Article

Journal of Educational Computing Research
2023, Vol. 61(2) 390–415

DOI: 10.1177/07356331221113613
journals.sagepub.com/home/jec

Mining Large Open Online Learning Networks: Exploring Community Dynamics and Communities of Performance

Wanli Xing[1] and Hanxiang Du[1]

Abstract

Online learning communities are becoming increasingly popular as they are known to support collaborative dialogue and knowledge building. Previous studies have typically focused on small, closed learning communities from an individual, static, and aggregated perspective. This research aims to advance our understanding of open and large online learning networks by exploring and characterizing community level dynamics and communities of performance. To achieve this goal, we mined a large open online learning network of over 30,000 students and approximately one million posts. First, we analyzed overall community network development by building temporal social networks. Subsequently, we studied sub-community dynamics using community detection algorithms, and, following that, investigated the interaction between community dynamics and communities of performance using best colleague correlation and Kruskal-Wallis test. Results found that large open online learning communities begin with a very large network having numerous small sub-communities. These communities consist of students who are similar in performance with strong links. The overall network size gradually shrinks, as does the number of sub-communities, and these communities evolve over time for their membership formulation with students who are more different in performance with weaker ties. Theoretical, practical, and methodological implications are then discussed. This study pushed the online learning

[1]School of Teaching & Learning, University of Florida, Gainesville, FL, USA

Corresponding Author:
Wanli Xing, School of Teaching & Learning, University of Florida, College of Education, Gainesville, FL 32611, USA.
Email: wanli.xing@coe.ufl.edu

community research to examine large and open networks by taking a more community-based and dynamic view of investigation.

Keywords
online learning communities, social network analysis, learning analytics, educational data mining, online learning

Introduction

There has been a considerably increased interest in recent years in the use of online communities for learning. With the prevalence of virtual learning environments, and "web 2.0" technologies, online communication is now widely used to support learning and build communication in various learning situations. McConnell (2006) defines online learning communities (OLCs) as cohesive communities that embody a culture of learning in which members are involved in a collective effort towards reaching an understanding of the material provided. This social phenomenon of the community on individual learning traces its roots to constructivism which proposing that knowledge is constructed within the social milieu (Cunningham, 1996). In this approach, learning is conceptualized as a participatory and social process in which a series of multistranded interpersonal transactions mediate knowledge exchange and construction (Engestrom, 1993). Online learning communities provide a learning atmosphere and a supportive system for collaborative dialogue and knowledge building by acquiring, generating, analyzing and structuring information (Carlen & Jobring, 2005; Studente, 2021, p. 273). It is important for us to understand and characterize OLCs so that we may respond and foster these communities effectively.

Previous studies have traditionally focused on small and closed OLCs. Their research has investigated OLCs embedded in online courses with a small number of students, e.g., online discussion forum (Lai, 2015; Mtshali et al., 2020), or an integrated social media like Twitter (Ayu et al., 2021; Peters et al., 2019). Oftentimes, participation in these OLCs is expected or required as part of the formal curriculum in a course. By comparison, open OLCs have usually been regarded as a natural environment for informal learning (Macia & García, 2016). Students with common interests that interact and work together for collaborative learning and production. Open OLCs must have a large number of participants to reach critical mass (Markus, 1987; Yao et al., 2021) and in order to be successful and sustainable. Rather than the required and/or expected steady participation in closed OLCs, the main learning paradigm for large open OLCs is legitimate peripheral participation (Lave & Wenger, 1991). As Nielsen (2006) explained in his 90:9:1 model, 90% of the participants only view content, 9% edit content, and 1% actively create new content. Much additional work is needed to better understand large open OLCs so that current education can be enriched by connecting schools with society and connecting formal with collaborative informal learning.

Methodologically, the documented research thus far has often applied qualitative methods such as observations, surveys, and qualitative content analysis of learning communities (Dille & Røkenes, 2021; Ke & Hoadley, 2009). The study is made possible nowadays by the automatic collection of electronic traces of OLC data, spanning a substantial number of diverse participants over extended time periods (Sundaram et al., 2012). Notably, this large-scale study can be conducted at a comparatively low cost, requiring little human supervision or manual work. One of the automatic data analytics methods popular in the OLC community is social network analysis (SNA). Current research and practice usually (1) focus on individuals, the interaction between network metrics and individual student learning outcomes (e.g., Xu et al., 2021; De-Marcos et al., 2016); (2) derive from a static perspective, overall network structure, and metrics (e.g., Cadima et al., 2012), rather than from a dynamic, community standpoint. However, OLCs, especially large open OLCs, are complex evolving social networks, they can be created or organically born, and they grow, shrink, split, merge, and disappear (Arslan et al., 2022; Chen et al., 2015; Dasgupta & Gupta, 2016; Morais et al., 2020). Connections between students are established and change over time. It is important to understand OLCs from a temporal perspective rather than solely from a static and aggregated angle. Also, while research centering on individuals in OLCs can support wide range of applications, e.g., personalized recommendations, community-level studies can allow for the identification of strategies that highlight the main properties of the network at a higher, collective, macro level that can facilitate scalability.

In this work we move from analyzing small and closed OLCs using an individual, static, and aggregated perspective to advance our understanding of large open OLCs by characterizing community-level dynamics and community of performance. Specifically, we collected and analyzed about one million forum posts from over 30,000 students in a large open online math learning community over a two-and-a-half-year period. To begin, we characterized the overall community development using SNA, and then examined the sub-community temporal dynamics using community detection algorithms, and last investigated the communities of performance by applying best colleague correlation and the Kruskal-Wallis (KW) test to reveal the interaction between community dynamics and communities of performance. Through systematic analysis, we generated new understandings regarding how large open OLCs develop over time from an overall network perspective, sub-community perspective, and communities of performance perspective. These community-level insights provided powerful theoretical, practical, and methodological implications for designing, fostering, and sustaining open, large OLCs.

Background

Theoretical Foundation

The enormous success of OLCs has attracted many researchers to explore the theories behind them. Communities of Practice (CoP), coined by Lave and Wenger (1991), are mentioned in numerous studies. From a CoP perspective, OLCs can be conceptualized as "a System between people, activities and the world; developing with time, and in relationship to other tangential and overlapping CoP." In a CoP, learning is participation and knowledge is constructed within the socialization process. Communities of Inquiry (CoI), developed by Garrison et al. (1999), also serve a guide for OLC practice and research, and inform methodologies and approaches to OLC design and delivery. Garrison (2017) conducted a comprehensive account of the research and developments in CoI frameworks, acknowledging the need for further exploration and validation for the structural investigation of CoIs as a social presence is the underlying presence in OLC. Social cognitive theory (Bandura, 1986), which argues that individual thought, affect, and action can be affected by observing others within the context of social interactions, is also used when studying OLCs. Social cognitive theory has been used to study how personal and environmental factors act as key drivers of loyalty behavior in OLCs (Lin, 2010). Further theoretical support for OLCs can be traced to collaborative learning (Chatterjee & Correia, 2020; Goda & Yamada, 2013), including enhancing motivation, learning achievement, and satisfaction.

As the current research and perspective about OLCs has emphasized more individual motivation, participation, and performance, we investigated the theoretical underpinnings in order for us to examine community-level dynamics. From a network science standpoint, the structural characteristics of an OLC manifest accumulated behaviors of individual leaners (Johnson et al., 2014). People do not act randomly in a community, but instead behave in response to shared motivations, practices, and tools that lead to the emergence of structural regularities (Monge et al., 2003). Community structure matters as it both shapes and reflects behavior (Orlikowski & Iacono, 2001). As both Holme and Saramaki (2019) and Barabási (2009) have suggested, network science must now gain an understanding of the processes that occur in community networks that shape their structure. Similarly, organizational researchers focus on identifying the mechanisms that drive network outcomes and illustrating the processes behind community emergence, dynamics, and evolution (Ahuja et al., 2002; Zaheer & Soda, 2009). Theoretically, this research calls for a shift from an individual and static view of OLCs to one that is community-based and dynamic.

Social Network Analysis

Social network analysis, often used in social learning analytics, is an interdisciplinary technique consisting of various quantitative analytics methods (Jan et al., 2019). It is performed on networks of relations between human or non-human entities such as documents and organizations. These different entities in a network are represented by nodes and their relations are represented by lines between the nodes, depicting them as a graph or network. Networks can be directed with arrowed lines connecting nodes to identify the starter and receiver of a relationship, and with the degree of weight, or thickness, of a line or an arrow indicating the strength of the relationship (Jan & Vlachopoulos, 2019). Social network analysis's methodological distinctness is due to its visual representation of data; its emphasis of relations between nodes as opposed to individual attributes; its examination of node activities according to the structure of the relational networks; its study of information flow between nodes; and its application at the individual (micro) and network (macro) level (Borgatti et al., 2013; Wasserman & Faust, 1994). Social network analysis has been widely used to study complex social interactions in various fields, for example, communication (e.g., Cho & Lee, 2008), healthcare (e.g., Baktha et al., 2017), engineering (e.g., Eissa et al., 2021), learning theories (e.g., Li et al., 2020) economics (Mathar & Gaur, 2020), and political science (Ward et al., 2011).

Social network analysis has been increasingly applied to the field of education (see review in Yassine et al., 2021). It includes but is not limited to: general engagement analysis for examining the emotional, cognitive, and behavioral interactions between learners and/or resources, behavior assessment which centers on diagnosing learner behaviors to identify any patterns of disorder that may need support and intervention, performance prediction to predict learners' achievement to aid in enhancing the teaching and learning processes, and recommender system development focusing on design filtration systems that provide personalized recommendations to assist learners based on their interest, preferences, and interaction patterns. These SNA studies in education span various contexts including learning management systems, Massive Open Online Courses (MOOCs), social learning environments, blended courses, intelligent tutoring systems, and game-based learning environments.

Empirical Research in OLCs

Many studies on OLCs have focused on the small and close communities embedded in certain courses. For instance, Kear et al. (2014) researched the role of personal profiles to enhance social presence in an OLC. They conducted two rounds of studies in online community contexts with 195 students and 29 students in confined course contexts by analyzing both quantitative questionnaire responses and qualitative answers to open questions. Gökçearslan and Alper (2015) studied the effects of locus of control on learners' sense of community and academic success in the context of OLCs consisting of 68 students in an online programming languages course for preservice teachers. They used various questionnaires such as sense of community and locus of control scale. They then examined the interactions of these instruments with student performance scores in the classroom. Abdelmalak (2015) conducted action research to explore student perspectives regarding using Web 2.0 technologies (e.g., Twitter, Google Docs, Wikis etc.) to develop a community of learners. Abdelmalak collected a variety of data from 25 graduate students in an online course including students' reflective journaling activities, researcher's field notes, and students posts and comments to understand student experiences and perspectives in the OLCs. Similar studies can also refer to Auyeung (2004); Vlachopoulos (2012); Lai (2015); DeKorver et al. (2020); Chen and Chen (2021); Wojcik et al. (2021); Chen and Swan (2020).

Given the rise of MOOCs and social media, some researchers have begun to explore large OLCs with less examination from a network perspective. For example, Huang et al. (2014) studied 44 MOOC forum communities with over 70,000 forum posts to understand superposter behavior primarily using descriptive statistics and forum content analysis. Cohen et al. (2019) analyzed one MOOC forum with 27,322 learners to explore and understand their participation patterns and factors correlated with student participation using text mining and statistical analysis. Xing and Gao (2018) analyzed over 600,000 Tweets from about 70,000 users in a learning community of professional teachers to understand their commitment in the large open OLC using text classification and survival analysis. Staudt Willet (2019) applied text mining to study over 1.2 million tweets in the same learning community to understand how and why educators use Twitter. Carpenter et al. (in press) analyzed over 2.6 million tweets in 15 education-related Twitter communities to understand the landscape of professional educator activity on Twitter using various statistical and data mining methods. A recent published study examined around half a million tweets and applied SNA to understand how communities were developed in informal online professional (Du et al., 2022).

Quite a few studies have employed SNA to examine OLCs. However, many of these studies examined small and/or closed communities. Jan et al. (2019) conducted a review of SNA studies of learning communities using CoP. Out of the 10 studies reviewed, nine had less than 100 students and the remaining study did not report the participant number. In addition, most studies have used similar social network measures to describe the structural properties in terms of shape and cohesion (e.g., centralization, density). These studies used students' network positions and properties to relate to their academic performance. A number of studies applied SNA to MOOC forums, which tend to have a larger number of participants (Boroujeni et al., 2017; Han et al., 2021; Liu et al., 2021; Poquet et al., 2018). Nevertheless, MOOCs, in essence, are still closed, course-based OLCs and have pre-existing structures and requirements for these communities as compared to large open online communities. More studies are required to reveal the community dynamics and community of performance for large open OLCs from a temporal perspective.

Summary and RQ

Previous studies have mostly centered on small and/or closed OLCs and have examined individual behavior and participation. These studies have often applied qualitative and text analysis methods. For those that did use SNA, static metrics were used instead of looking to understand communities from a dynamic perspective. Examining open and large OLCs from a dynamic network perspective can potentially advance our theoretical understanding of learning communities and produce new methodological paradigms to study these communities. The following research questions guided our study:

(1) How do open online learning networks develop over time as a whole?
(2) How do open online learning networks develop over time from a sub-community perspective?
(3) How do open online learning network dynamics interact with communities of performance?

Methods

Research Context and Data Collection

Algebra Nation (AN) is a dynamic online platform that helps students master Algebra – the gateway math course that has implications for students' success in secondary school and beyond. AN provides 24-hr access to high-quality instructional videos, workbooks, collaborative learning tools, discussion forums, and adaptive assessments and support for hundreds of thousands of students across six states who utilize the platform every year. We collected data by querying the database of AN. We extracted posts and threads published from Fall semester of 2017 to Spring semester of 2019, created by 31,343 students from 2771 Florida schools. We set this date criterion with the aim of examining students' recent discussion forum activities before school closures due to COVID-19. Meanwhile, the platform does not require users to self-report gender information, leaving the participants' demographic information unknown. We operationalized a post as top-level discussion that can initiate continuous discussions and a thread as the response(s) to a post. Students can view and comment on peers' posts from the same course regardless of their region or school. All students of Algebra I, though from different schools, have access to other students' discussion interactions in Algebra I. Student assessments were also extracted from the database. Each assessment included 10 assessment items and was designed to evaluate student learning in a specific learning module. Data collection received Institutional Review Board approval from the authors' institute.

Data Preprocessing

We developed networks based on post-reply activities in the virtual discussion forum. The nodes of the network represent students, and the edges represent discussion forum interactions between students. In this study, we consider a reply to a comment as the explicit indicator of interactions between two students. If two students interacted more than once, multiple edges were drawn between the two students. In other words, the edge is weighted. A student learning procedure was partitioned every 2–3 months to examine the evolving interaction networks across the course. To further investigate the extent to which the differences in the network dynamics interact with communities of

Table 1. Number of Posts and Threads.

Section	Number of posts	Number of threads
2017C	115,292	14,218
2017D	35,015	3685
2018A	81,785	9478
2018B	65,012	7952
2018C	45,707	5411
2018D	73,081	7288
2019A	55,758	5718

performance, we examined the networks on the dataset with all students who registered the course and only students with valid performance on course modules, respectively. Specifically, we developed two types of social networks, one with all students who had ever posted or replied to a comment between September 01, 2017 and March 31, 2019, and one for students with valid performance. We considered students who participated in at least one module of the course to be students with valid performance.

For easy reference, we used session indexes to represent the seven periods from September 01, 2017 to March 31, 2019. We use the letters A, B, C, D to represent the four periods of a year in order. In other words, 2018A is the first 3-month period in year 2018 while 2018D is the final period. The number of posts and threads of the data set can be found in Table 1. The number of posts for each period lies between 35,015 and a maximum of 115,292.

Data Analysis

Figure 1 provides an overview of the methodology of our study. Using AN discussion forum posts, we first apply SNA methods to develop social networks for each period to understand the dynamics at the social network level, then conduct community detection to uncover meaningful communities within each network to examine sub-community dynamics, and finally examine the interaction between network dynamics and communities of performance using social best colleague correlation and the KW test. All analysis was conducted using Python 2.7.

Social Network Analysis. A social network represents the relationship among a group of actors via nodes and edges, where each node represents one actor and each edge connecting two nodes indicates a certain type of relationship. In this study, we developed seven social platform networks based on students' posts, each using a 2- to 3-month period of data. In each network, a single student is considered to be one node. If any two students have directly communicated with each other once, we create a pair of connected nodes to represent the students by adding an edge between them. If two students have communicated more than once, more edges are added. In the context of

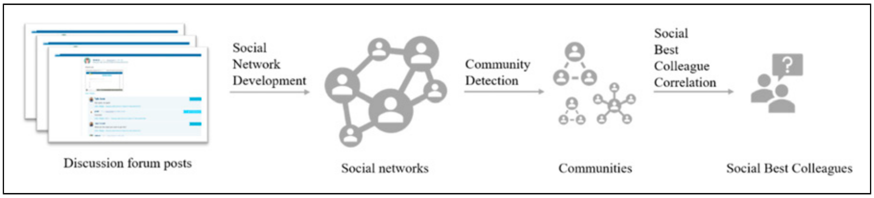

Figure 1. Methodology overview.

social network graphs, such a network graph—which allows multiple edges among two nodes—is known as a weighted graph. If a student only posted a comment yet never replied to others', the student will become an isolated node in the social network graph. If a student did not actively create a thread and only replied to others' comments, the student will be presented as a connected node in the social network graph. In a weighted social network graph, weighted edges indicate the level of connection. In our case, the weight of an edge represents the number of communications between two connected students. Python *networkx* is a powerful package to help create and study large-scale networks.

Community Detection. Community detection describes the process of identifying meaningful subgroups within a complex network structure, while each subgroup is considered to be a community. Community, which demonstrates the tendency for a group of members to cluster, is a useful structure to study in SNA (Haythornthwaite, 2009). In a social network, communities consist of individuals who have a relationship to each other (Girvan & Newman, 2002). Communities are seen as meaningful subgroups in a massive learning network, and differ from one another (e.g., Brown, Lynch, Eagle, et al., 2015; Xu et al., 2018). Community detection could provide the dynamic view of how student groups communicate over time (Xu et al., 2018). We conducted community detection in the AN platform course network to examine how student communities evolve over time.

This study applies Clauset-Newman-Moore greedy modularity maximization—a commonly used modularity-based detection algorithm—to detect communities in each course social network (Clauset et al., 2004). Research indicates that this algorithm can detect meaningful and reliable communities in large social networks with thousands or even millions of actors (Camacho et al., 2020; Clauset et al., 2004). Modularity is a parameter for measuring the cleanness and strength of the division in community detection, lying between -1 and 1. A high modularity indicates that members of the same community have close connections with each other and weak connections with members of other communities. Initially, the algorithm assumes that each node belongs to its own community. The algorithm then merges the pair of nodes which increase modularity to the greatest degree. The algorithm will repeat the merging step until only one community remains and finally, choose the best partition results which have the maximum modularity.

Social Best Colleague Correlation. Previous studies found a positive relationship between student academic performance and the academic performance of their close friends in both traditional classrooms and MOOCs (Brown, Lynch, Wang, et al., 2015; Fire et al., 2012). Fire et al. (2012) developed a social network based on students' interactions in a traditional classroom and analyzed a series of features extracted from the classroom social networks. They found a high positive relationship between the grade of a student and their best friends' grades. Brown, Lynch, Wang, et al. (2015) used discussion forum interactions to identify a student's social best friend and reported a similar positive relationship between students' grades in a MOOC.

We are also looking to discover whether such a relationship exists on the AN learning network and whether this relationship changes over time. Based on students' interaction frequency on discussion forums, we were able to identify their social "best friend", which we call "social best colleague." Interaction frequency is the number of interactions on discussion forums between any two students. We consider the reply to a comment to be the explicit indicator of an interaction. In other words, a student's social best colleague is their peer who has interacted with this student the most through discussion forums. Mathematically, the social best colleague *BC* for a student *ST* is their most connected neighbor, $neighbor_i$ with a score s_i, where *ST* has a collection of n neighbors $\{neighbor_1, neighbor_2, \ldots, neighbor_n\}$ with a score $s_1, s_2, \ldots, s_n$, respectively, for $n \in \boldsymbol{N}^*$ and $1 \leq i \leq n$:

$social\text{-}best\text{-}colleague\ (ST) = neighbor_i$, where the subscript i:

$i = m$, where m is the subscript of the highest score k_m, and k_m:

$k_m = max\{s_1, s_2, \ldots, s_n\}$.

The score s_j for a participant's $neighbor_j$ is defined as the degree, or, in other words, the number of interactions between them. In the case that a student has more than one most connected peers, we randomly choose one of the peers as the student's social best colleague. Then, we perform a correlation to examine the relationship between the student's grades and their social best colleague's grades for each network. To further investigate the network communities and communities of performance, we applied the KW test to assess the interaction of network dynamics and communities of performance over time. The KW test by ranks is a non-parametric method to test whether samples originate from the same distribution. Let $n_i (i = 1, 2, \ldots, k)$ be the sample sizes for each of the k groups in the data. Then, rank the combined sample and compute R_i, the sum of the ranks for group i. The KW test statistic is

$$\mathrm{H} = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1)$$

Table 2. Number of Edges and Nodes for Seven Social Networks.

	Number of edges		Number of nodes	
Course social network	*ALL*	*Non_Zero*	*ALL*	*Non_Zero*
2017C	56,830	26,426	10,128	5276
2017D	13,833	6330	2759	1476
2018A	28,150	13,789	5707	3135
2018B	18,727	11,282	4020	2575
2018C	15,472	10,231	3266	2213
2018D	21,942	11,940	3169	1943
2019A	10,527	7146	2123	1381

Results

Community Network Dynamics for RQ1

We developed two types of networks: *ALL* course social networks consist of all students as long as they have posted a comment on discussion forums within a period of time, while *Non_Zero* course social networks remove those students who posted but did not have a valid performance score within that period. The number of edges and nodes for seven social networks are shown in Table 2. For both *ALL* and *Non_Zero* course social networks, 2017C has the largest number of students and communications. After removing students who did not have a valid score, the course social network shrinks. Specifically, both the number of nodes and the number of edges decrease by 30–50%. This change indicates a large number of students who are active and vibrant in the AN platform at the beginning and a decrease in these activities over time.

To assess how the social network changes over time overall, we calculate three types of social network dynamics: Network+, Network-, and Network=, representing the number of students who joined the network, who left the network, and who remained in the network, respectively. Figure 2 visualizes the social network dynamics between 2018D and 2019A. In the two social network graphs, each node represents a single student. The node size scales to the interaction frequency between a student and their peers: the larger the node, the more interaction exists. The weight of an edge stands for the interaction frequency between a pair of nodes: the thicker the edge, the more interaction between the two students. Meanwhile, we use colors to visualize the social network dynamics: green nodes are students who remain in both social networks (Network=), blue nodes in Figure 2(a) are students who leave during 2019A, while orange nodes in Figure 2(b) are students who join the network during 2019A. While most of the large-size nodes in Figure 2(a) are green, several are blue. This suggests that students who were active in discussion forum of the AN platform in 2018D became less active in 2019A.

The results of social network dynamics are shown in Table 3 (*ALL*) and Table 4 (*Non_Zero*). Figure 3 visualizes the trends of new participants for both *ALL* and

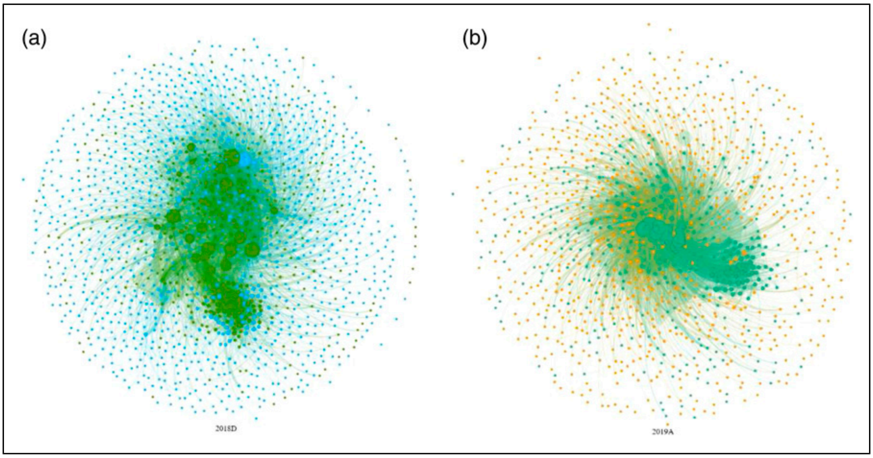

Figure 2. Social network dynamics between (a) 2018D and (b) 2019A.

Table 3. Network Dynamics for *ALL* Course Social Networks Over the Years.

Interval	Network+	Network-	Network=
2017C – 2017D	1391	8760	1368
2017D – 2018A	4788	1840	919
2018A – 2018B	2883	4570	1137
2018B – 2018C	3164	3918	102
2018C – 2018D	2285	2382	884
2018D – 2019A	714	2455	1409

Table 4. Network Dynamics for *Non_Zero* Course Social Networks Over the Years.

Interval	Network+	Network-	Network=
2017C – 2017D	640	4440	836
2017D – 2018A	2583	924	552
2018A – 2018B	1821	2381	754
2018B – 2018C	2135	2497	78
2018C – 2018D	1308	1578	635
2018D – 2019A	863	1425	518

Non_Zero networks. The results show a high volume of dynamics over time. Taking 2017C–2017D (see Table 3) as an example, 1391 students who did not engage discussion forum interaction in 2017C joined the course social network in 2017D; 8760 students who engaged discussion forum interaction in 2017C left the course social network in 2017D; 1368 students were present in both 2017C and 2017D course social

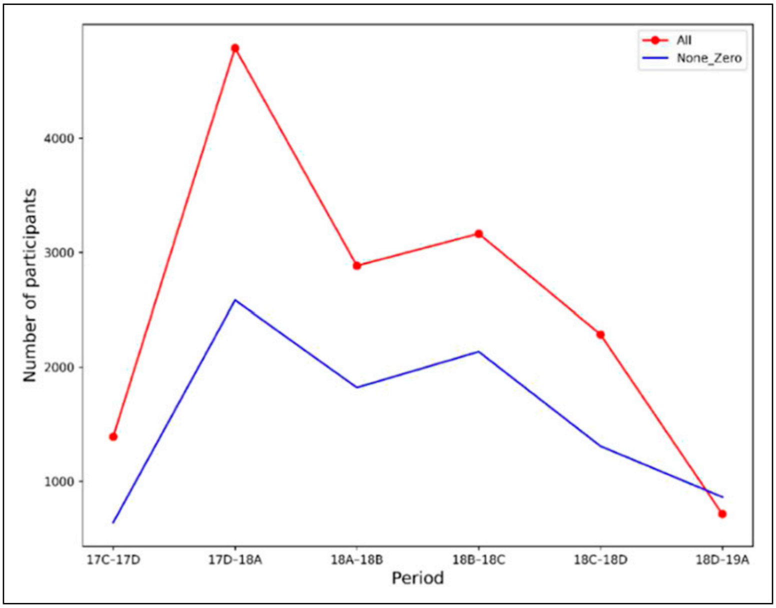

Figure 3. New participating students by period.

networks. We observe an overall decrease in both the total number of posted comments and the total number of students who posted a comment over time regardless of network type (see Figure 3). For both *ALL* course social networks, Network+ is sometimes higher than Network- and sometimes lower than Network- (Table 3). However, when we remove those students who did not complete any assessments, Network- is higher for most years (Table 4). We saw the high dynamics at the beginning of this large open online learning network and such network, and its dynamics decreased over time. Given that the AN forum is mainly a question-and-answer community focusing on K-12 algebra, we suspect this large and open online learning community may eventually evolve into a stable knowledge repository as most of the questions may already be answered and the need for new posts decreases over time.

Sub-Community Dynamics for RQ2

To examine how sub-communities were developed on the AN community over time, we applied Clauset-Newman-Moore greedy modularity maximization to detect communities in each social network. Table 5 shows the community detection results for both *ALL* and *Non_Zero* social networks. During the majority of the time period, *ALL* social networks have a higher number of detected communities than *Non_Zero* social networks, while *Non_Zero* social networks have more students in each detected community. For two periods (2018A and 2018B), *Non_Zero* social networks have a

Table 5. Number of Detected Communities Over Time.

	Number of communities	
Course social network	All	Non_Zero
2017C	37	11
2017D	13	7
2018A	7	9
2018B	6	7
2018C	9	4
2018D	4	2
2019A	2	2

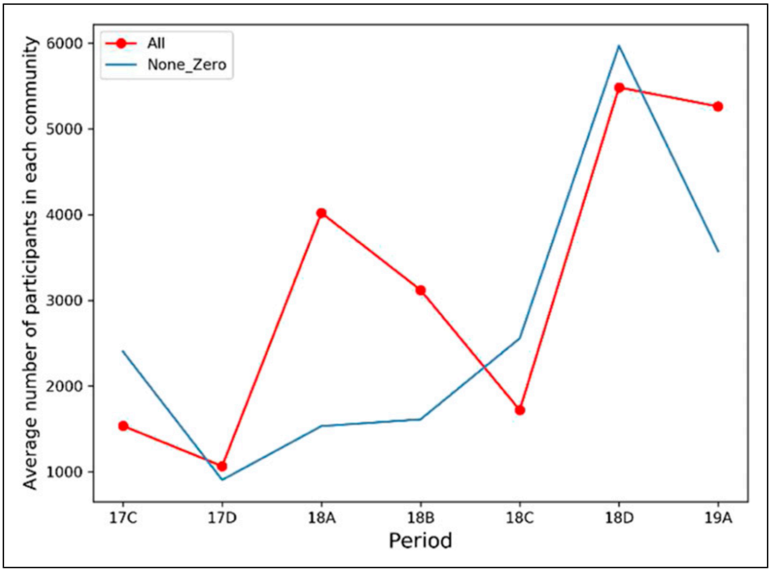

Figure 4. The average number of participants in each detected community over the years.

higher community number. Figure 4 shows the average number of students in each community during the entire time period assessed. For the *ALL* social network, the average number of students in each community ranges between 1000 and 5,500, while the value for *Non_Zero* course social networks range between 800 and 6000. In addition to the strong dynamics of both types of course social networks, we also observe an overall increase in the average number of students in each community (see Figure 4), indicating that students developed larger and larger communities over the 2 years.

Table 6. Community Dynamics for *ALL* Course Social Networks Over 2 Years.

Interval	Cum_Avg+	Cum_Avg-	Cum_Avg=
2017C – 2017D	1538.89	5221.99	619.53
2017D – 2018A	4698.78	1470.93	754.58
2018A – 2018B	2527.36	4571.76	944.05
2018B – 2018C	2558.35	3654.13	87.0
2018C – 2018D	2555.10	1739.24	603.32
2018D – 2019A	1403.0	2448.0	714.0

Similar to network dynamics, we also calculate the dynamics at community level to examine how the community changes over time. As there are multiple communities for each course social network, we calculate Cum_Avg+, Cum_Avg-, and Cum_Avg+, representing the average number of students who join the community, the average number of students who left the community, and the average number of students who remained for each student. The community dynamic results are shown in Table 6 (*ALL*) and Table 7 (*Non_Zero*). Taking 2017C – 2017D from Table 6 as an example, for each student, an average number of 1538.89 students who were not in the same community in 2017C joined their communities in 2017D, an average number of 5221.99 students who shared the same community in 2017C left the community in 2017D, and an average number of 619.52 students were in the same community with the student for both 2017C and 2017D.

Overall, we also observed high sub-community dynamics for both *ALL* and *Non_Zero* social networks over time. Specifically, Cum_Avg+ or Cum_Avg- is at least two times of Cum_Avg= over the years. Meanwhile, Cum_Avg+ is higher than Cum_Avg- for some periods, while Cum_Avg- is higher than Cum_Avg+ for some other periods. We believe the results of the network dynamics and community dynamics indicate that students actively interact with peers across the AN platform, developing communities and changing community membership quite often. The trend of decreasing number of sub-communities and community dynamics is in line with our hypothesis from the overall network analysis that this large and open online learning network in AN becomes a huge knowledge base covering an increasing number of questions and topics, and in turn requires fewer new postings over time.

Community Dynamics and Communities of Performance for RQ3

Previous studies indicate that students' grades are positively correlated with the grades of their closest peers in both traditional courses and MOOCs (Brown, Lynch, Wang, et al., 2015; Fire et al., 2012). We also wonder as to whether such a relationship exists in the AN platform when students interact with each other across curriculums. As mentioned previously, a large number of students use the AN platform to interact with peers only. These students do not have a valid performance score and yet could

Table 7. Community Dynamics for *Non_Zero* Course Social Networks Over 2 Years.

Interval	Cum_Avg+	Cum_Avg-	Cum_Avg=
2017C – 2017D	670.32	2664.03	306.29
2017D – 2018A	2705.64	639.10	396.36
2018A – 2018B	1790.62	2347.88	750.01
2018B – 2018C	1791.85	2442.92	68.54
2018C – 2018D	1492.93	1196.10	445.02
2018D – 2019A	858.0	1423.0	518.0

Table 8. Social Best Colleague Correlation Coefficients for *ALL* and *Non_Zero* Course Social Networks Over Time.

Course social network	All	Non_Zero
2017C	.11**	.06**
2017D	.09**	.09**
2018A	.10**	.04*
2018B	.03	.13
2018C	.04	.02
2018D	.08**	.08**
2019A	.05	.03

**< .01 *: < .05.

potentially become others' social best colleague due to their active participation in discussion forums. As it does not make sense to calculate the correlation between two students' grades when one does not obtain a grade, we only included students with a valid performance score during the calculation. A Spearman-Pearson correlation is performed to examine the relationship between a student's grade and their social best colleague's grade for both *ALL* and *Non_Zero* course social networks, as shown in Table 8.

For the starting time, students' grades are statistically significantly positively correlated with the grades of their social best colleague in both *ALL* and *Non_Zero* course social networks. However, we did not observe such a significant and strong correlation in 2018B, 2018C, and 2019A, potentially indicating a different pattern of student performance and interaction behaviors on discussion forums of the AN platform at subsequent timeframes. This is somewhat different from previous studies, which stated a positive correlation of performance between students and their closest friends (e.g., Brown, Lynch, Wang, et al., 2015; Fire et al., 2012). This analysis showed that students interact more with those are similar intensively and that this interaction becomes weaker over time when students communicate with those who are dissimilar.

To investigate the interaction of sub-community dynamics and communities of performance, we further applied the KW test to assess the differences in student

Table 9. The KW Test by Ranks Coefficients for *ALL* and *Non_Zero* Course Social Networks Over Time.

Course social network	All	Non_Zero
2017C	118.42**	26.33**
2017D	36.35**	66.34**
2018A	35.78**	8.38
2018B	24.88**	5.80
2018C	4.02	29.77**
2018D	1.36	.29
2019A	3.66	.05

Note. KW = Kruskal-Wallis.
**: < .01.

performance within each community over time. Results are shown in Table 9 which demonstrate some notable patterns. First, we observe statistically significant differences on student performance among communities for four out of seven *ALL* course social networks (from 2017C to 2018B), as well as three out of seven *Non_Zero* course social networks (2017C, 2017D, and 2018C). The coefficient becomes smaller over time generally. Second, the effect of removing students without valid performance on the test is mixed, leading to a transition from significant to non-significant (2018A and 2018B), from non-significant to significant (2018C), or no transition (2017C, 2017D, 2018D, and 2019A). This result is in line with the above social best colleague correlation in which students tend to interact intensively with those who have similar performance first and formulate a community and then interact with those who are different in performance over time in a less intensive way.

Discussion

This study, to our best knowledge, is the first of its kind to explore and characterize large open OLCs from a dynamic and community perspective. In comparison, most previous studies have centered on small and closed OLCs and many of them have used qualitative analysis and relied on text analysis methods (Dille & Røkenes, 2021; Ke & Hoadley, 2009). While some studies have examined large OLCs such as MOOC forums (Boroujeni et al., 2017; Cohen et al., 2019; Han et al., 2021), they are still learning communities in a closed form. A few exploratory studies did investigate large open OLCs such as Twitter-based teacher professional development networks (Carpenter et al., in press; Staudt Willet, 2019; Xing & Gao, 2018). These studies focus mainly on their discussion content instead of community formulation and development. There are many network studies about OLCs in the field of education (Jan et al., 2019). But their studies oftentimes are from an individual, static, and accumulated view of the social networks. Our study presents a dynamic network view of large open OLCs and the results obtained advanced our understanding of OLCs.

As an exploratory study, the findings from our research are compelling. The large open learning community as a whole starts with a very large network with numerous small sub-communities. The network size shrinks gradually over time as does the number of sub-communities. In the meantime, the sub-community size becomes much larger to the extent that the whole network evolved into one and/or two large, stable sub-communities. This may reflect that large open OLCs may eventually become a knowledge repository as suggested in Houda et al. (2019). One main source of knowledge lies in social networks because they contain human experiences and information. Given the context we have as a K-12 Algebra Learning community, most, if not all, of the questions related to K-12 Algebra learning may have been raised and answered. Therefore, many students may just view and check OLC posts to find their answers without having to start a new thread for their questions. This hypothesis is also supported by the overall trend of dynamics decrease for both the whole networks and community networks.

Another finding worth noticing is related to communities of performance. By comparing the whole network and non-zero networks (removing students without valid performance scores), we found that network dynamics are usually much more stable for both the whole network results (Table 3 vs. Table 4) and sub-community network results (Table 6 vs. Table 7) – proportionally, the number of students joining, remaining, or leaving a community are much smaller for non-zero networks as compared to the entire network. Also, we found that students tend to initially bond and form a community with those having similar performance and will gradually interact with those who are not similar. This means that an OLC is not always constituted of members with similar characteristics, which counters the popular and static view of students tending to communicate with other students who are similar to them (Brown, Lynch, Wang, et al., 2015; Fire et al., 2012). Instead, OLCs evolve over time in their community formulation and student communications.

Methodologically, this study pushed the OLCs field to expand from individual and static analysis of social learning networks to take a more community and dynamic level analysis. Previous studies have overwhelmingly focused on individual-level analysis, e.g., examining individual participation, network position and interaction with their performance, and on static overall network characteristics, network centrality and cohesion (Jan & Vlachopoulos, 2019). Our research studied the dynamic development of the OLC from an overall network perspective, a sub-community perspective, and examined how members dynamically shifted between those communities. In addition, using a community detection algorithm and best colleague correlation, we were able to characterize the network dynamics from community of performance perspective which can be easily used by other researchers aiming to do so. These methodological approaches can be easily adopted by researchers who are interested in this area and social networks in general.

This research has both theoretical and practical implications. Theoretically, it called for more studies of dynamic structural characteristics of OLCs to understand how community structure shapes and reflects participant behavior and achievement as well as investigating the process that takes place in the OLCs that leads to the emergence of structural regularities. In this way, our basic understanding of OLCs from a network science perspective can be advanced. Practically, the findings of this study can aid forum designers in developing network-level strategies to foster community dynamics, development, and sustainability. For instance, given that both the overall network and sub-community network are much more vibrant early on, forum designers can provide more incentives at the initial stages of community formation rather than later to spur the accumulation of topics and knowledge for the community. In this way, the OLCs may quickly evolve into a more stable community as an open source repository. Also, given the best colleague regression, students are most likely to connect with peers with similar performance early on and then with peers with performance more different than their own later, when we design intelligent peer recommenders for the open OLC (for Question & Answering or promoting peer support), we can consider recommending peers with closest performance early on. As they develop into a more stable sub-community, this intelligent recommender can suggest more peers (as ties become weaker) with more difference performance. Similarly, as students tend to sub-communicate more with those who are similar in terms of performance earlier on and later with those who are dissimilar, forum designers can create corresponding strategies to foster these communications in the community over time. That is, forum facilitators/forum designers/teachers can provide similar information, content, and incentives early on when students are similar in terms of their performance but provide more choices to students as they evolve into stable sub-communities catering the different performance needs of the communities.

There are several limitations related to this study. First, our study only examined one specific large open OLC. The AN forum primarily focuses on questions and answers, but many other learning forums and communities are mixed with social and information exchange. Therefore, the findings of this study may not be able to generalize to other learning communities. Second, this study revealed many dynamic network characteristics for large open OLCs. However, there may be outer factors that influence the network dynamics and development, for example, AN maintenance/dysfunction, school closure, etc. which may influence analysis results. Third, the significance of identified correlations could potentially be affected by the large sample size of this study, as suggested by Sullivan and Feinn (2012). Fourth, communities of performance are probably oversimplified as they are measured by student scores on the AN platform. Performance is a complex construct and scores on a certain platform in itself may not reflect a holistic view of a student's performance.

Conclusion

This study examined a large open OLCs from a dynamic network perspective and generated new knowledge in order to achieve a better understanding of learning communities. The results revealed various temporal network characteristics related to how large open OLCs develop and evolve from the overall network, sub-community, and communities of performance perspectives. Our research hopes to foster additional OLC research by taking a dynamic and community network view. There are several directions for future research: First, the learning network in this work was primarily constructed as a result of posting behavior. Future studies can build the social network by incorporating the posting content and even the viewing behavior (e.g., incorporating eye-tracking data) and further examine the network dynamics. Second, researchers can also conduct qualitative studies to gain an understanding of why students communicate with others in a certain way to formulate strong ties and communities and why such communication changes over time. Third, our work only examined one large open OLC. Other studies can apply similar methods to other settings and examine the transferability and generalizability of the findings.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported by the Institute of Education Sciences, US Department of Education, through Grant R305C160004 to the University of Florida, the University of Florida AI Catalyst Grant -P0195022, and the University of Florida Informatics Institute Seed Funding. The opinions expressed are those of the authors and do not represent the views of the University of Florida, Institute of Education Sciences, or those of the US Department of Education.

ORCID iDs

Wanli Xing https://orcid.org/0000-0002-1446-889X
Hanxiang Du https://orcid.org/0000-0002-9081-0706

References

Abdelmalak, M. M. M. (2015). Web 2.0 technologies and building online learning communities: Students' perspectives. *Online Learning*, *19*(2), n2. https://doi.org/10.24059/olj.v19i2.413

Ahuja, M., Chudoba, K. M., George, J. F., Kacmar, C., & McKnight, H. (2002, January). Overworked and isolated? Predicting the effect of work-family conflict, autonomy, and workload on organizational commitment and turnover of virtual workers. In Proceedings of the 35th Annual Hawaii International Conference on System Sciences, Big Island, HI, USA, 10 January 2002 (pp. 3586–3593). IEEE.

Arslan, O., Xing, W., Inan, F. A., & Du, H. (2022). Understanding topic duration in Twitter learning communities using data mining. Journal of Computer Assisted Learning, *38*(2), 513–525.

Auyeung, L. H. (2004). Building a collaborative online learning community: A case study in Hong Kong. *Journal of Educational Computing Research*, *31*(2), 119–136. https://doi.org/10.2190/ycm8-xkdy-qwn2-gpeh

Ayu, D. P., Raja, P., & Sholihah, L. (2021). The effect of online learning through Twitter on students' writing. *Unila Journal of English Teaching (U-JET)*, *10*(1), 96–106. https://doi.org/10.23960/UJET.v10.i1.202112

Baktha, K., Dev, M., Gupta, H., Agarwal, A., & Balamurugan, B. (2017). Social network analysis in healthcare. In *Internet of things and big data technologies for next generation healthcare* (pp. 309–334). Springer.

Bandura, A. (1986). *Social foundations of thought and action*. Prentice Hall.

Barabási, A. L. (2009). Scale-free networks: A decade and beyond. *Science*, *325*(5939), 412–413. https://doi.org/10.1126/science.1173299

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). *Analyzing social networks* (Kindle Edition). SAGE Publications.

Boroujeni, M. S., Hecking, T., Hoppe, H. U., & Dillenbourg, P. (2017, March). Dynamics of MOOC discussion forums. In Proceedings of the seventh international learning analytics & knowledge conference, Vancouver, Canada, 13 March, 2017 (pp. 128–137).

Brown, R., Lynch, C., Wang, Y., Eagle, M., Albert, J., Barnes, T., Baker, R., Bergner, Y., & McNamara, D. (2015). Communities of performance & communities of preference. In CEUR Workshop Proceedings, *1446*.

Brown, R., Lynch, C. F., Eagle, M., Albert, J., Barnes, T., Baker, R., Bergner, Y., & Mcnamara, D. (2015). Good communities and bad communities: Does membership affect performance? *Proceedings of Educational Data Mining*, 612–613.

Cadima, R., Ojeda Rodríguez, J., & Monguet Fierro, J. M. (2012). Social networks and performance in distributed learning communities. *Educational Technology and Society*, *15*(4), 296–304. Retrieved from https://www.jstor.org/stable/jeductechsoci.15.4.296

Camacho, D., Panizo-LLedot, A., Bello-Orgaz, G., Gonzalez-Pardo, A., & Cambria, E. (2020). The four dimensions of social network analysis: An overview of research methods, applications, and software tools. *Information Fusion*, *63*(1), 88–120. https://doi.org/10.1016/j.inffus.2020.05.009

Carlen, U., & Jobring, O. (2005). The rationale of online learning communities. *International Journal of Web Based Communities*, *1*(3), 272–295. https://doi.org/10.1504/ijwbc.2005.006927

Carpenter, J., Tani, T., Morrison, S., & Keane, J. (in press). Exploring the landscape of educator professional activity on Twitter: An analysis of 16 education-related Twitter hashtags. *Professional Development in Education*, 1–22. https://doi.org/10.1080/19415257.2020.1752287

Chatterjee, R., & Correia, A. P. (2020). Online students' attitudes toward collaborative learning and sense of community. *American Journal of Distance Education*, *34*(1), 53–68. https://doi.org/10.1080/08923647.2020.1703479

Chen, B., Chen, X., & Xing, W. (2015, March). "Twitter Archeology" of learning analytics and knowledge conferences. In Proceedings of the Fifth International Conference on learning analytics and knowledge (pp. 340-349). March 16–20, New York, USA.

Chen, C., & Chen, M. (2021). The research of social network analysis on college students' interactive relations. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, *15*(2), 49–59. https://doi.org/10.4018/ijcini.20210401.oa5

Chen, C. C., & Swan, K. (2020). Using innovative and scientifically-based debate to build e-learning community. *Online Learning*, *24*(3), 67–80. https://doi.org/10.24059/olj.v24i3.2345

Cho, H., & Lee, J. S. (2008). Collaborative information seeking in intercultural computer-mediated communication groups: Testing the influence of social context using social network analysis. *Communication Research*, *35*(4), 548–573. https://doi.org/10.1177/0093650208315982

Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *70*(6), 066111. https://doi.org/10.1103/PhysRevE.70.066111

Cohen, A., Shimony, U., Nachmias, R., & Soffer, T. (2019). Active learners' characterization in MOOC forums and their generated knowledge. *British Journal of Educational Technology*, *50*(1), 177–198. https://doi.org/10.1111/bjet.12670

Cunningham, D. J (1996). Time after time. In W. Spinks (Ed.), *Semiotics 95* (pp. 263–269). Lang Publishing.

Dasgupta, S., & Gupta, A. (2016). Analyzing community dynamics in social media. In Proceeding of the 1st VLDB Workshop on Social Data Analytics and Management (SODAM), Delhi, Sep. 9th.

DeKorver, B., Chaney, A., & Herrington, D. (2020). Strategies for teaching chemistry online: A content analysis of a chemistry instruction online learning community during the time of COVID-19. *Journal of Chemical Education*, *97*(9), 2825–2833. https://doi.org/10.1021/acs.jchemed.0c00783

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

Dille, K. B., & Røkenes, F. M. (2021). Teachers' professional development in formal online communities: A scoping review. *Teaching and Teacher Education*, *105*, 103431. https://doi.org/10.1016/j.tate.2021.103431

Du, H., Xing, W., & Zhu, G. (in press). Mining teacher informal online learning networks: Insights from massive educational chat tweets. *Journal of Educational Computing Research*. https://doi.org/10.1177/07356331221103764

Eissa, R., Eid, M. S., & Elbeltagi, E. (2021). Current applications of game theory in construction engineering and management research: A social network analysis approach. *Journal of Construction Engineering and Management*, *147*(7), 04021066. https://doi.org/10.1061/(asce)co.1943-7862.0002085

Engeström, Y. (1993). Developmental studies of work as a testbench of activity theory: Analyzing the work of general practitioners. In S. Chaiklin & J. Lave (Eds), *Understanding practice: Perspectives on activity and context* (pp. 64–103). Cambridge University Press.

Fire, M., Katz, G., Elovici, Y., Shapira, B., & Rokach, L. (2012). Predicting student exam's scores by analyzing social network data. *Lecture notes in computer science (Including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)* (Vol. 7669). LNCS. https://doi.org/10.1007/978-3-642-35236-2_59

Garrison, D. R. (2017). *E-Learning in the 21st century*. Routledge.

Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, *2*(2–3), 87–105. https://doi.org/10.1016/s1096-7516(00)00016-6

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(12), 7821–7826. https://doi.org/10.1073/pnas.122653799

Goda, Y., & Yamada, M. (2013). Application of CoI to design CSCL for EFL online asynchronous discussion. In *Educational communities of inquiry: Theoretical framework, research and practice* (pp. 295–316). IGI Global.

Gökçearslan, Ş., & Alper, A. (2015). The effect of locus of control on learners' sense of community and academic success in the context of online learning communities. *The Internet and Higher Education*, *27*, 64–73. https://doi.org/10.1016/j.iheduc.2015.06.003

Han, Z. M., Huang, C. Q., Yu, J. H., & Tsai, C. C. (2021). Identifying patterns of epistemic emotions with respect to interactions in massive online open courses using deep learning and social network analysis. *Computers in Human Behavior*, *122*, 106843. https://doi.org/10.1016/j.chb.2021.106843

Haythornthwaite, C. (2009). Building social networks via computer networks: Creating and sustaining distributed learning communities. *Building virtual communities*. https://doi.org/10.1017/cbo9780511606373.011

Holme, P., & Saramäki, J. (Eds.), (2019). *Temporal network theory* (Vol. 2). Springer.

Houda, S., Naila, A., & Samir, B. (2019). Knowledge management and reuse in virtual learning communities. *International Journal of Emerging Technologies in Learning*, *14*(16), 23. https://doi.org/10.3991/ijet.v14i16.10588

Huang, J., Dasgupta, A., Ghosh, A., Manning, J., & Sanders, M. (2014, March). Superposter behavior in MOOC forums. In Proceedings of the first ACM conference on Learning@ scale conference, March 4-5, Atlanta, GA, USA (pp. 117–126).

Jan, S. K., & Vlachopoulos, P. (2019). Social network analysis: A framework for identifying communities in higher education online learning. *Technology, Knowledge and Learning*, *24*(4), 621–639. https://doi.org/10.1007/s10758-018-9375-y

Jan, S. K., Vlachopoulos, P., & Parsell, M. (2019). Social network analysis and learning communities in higher education online learning: A systematic literature review. *Online Learning*, *23*(1), 249–264. https://doi.org/10.24059/olj.v23i1.1398

Johnson, S. L., Faraj, S., & Kudaravalli, S. (2014). Emergence of power laws in online communities. *Mis Quarterly*, *38*(3), 795–808. https://doi.org/10.25300/misq/2014/38.3.08

Ke, F., & Hoadley, C. (2009). Evaluating online learning communities. *Educational Technology Research and Development*, *57*(4), 487–510. https://doi.org/10.1007/s11423-009-9120-2

Kear, K., Chetwynd, F., & Jefferis, H. (2014). Social presence in online learning communities: The role of personal profiles. *Research in Learning Technology*, *22*. https://doi.org/10.3402/rlt.v22.19710

Lai, K. W. (2015). Knowledge construction in online learning communities: A case study of a doctoral course. *Studies in Higher Education*, *40*(4), 561–579. https://doi.org/10.1080/03075079.2013.831402

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

Li, S., Du, H., Xing, W., Zheng, J., Chen, G., & Xie, C. (2020). Examining temporal dynamics of self-regulated learning behaviors in STEM learning: A network approach. *Computers & Education,* 158, 103987.

Lin, C. P. (2010). Learning virtual community loyalty behavior from a perspective of social cognitive theory. *International Journal of Human-Computer Interaction*, *26*(4), 345–360. https://doi.org/10.1080/10447310903575481

Liu, Z., Zhang, N., Peng, X., Liu, S., Yang, Z., Peng, J., Su, Z., & Chen, J. (2021). Exploring the relationship between social interaction, cognitive processing and learning achievements in a MOOC discussion forum. *Journal of Educational Computing Research*, *60*(1), 132–169. https://doi.org/10.1177/07356331211027300

Macia, M., & García, I. (2016). Informal online communities and networks as a source of teacher professional development: A review. *Teaching and Teacher Education*, *55*, 291–307. https://doi.org/10.1016/j.tate.2016.01.021

Markus, M. L. (1987). Toward a "critical mass" theory of interactive media: Universal access, interdependence and diffusion. *Communication Research*, *14*(5), 491–511. https://doi.org/10.1177/009365087014005003

Mathar, D., & Gaur, M. (2020). Economic value of product & digital network using regression and social network analysis. *International Journal of Computer Engineering and Technology*, *11*(1), 28–38. Available at SSRN: https://ssrn.com/abstract=3639072

McConnell, D. (2006). *E-learning groups and communities of practice*. Open University Press.

Monge, P. R., Contractor, N. S., Contractor, P. S., Peter, R., & Noshir, S. (2003). *Theories of communication networks*. Oxford University Press.

Morais, G. M., Santos, V. F., & Gonçalves, C. A. (2020). Netnography: Origins, foundations, evolution and axiological and methodological developments and trends. *The Qualitative Report*, *25*(2), 441–455. https://doi.org/10.46743/2160-3715/2020.4227

Mtshali, M. A., Maistry, S. M., & Govender, D. W. (2020). Online discussion forum: A tool to support learning in business management education. *South African Journal of Education*, *40*(2), 1–9. https://doi.org/10.15700/saje.v40n2a1803

Nielsen, J. (2006). Participation inequality: Encouraging more users to contribute. http://www.useit.com/alertbox/participation_inequality.html

Orlikowski, W. J., & Iacono, C. S. (2001). Research commentary: Desperately seeking the "IT" in IT research—a call to theorizing the IT artifact. *Information Systems Research*, *12*(2), 121–134. https://doi.org/10.1287/isre.12.2.121.9700

Peters, A. M., Crane, D., & Costello, J. (2019). A comparison of students' twitter use in a postsecondary course delivered on campus and online. *Education and Information Technologies*, *24*(4), 2567–2584. https://doi.org/10.1007/s10639-019-09888-1

Poquet, O., Dowell, N., Brooks, C., & Dawson, S. (2018, March). Are MOOC forums changing? In Proceedings of the 8th International Conference on Learning Analytics and Knowledge, March 7-9, Sydney, Australia (pp. 340–349).

Staudt Willet, K. B. (2019). Revisiting how and why educators use Twitter: Tweet types and purposes in# Edchat. *Journal of Research on Technology in Education*, *51*(3), 273–289. https://doi.org/10.1080/15391523.2019.1611507

Studente, S. (2021). Staying connected: Minimizing isolation and building learning communities via chatbot technology. *The impact of COVID-19 on teaching and learning in higher education* (pp. 51–75). Nova Science.

Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the P value is not enough. *Journal of Graduate Medical Education*, *4*(3), 279–282. https://doi.org/10.4300/jgme-d-12-00156.1

Sundaram, H., Lin, Y. R., De Choudhury, M., & Kelliher, A. (2012). Understanding community dynamics in online social networks: A multidisciplinary review. *IEEE Signal Processing Magazine*, *29*(2), 33–40. https://doi.org/10.1109/msp.2011.943583

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

Ward, M. D., Stovel, K., & Sacks, A. (2011). Network analysis and political science. *Annual Review of Political Science*, *14*(1), 245–264. https://doi.org/10.1146/annurev.polisci.12.040907.115949

Wasserman, S., & Faust, K. (1994). *Social network analysis*. Cambridge University Press.

Wojcik, D. J., Ardoin, N. M., & Gould, R. K. (2021). Using social network analysis to explore and expand our understanding of a robust environmental learning landscape. *Environmental Education Research*, *27*(9), 1263–1283. https://doi.org/10.1080/13504622.2021.1905779

Xing, W., & Gao, F. (2018). Exploring the relationship between online discourse and commitment in Twitter professional learning communities. *Computers & Education*, *126*, 388–398. https://doi.org/10.1016/j.compedu.2018.08.010

Xu, T., Wu, Q., & Xu, Z. (2021, March). The impact of online learners' social interaction on learning achievement based on social network analysis. In 2021 9th International Conference on Information and Education Technology (ICIET), Okayama, Japan, 27–29 March 2021 (pp. 232–241). IEEE.

Xu, Y., Lynch, C. F., & Barnes, T. (2018). How many friends can you make in a week? Evolving social relationships in MOOCs over time. Proceedings of the 11th International Conference on Educational Data Mining, EDM 2018, July 15-18, Buffalo, NY, USA.

Yao, Z., Yang, D., Levine, J. M., Low, C. A., Smith, T., Zhu, H., & Kraut, R. E. (2021). Join, stay or go? A closer look at members' life cycles in online health communities. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–22. https://doi.org/10.1145/3449245

Yassine, S., Kadry, S., & Sicilia, M. A. (2021). *Detecting communities using social network analysis in online learning environments: Systematic literature review* (p. e1431). Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery.

Zaheer, A., & Soda, G. (2009). Network evolution: The origins of structural holes. *Administrative Science Quarterly*, *54*(1), 1–31. https://doi.org/10.2189/asqu.2009.54.1.1

Author Biographies

Wanli Xing is an Assistant Professor of Educational Technology at the University of Florida. His research interests are artificial intelligence and learning analytics in STEM and online learning.

Hanxiang Du is a PhD candidate at the University of Florida. Her research interests are learning analytics and computer science education.