We then develop a typology of discussion board networks using principal component analysis and k-Means clustering to arrive at three clusters: dense discussion; distinct discussion groups; and discussion brokers and hubs.

CCS CONCEPTS

• **Applied computing → Learning management systems**; • **Networks → Overlay and other logical network structures**.

KEYWORDS

discussion boards, social network analysis, higher education

ACM Reference Format:
Kimberly Williamson and René F. Kizilcec. 2022. Large-scale Analysis of Discussion Networks in College Courses. In *Proceedings of the Ninth ACM Conference on Learning @ Scale (L@S '22), June 1–3, 2022, New York City, NY, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3491140.3528321

1 INTRODUCTION

Online discussion boards are widely used in higher education to facilitate student discourse, community, and help-seeking, especially in large lecture-based courses [21]. They are generally believed to enhance students' course experience [4, 6, 35] and their learning outcomes [21]. However, how a discussion board is used by students in a course can vary substantially and lead to different outcomes. To understand this variation and its consequences, we need to develop a measure of discussion networks that can scale across hundreds of courses and capture variation along interpretable dimensions. Social network analysis (SNA) techniques, in particular, have been used to examine patterns in academic achievement [10, 18], student peer relationships [28, 36], and instructional design [26, 27]. However, many educational research studies that use digital trace data are not scaled or scalable to multiple learning contexts to understand general patterns in student behavior. While some modifications to algorithms and models may still be needed to fit different contexts, there is significant value in establishing a standard typography of course structures in higher education [7]. This typography could be used to compare classrooms across various contexts and understand the relationship between these structures and student outcomes.

We demonstrate the analytic process and scientific value of scaling classroom structures research by developing a typology of discussion networks in college courses.We use SNA techniques to quantify and explore the network structure of 616 course discussion boards at a selective research university in the United States. Using a range of composite metrics from the SNA literature, we examine variation in discussion boards and develop a typology of discussion board networks using principal component analysis and k-Means clustering. This work contributes a general typography of online discussion networks to the education literature.

Online discussion boards for college courses enable asynchronous social interactions between course participants using a standard set of interaction types, primarily writing posts and replying to posts. Social Network Analysis is a method for exploring and analyzing the structure of social interactions, which makes it a natural method to choose for studies of social interactions on discussion boards [23, 33]. SNA relies on the two core components of a network: nodes and edges. Network nodes can be defined in several ways (students, posts, assignments, courses, etc.) and connected with edge relationship, such as students replying to discussion posts or students enrolling in the same courses. Once a network is defined by its nodes and edges (and possibly edge weights and direction), the SNA literature provides a rich set of metrics to describe the properties of a network. The following network metrics are widely used to describe the structure of networks:

Homophily is the tendency of nodes with similar characteristics to be more closely connected than nodes with different characteristics [22]. If nodes are students, their characteristics can be socio-demographic, academic, or other background information. A discussion network with positive socio-demographic homophily is one where students with shared characteristics reply to each other.

Reciprocity is the likelihood that nodes in a directed graph are mutual [33]. For student nodes in a discussion network, reciprocity can measure the likelihood that two students reply to each others' posts. A discussion network with a high reciprocity is one where the students reply back to each other often.

Density is the ratio of actual ties to the potential ties [33]. For each pair of nodes in a course, there is a chance for two directed links, and density would be the percentage of those possible ties present in the network. A discussion network with a high density is one where all nodes talk to each other.

L@S '22, June 1–3, 2022, New York City, NY, USA

ACM ISBN 978-1-4503-9158-0/22/06…$15.00
https://doi.org/10.1145/3491140.3528321

Diameter is the longest distance between any two students in a course. A discussion network with a short diameter is one where most students have directly replied to all other students discussion post at least once.

Transitivity is a probability measure for the existence of interconnected communities of three or more nodes. For example, for each triplet of nodes in a course (students A, B, and C), transitivity is the probability that A and C will be linked if B is linked to both A and C. A discussion network with a high transitivity is one where students tend to form clusters with each other, where all nodes of the cluster reply to each other.

Degree Centralization is a distribution score for node degree centrality [33]. There are three measures of degree centralization for a directed network graph: In, Out, and All degree. In-degree centrality is the number of other nodes in the course that replied to their posts. Out-degree centrality for a student is the number of other nodes who replied to their posts. All-degree requires both an In and Out edge for each node pair. The degree centralization for each degree metric measures how evenly distributed the centrality is amongst the nodes in the course. An example of a discussion network with a high degree centralization is one where all nodes have at least replied once to a single node.

Closeness Centralization is a distribution score for node closeness centrality [33]. Closeness for a student is the inverse of the average path distance between the student and every other node. Path distance is the number of links needed to connect two nodes in a course. Closeness Centralization measures how evenly distributed the closeness centrality is amongst the nodes in the course. A discussion network with a high closeness centralization is one where a single node is the only replier or repliee.

Betweenness Centralization is a distribution score for node betweenness centrality [33]. The importance of each student for keeping the network connected is a students's betweenness centrality. A student with high betweenness could be a broker of knowledge across different discussion communities in the course. Betweenness centralization measures how evenly distributed the betweenness centrality is amongst the nodes in the course. A discussion network with a high betweenness centralization is one where a single node is the only replier or repliee.

Most research on networks in education has studied either course-level or university-level networks. These studies show that there is precedence for applications of SNA in higher education, and that SNA techniques can be used to understand various social phenomena in higher education. However, while previous network research has occurred on the individual classroom scale and the university scale, we are not aware of studies that examine course-level networks at the university scale. One reason for the limited research on university course networks may be the lack of standardized data collection across multiple courses at a university. The widespread adoption of learning management systems has led to many classroom functions being assisted by technology, including submitting homework, reporting grades and feedback from instructors, and online written communication via discussions boards [1, 11, 32]. Educational researchers have recently started to use the data collected by learning management systems [3, 16, 17].

Several in-depth network analysis studies have highlighted various interaction types and patterns among students [14, 28]. These studies, frequently focusing on a small number of courses, explored peer relationships [14], cheating [31], and student leadership in course groups [29, 36]. A number of network studies in education have examined social interactions in the context of massive open online courses (MOOCs), which have thousands of learners in their discussion boards [2, 9, 13, 15, 19, 37]. With the large number of learners in MOOCs, SNA has proven effective at distilling insights such as how discourse [9, 35] and course structure [2] affect learners' course performance. For example, Dowell et al. [9] explored how learners' discourse is related to their network position and academic performance. They found that discourse patterns of high performers differed from those in central network positions, raising questions about whether centrality measures alone can explain course phenomena. This work inspired us to explore which discussion network measures are associated with academic achievement.

University-wide network analysis studies have examined issues of student retention [10], organizational structure [20], and most recently the spread of COVID-19 [8, 34]. For example, Eckles and Stradley [10] used student organization rosters, class registration, and roommate assignments to construct a network of relationships among first-year students. The researchers found that centrality and density measures were significant in predicting the likelihood of a student being retained for the next semester. These studies demonstrate the usefulness of SNA for answering research questions involving large, diverse samples like an entire university community. Our study uses university-wide data but applies classroom-level SNA to many courses on campus.

Our overarching goal is to understand the variation in course discussion networks. To accomplish this goal, we need to develop a scalable approach to quantify variation in discussion networks. We did not find prior work on which network-based measures are effective at characterizing different structures of course discussion networks. Nearly all SNA studies in education focus on egocentric metrics, such identifying students who are leaders [36] or how a given student's network position relates to their retention in college [10]. In contrast, our goal is to characterize the entire network structure (i.e. one value per network instead of one value per student) to compare many networks. This leads us to pose the following **research question**: Which network-based measures capture variation in student interactions on course discussion boards?

2 METHODOLOGY

2.1 Dataset and Context

The study sample consists of Canvas discussion post entries for courses offered in a two-year period (2019-2021) at a selective research university in the United States. This period was chosen because the Fall 2019 semester was the first semester where Canvas was the official LMS, and Spring 2021 was the last full semester of data in the dataset. We only analyzed Fall and Spring courses with at least 50 reply posts in a course over the semester to ensure a critical mass of students and communication to construct a network. Our final sample consist of 51,875 student-course observations for 17,330 students and 616 courses across 102 course subject areas. Average course enrollment is 84.2 (SD=138) and the average number of replies is 256 (SD=398). We further obtained student enrollment information and sociodemographic data from the Student Information System (SIS). The student demographics are 57.2% female, 42.8% male, 23.8% URM, 35.3% White, 19.2% Asian, 13.4% Hispanic, 12.8% International, 7% Unknown, 6.9% Black, 5% Two or More Races, 0.3% American Indian, 0.1% Hawaiian/Pacific Islander; and 15.5% are first-generation college students.

We created a directed edge list from the Canvas discussion data for each course using a Direct Reply characterization to define network ties [12]. Thus, a directed edge is drawn if a person replies to another. The data was analyzed using the iGraph [5] R package. We calculated a set of standard network metrics (reviewed above) for each course in the dataset:

Homophily: Four measures of homophily were calculated for this study based on sociodemographic characteristics (Gender, Race, Ethnicity, First Generation, Citizenship). Given that these measures were categorical, we used Newman [25]'s assortativity coefficient to calculate homophily scores for each characteristic.

Reciprocity: Reciprocity was calculated as the number of edges that pointed in both directions divided by the maximum number of possible double pointed edges, given at least one edge.

Density: Density was calculated as the proportion of actual ties relative to the maximum number of ties possible.

Diameter: The diameter was calculated by finding the longest shortest path distance between any two nodes in the network. Each node needed to connect the two nodes raises the diameter by 1.

Transitivity: A global transitivity measure was calculated for each course. We measured transitivity by counting the number of edges in a closed triplet of nodes divided by the maximum amount of edges possible if all node triplets were closed. Closed in this calculation means that all three nodes in a triplet have an edge with the other two nodes.

Degree Centralization: We calculated three measures for Degree Centralization, one calculating the distribution of in degree centrality scores, another for the distribution of the out degree centrality scores, and the last measure being the distribution of centrality scores for a network where only mutual ties are considered. A mutual tie is created in a directed network when two nodes have a double pointed edge.

Closeness Centralization: We calculated three measures for Closeness Centralization: in closeness centrality scores, out closeness centrality scores, and centrality scores for a network where only mutual ties are considered.

Betweenness Centralization: As opposed to the other two measures of centralization that are directional (in or out), betweenness centralization is one measure for a directed network graph.

We calculated the network metrics for each course discussion network to examine their distributions and correlations. Figure 1 shows the distribution of all metrics we considered (rescaling homophily metrics from [-1,1] to [0,1] for ease of presentation).

3 FINDINGS

We find that all metrics exhibit moderate levels of variation across courses, which indicates that discussion networks differ along key dimensions. Density, reciprocity, and transitivity have medians below 0.5, which indicates that most discussion networks are sparse. The higher levels of transitivity compared to density and reciprocity are indicative of high clustering within many courses. Unlike the

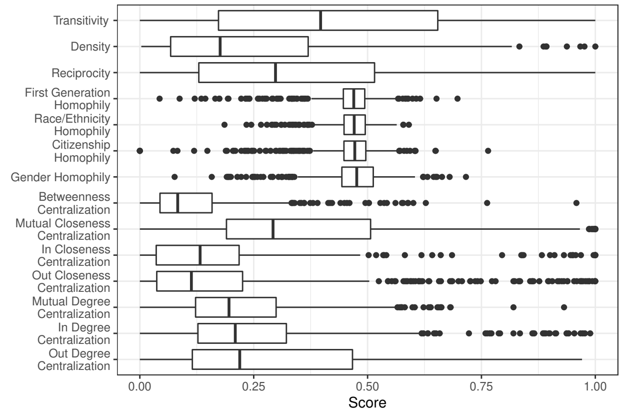

Figure 1: Distribution of each discussion network metric across 616 college courses.

centralization metrics, which show many outliers, there appears to be a wider spread of values for density, reciprocity, and transitivity. We also find that closely related metrics of centralization exhibit substantial variation in their medians and inter-quartile ranges, suggesting that they are capturing different sources of variation for the networks. In contrast, the four homophily metrics show much less variability across courses and have a similar median. A few courses exhibit assortativity based on sociodemographic characteristics, but most courses have a homophily score between -.2 and .2 (rescaled to [0.4,0.6] in Figure 1). Since most of the courses in this sample contained numerical gender parity, looking at courses on the extremes of gender homophily may offer insights regarding inequities. In cases of unbalanced sociodemographic characteristics, reviewing homophily scores alone is less informative.

To understand how the different metrics are associated with each other, we examine their bivariate correlations and find multiple clusters of metrics. The centralization metrics are highly intercorrelated, which is expected as they measure similar aspects of the network structure. Reciprocity, transitivity, and density are correlated positively with each other, but negatively with the number of students enrolled in a course. To address this course enrollment bias for the rest of the analysis, we normalize each of these three metrics by dividing by the course enrollment (i.e. number of nodes in the network). The presence of strong positive and negative correlations between many network features indicates that the dimensionality of this metric space can be reduced. We performed a principal component analysis over all network metrics and found that over 50% of the variance is accounted for by the first two components alone. We continue with a two-component representation of the metrics, but note that a three-component representation yields similar network clusters in the end. To understand what the two components represent, we examine which metrics contribute the most to each of them. We find that the first principal component is highly related to centralization metrics, while the second component is most related to reciprocity, density, and transitivity.

To identify clusters of discussion networks, we use k-Means clustering over the two principle components. We determine the optimal number of clusters using the within-cluster sums of squares,

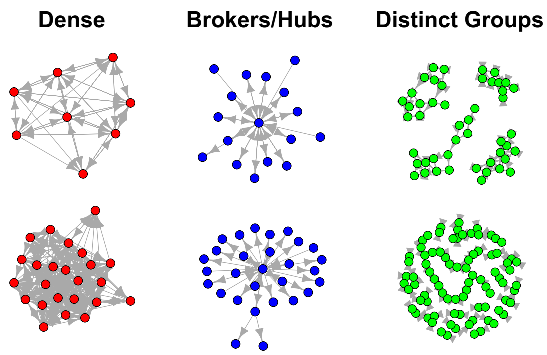

Figure 2: Examples of discussion networks for courses in each network type based on clustering.

average silhouette and gap statistics for values of k between 2 and 10. This consistently yielded $k = 3$ as the optimal number of clusters. These three clusters are the basis for the proposed discussion network typology. Upon closer examination of the three clusters of discussion networks, we labeled them as dense discussions (red), distinct discussion groups (green), and discussion brokers and hubs (blue). We now describe the nature of each type of discussion network visualized in Figure 2.

Dense discussion networks are characterized by the number of replies to multiple students. The distribution of edges suggests that there are no set instructions/rules in the course for who students should communicate with on the discussion board. These courses have high levels of density, reciprocity, and transitivity. Figure 2 shows examples of two courses with this type of discussion network. Both networks are dense but the bottom one has a few students on the edge of the diagram with few connections, while most students are concentrated in the middle. The students on the periphery could have withdrawn from the course, explaining their minimal links, or this could be a sign of disengagement. We hypothesize that this network type indicates courses that grade course discussion.

Distinct discussion groups tend to have higher homophily scores, which indicates that students only interact with a small number of peers in the course and that the course does not have a diverse sociodemographic enrollment. For example, if a course with this network type has only five students out of 100 who identify as students of color, most groups will look racially homogeneous. Figure 2 shows examples of two courses with this network type. The one on top shows distinct clusters with most discussion within and little discussion outside of particular groups. The other course is more connected across groups, but groups are still clearly present. This structure could represent courses where groups change over the semester, but only a few students actually switch groups.

Discussion brokers and hubs are networks with high centralization scores, which indicates that there are a few nodes that are very central to the network, while most nodes have few edges. Figure 2 shows examples of two courses with this network type. Both networks appear to have a topic leader structure, with most students replying to the posts of only a few other nodes. The leader could be an instructor, teaching assistant, or students asking about topics of particular interest. Another interpretation of this network structure is that students take turns leading the course for a week, and the students can choose what technology to use for their week. In this case, maybe only a few students choose to use Canvas, which makes them stand out as topic leaders in the network on Canvas.

4 DISCUSSION

This study contributes new insights into the anatomy of student discussions in college courses. We used well-established network measures from social network analysis to develop a scalable typology of discussion networks: Distinct Groups, Dense Groups, and Brokers and Hubs. This typology can be used to understand how instructors actively or passively shape peer interactions in their classes. This has implications for the student experience in the course, specifically their sense of community and peer support, which are known to influence student success in higher education [30]. The social network structure could also influence inequities, for instance for first-generation college students who benefit from being connected with peers in the class. Structured learning environments have been found to promote equitable participation [24]. The three distinctive types of course discussion structures we found may generalize to other institutions and enable similar inquiries. Our work demonstrates the utility of SNA metrics to capture variation between different course discussion boards and how this variation could be used in future work to explore connections with student achievement and sociodemographic achievement gaps.

Course instructors and teaching assistants have substantial control over the course and we expect that the observed discussion board structure reflects the intentions of the teaching team. The typology we presented here could provide instructors with feedback on whether students are interacting with each other in the class as intended. Instructors could also use results from this study to compare different instructional practices related to discussion posts. For example, it could help an instructor evaluate the outcomes of trying out new guidelines for the use of discussion boards in their course. These insights provide new feedback loops that can help inform and refine instructional practices for social learning and support in courses. The results also have practical implications for academic leaders who seek to encourage the formation of social ties between students during their college experience. An overview of course-level social networks could provide insight into how the social experience relates to students' college persistence and issues of equity and inclusion. Future research can offer insight into how discussion board structure and achievement outcomes are related.

We note two limitations of this initial study and directions for future work. First, the context of this research is a selective research university during a period of time that included in-person instruction and a transition to online learning due to the COVID-19 pandemic. These contextual factors influence our findings and we therefore call for replications of this analysis at other institutions with access to LMS data. Second, while using replies of a discussion board to create a network depicts one type of structure, more information in terms of the content of the posts could lead to another type of structure. Future research could examine the content of the posts to create different types of links such as characterizing supportive, dismissive, and other types of connections between students.

REFERENCES

[1] Nor Azura Adzharuddin and L Hwei Ling. 2013. Learning management system (LMS) among university students: Does it work. *International Journal of e-Education, e-Business, e-Management and e-Learning* 3, 3 (2013), 248–252.

[2] Mina Shirvani Boroujeni, Tobias Hecking, H. Ulrich Hoppe, and Pierre Dillenbourg. 2017. Dynamics of MOOC discussion forums. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference (LAK '17)*. Association for Computing Machinery, 128–137. https://doi.org/10.1145/3027385.3027391

[3] Magdalena Cantabella, Raquel Martínez-España, Belén Ayuso, Juan Antonio Yáñez, and Andrés Muñoz. 2019. Analysis of student behavior in learning management systems through a Big Data framework. *Future Generation Computer Systems* 90 (2019), 262–272.

[4] Bonnie J. Covelli. 2017. Online Discussion Boards: The Practice of Building Community for Adult Learners. *The Journal of Continuing Higher Education* 65, 2 (May 2017), 139–145.

[5] Gabor Csardi and Tamas Nepusz. 2006. The igraph software package for complex network research. *InterJournal* Complex Systems (2006), 1695. https://igraph.org

[6] Corinne Dalelio. 2013. Student Participation in Online Discussion Boards in a Higher Education Setting. *International Journal on E-Learning* 12, 3 (Jun 2013), 249–271.

[7] Dan Davis, Daniel Seaton, Claudia Hauff, and Geert-Jan Houben. 2018. Toward large-scale learning design: categorizing course designs in service of supporting learning outcomes. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale (L@S '18)*. Association for Computing Machinery, 1–10. https://doi.org/10.1145/3231644.3231663

[8] Gina Deom, Stefano Fiorini, Mark McConahay, Linda Shepard, and Julie Teague. 2021. Data-Driven Decisions: Using Network Analysis to Guide Campus Course Offerings. *College and University* 96, 4 (2021), 2–13.

[9] Nia M. Dowell, Oleksandra Skrypnyk, Srecko Joksimovic, Arthur C. Graesser, Shane Dawson, Dragan Gaševic, Thieme A. Hennis, Pieter de Vries, and Vitomir Kovanovic. 2015. *Modeling Learners' Social Centrality and Performance through Language and Discourse.* International Educational Data Mining Society. https://eric.ed.gov/?id=ED560532

[10] James E. Eckles and Eric G. Stradley. 2012. A social network analysis of student retention using archival data. *Social Psychology of Education* 15, 2 (June 2012), 165–180. https://doi.org/10.1007/s11218-011-9173-z 110.

[11] Marissa R. Fearnley and Johnny T. Amora. 2020. Learning Management System Adoption in Higher Education Using the Extended Technology Acceptance Model. *IAFOR Journal of Education* 8, 2 (Jul 2020), 89–106.

[12] Ed Fincham, Dragan Gašević, and Abelardo Pardo. 2018. From Social Ties to Network Processes: Do Tie Definitions Matter? *Journal of Learning Analytics* 5, 2 (Aug 2018), 9–28.

[13] Nabeel Gillani, Taha Yasseri, Rebecca Eynon, and Isis Hjorth. 2014. Structural limitations of learning in a crowd: communication vulnerability and information diffusion in MOOCs. *Scientific Reports* 4, 1 (Sep 2014), 6447. https://doi.org/10.1038/srep06447 75.

[14] Daniel Z. Grunspan, Benjamin L. Wiggins, and Steven M. Goodreau. 2014. Understanding Classrooms through Social Network Analysis: A Primer for Social Network Analysis in Education Research. *CBE—Life Sciences Education* 13, 2 (June 2014), 167–178. https://doi.org/10.1187/cbe.13-08-0162 249.

[15] Jonathan Huang, Anirban Dasgupta, Arpita Ghosh, Jane Manning, and Marc Sanders. 2014. Superposter behavior in MOOC forums. In *Proceedings of the first ACM conference on Learning@ scale conference*. ACM, 117–126.

[16] Sandeep M Jayaprakash, Erik W Moody, Eitel JM Lauría, James R Regan, and Joshua D Baron. 2014. Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics* 1, 1 (2014), 6–47.

[17] Libor Juhaňák, Jiří Zounek, and Lucie Rohlíková. 2019. Using process mining to analyze students' quiz-taking behavior patterns in a learning management system. *Computers in Human Behavior* 92 (2019), 496–506.

[18] Reynol Junco. 2015. Student class standing, Facebook use, and academic performance. *Journal of Applied Developmental Psychology* 36 (Jan. 2015), 18–29. https://doi.org/10.1016/j.appdev.2014.11.001 323.

[19] René F Kizilcec. 2013. Collaborative learning in geographically distributed and in-person groups. In *AIED 2013 Workshops Proceedings Volume*, Vol. 67.

[20] Alexis V. Knaub, Charles Henderson, and Kathleen Quardokus Fisher. 2018. Finding the leaders: an examination of social network analysis and leadership identification in STEM education change. *International Journal of STEM Education* 5, 1 (Jun 2018), 26. https://doi.org/10.1186/s40594-018-0124-5 24.

[21] Kathleen A. Krentler and Laura A. Willis-Flurry. 2005. Does Technology Enhance Actual Student Learning? The Case of Online Discussion Boards. *Journal of Education for Business* 80, 6 (Jul 2005), 316–321.

[22] Miller McPherson, Lynn Smith-Lovin, and James M Cook. 2001. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology* 27, 1 (Aug 2001), 415–444. https://doi.org/10.1146/annurev.soc.27.1.415 19210.

[23] Peter R. Monge and Noshir S. Contractor. 2003. *Theories of communication networks*. Oxford University Press. 3047.

[24] Connor Neill, Sehoya Cotner, Michelle Driessen, and Cissy J Ballen. 2019. Structured learning environments are required to promote equitable participation. *Chemistry Education Research and Practice* 20, 1 (2019), 197–203.

[25] M. E. J. Newman. 2003. Mixing patterns in networks. *Physical Review E* 67, 2 (Feb 2003), 026126.

[26] Fan Ouyang and Cassandra Scharber. 2017. The influences of an experienced instructor's discussion design and facilitation on an online learning community development: A social network analysis study. *The Internet and Higher Education* 35 (Oct. 2017), 34–47. https://doi.org/10.1016/j.iheduc.2017.07.002 90.

[27] Kathleen Quardokus and Charles Henderson. 2015. Promoting instructional change: using social network analysis to understand the informal structure of academic departments. *Higher Education* 70, 3 (Sept. 2015), 315–335. https://doi.org/10.1007/s10734-014-9831-0 64.

[28] Bart Rienties, YingFei Héliot, and Divya Jindal-Snape. 2013. Understanding social learning relations of international students in a large classroom using social network analysis. *Higher Education* 66, 4 (Oct. 2013), 489–504. https://doi.org/10.1007/s10734-013-9617-9 149.

[29] Peter Shea, Suzanne Hayes, Sedef Uzuner-Smith, Mary Gozza-Cohen, Jason Vickers, and Temi Bidjerano. 2014. Reconceptualizing the community of inquiry framework: An exploratory analysis. *The Internet and Higher Education* 23 (Oct 2014), 9–17. https://doi.org/10.1016/j.iheduc.2014.05.002 151.

[30] Vincent Tinto. 1975. Dropout from higher education: A theoretical synthesis of recent research. *Review of educational research* 45, 1 (1975), 89–125.

[31] Alexandru Topîrceanu. 2017. Breaking up friendships in exams: A case study for minimizing student cheating in higher education using social network analysis. *Computers & Education* 115 (Dec 2017), 171–187. https://doi.org/10.1016/j.compedu.2017.08.008 20.

[32] Gloria Y. Washington. 2019. The Learning Management System Matters in Face-to-Face Higher Education Courses. *Journal of Educational Technology Systems* 48, 2 (Dec 2019), 255–275.

[33] Stanley Wasserman and Katherine Faust. 1994. *Social network analysis: methods and applications*. Cambridge University Press. 41197.

[34] Kim Weeden and Benjamin Cornwell. 2020. The Small-World Network of College Classes: Implications for Epidemic Spread on a University Campus. *Sociological Science* 7 (2020), 222–241.

[35] Alyssa Friend Wise and Yi Cui. 2018. Learning communities in the crowd: Characteristics of content related interactions and social relationships in MOOC discussion forums. *Computers & Education* 122 (Jul 2018), 221–242.

[36] Kui Xie, Gennaro Di Tosto, Lin Lu, and Young Suk Cho. 2018. Detecting leadership in peer-moderated online collaborative learning through text mining and social network analysis. *The Internet and Higher Education* 38 (July 2018), 9–17. https://doi.org/10.1016/j.iheduc.2018.04.002 42.

[37] Diyi Yang, Tanmay Sinha, David Adamson, and Carolyn Penstein Rosé. 2013. Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In *Proceedings of the 2013 NIPS Data-driven education workshop*, Vol. 11. 14.