CourseNetworking and community: linking online discussion networks and course success

Adrienne Traxler,¹ Andrew Gavrin,² and Rebecca Lindell³

¹*Department of Physics, Wright State University, 3640 Colonel Glenn Hwy, Dayton, OH, 45345*

²*Department of Physics, Indiana University Purdue University Indianapolis, 420 University Blvd., Indianapolis, IN, 46202*

³*Department of Physics and Astronomy, Purdue University, 525 Northwestern Ave., West Lafayette, IN, 47907*

Large introductory science courses are isolating for many students, and reducing this isolation is an important factor for student retention in college. Active learning courses often build community among students as an explicit goal, but many commuter or non-traditional students have limited on-campus time. Online discussion forums provide one tool for engaging students with each other outside of class time. This study uses social network analysis with forum data from an introductory physics course to examine students' positions in the class discussion network and link it to their final course grades. We find that, contrary to expectations, there is no strong correlation between forum network centrality and class outcomes. Possible reasons for this mismatch and future refinements to the model are discussed.

I. INTRODUCTION

Widespread integration of technology in learning environments has made online forums a common tool for promoting discussion. Forums provide a place for students to ask questions of each other or the instructor, trade information, or share resources such as news articles or interesting videos. Many instructors see forums as a tool for extending in-class discussion and fostering community among students [1]. The latter function can be especially valuable at institutions with many commuter or non-traditional students, who may have scarce on-campus hours for study groups but still benefit from increased engagement [2]. This paper focuses on the use of the CourseNetworking (CN) forum software in an introductory calculus-based physics course at a primarily commuter university.

Individual-focused measures of CN participation and course performance are detailed in a companion paper [3]. Here, we focus on the network of interactions captured by forum data—not just how often a student participated, but their conversation partners and overall prominence in the aggregated structure of threads. Past results from social network analysis, outlined below, suggest that the most central students in this network may experience better course outcomes [4–6].

II. BACKGROUND

In the context of university physics courses, social network analysis (SNA) has been used to contrast different classroom types [7], to explore connections between student networks and success measures [4–6], and to trace the development of student communities [8, 9]. More broadly, studies of online course discussion forums have linked student participation in a forum network with sense of course community [10] and found that forums with a high amount of student social structure can demonstrate higher levels of critical thinking [11]. If the CN forum serves its intended purpose as a social space for students to informally collaborate and connect to peers,

those who regularly post and reply to other students may see beneficial effects on their grades.

Classroom networks seem to show some degree of self-segregation, where over time students group by obvious characteristics such as gender or a shared recitation section [8]. In an online space, these grouping tendencies may recur, or student groups may be more homogeneous if they are based primarily on liking the content of other users' posts. Or there may be chance-based groupings, such as students with similar work schedules whose forum activity tends to be synchronized and who see each other's posts at the top of the page. Ideally, by the end of the semester, students will readily engage each other with comments and questions about the course material and related topics such as science in the news. The current first stage of exploring the forum network provides a foundation for future analysis using network community detection tools.

Community building among students is a valuable goal in its own right, with important links to persistence in college [12]. Here we particularly focus on academic success within the course as an outcome, and ask the research question: To what extent does student position in the forum posting network correlate with course success, as measured by final grade?

III. METHODS

A. Data collection and context

Data was collected in fall 2014 at a large, urban university in the Midwest with a large number of commuter and non-traditional students. The course was calculus-based introductory mechanics, with around 150 students, and met face-to-face weekly for two hours each of lecture, recitation, and laboratory. One of the authors (AG) was the instructor. The CourseNetwork online discussion forum was officially introduced the first week of class. Participation in forum discussions was worth up to 5% extra credit on the final grade. Though it was well incentivized in this fashion, many students kept posting after reaching maximum bonus points.

The first forum activity was in early July, an introduction poll posted by the instructor. The first student replies occurred in late July, and classes started in the last week of August. Classes ended in the second week of December, and the final student posts appeared in the fourth week of December after final grades were published. Data downloaded from the forum includes a unique student ID code for each post, whether the activity was a post, poll, or reply to one of those two, the timestamp, the text of the post, the number of pictures or other attachments, and the number of “likes” received by the post. Also available for each student are gender and their final grade in the course after removing bonus points from forum activity.

Future analysis will discuss post content and time-dependent structure. The current work focuses on students’ posting patterns over the whole semester, their resulting importance in the forum network structure, and how those positions correlate with their final grades. Accordingly, at this stage posts were not coded for content, only tagged as thread-starters (either posts or polls) and replies.

B. Network analysis

In some cases a reply to a post may be obviously targeted toward a single student, but more often the recipient is ambiguous or more general (anyone reading the forum, or perhaps anyone participating in that particular discussion thread). To analyze the data as a network object, we must first decide what constitutes a connection between students (also known as an edge). Existing literature does not provide strong guidance on this point, as other network analyses of asynchronous learning have relied on pre/post surveys rather than using forum data [13, 14], have assumed links only to the immediately preceding post [11], or have not specified the method of link generation [10]. We chose a method that has been previously used in analyzing scientific collaboration networks, and treated the data as a bipartite network [15, 16].

The threaded structure of CN data lends itself well to bipartite network analysis [16], where there are two types of nodes: actors (students) and events (discussion threads). Links in bipartite data can exist only between nodes of different types—for example, a student posting in a thread draws a link between that student-thread node pair, but two threads will not directly connect to each other. A student will have many outward-going links, to each thread in which they posted, and a thread will receive incoming links from the original poster and all students who replied.

A bipartite network can be collapsed using one of two projections called co-affiliation networks. In the actor co-affiliation network, which we examine here, two students are linked if they participated in the same discussion thread. If they shared multiple threads, which was often the case in this data, each additional link increases the edge weight. This representation privileges large threads (in which participants will link to many other students in the network) and posting

multiple times in the same thread (which generates higher-weight edges). We argue that, though these are certainly not the only measures of participation, they do reasonably indicate a higher level of forum activity.

With nodes and edges thus defined, we can estimate students’ network positions by calculating their centrality values. Centrality is a family of measures for classifying a node’s connectedness in a network. At the most basic level, degree centrality simply measures the number of connections possessed by a node. Another commonly used measure, PageRank [17], incorporates information about how well-connected a node’s neighbors are, reasoning that having better-connected contacts will confer a more useful position in the network. Bruun and Brewe [6] find that PageRank is positively correlated with students’ final course grades in an introductory course where the network was constructed from weekly surveys. They also identify two additional measures, target entropy (TE) and hide (H), that feature prominently in their analysis. Target entropy concerns the variety of messages passed to a node: a high-TE node is one with many connections and links among those connections, meaning that it may receive messages from a diversity of sources and thus has access to richer information. Students who participate in threads with a wide variety of other students will tend to have high target entropy. Hide measures the number of “steps” a person would have to take to reach a given student from elsewhere in the network. A large hide value indicates that a student is only tenuously connected, through one or a few low-activity threads. In this data, we would expect target entropy to correlate positively (if at all) with grades, and hide to have a negative correlation. Both of these correlations were found in the prior study of Bruun and Brewe [6].

All three centrality measures (PageRank, TE, and hide) are inherently interdependent types of data, because their values derive from surrounding nodes. Instead of making the standard statistical assumption of independence, we use a permutation method to evaluate correlations between each centrality measure and course grade [18]. This technique repeatedly resamples the data ($n = 10000$ repetitions here), creating a distribution of possible values that allows estimation of whether the actual measured correlation is unlikely to have occurred by chance.

IV. RESULTS

The complete data set includes 936 forum threads and 2376 replies, with 156 participants in the discussion (154 students, one instructor, and one CN staff member). Removing the instructor’s high post count, the average number of student posts and replies was 20.8, and each thread received an average of 2.5 replies.

When the actor projection of this bipartite network is taken, the resulting network has 156 nodes and 3814 edges (connections between students). Note that the number of edges and number of original threads is not the same because each

thread generates multiple edges connecting all its posters. There are twelve isolated nodes in the actor projection, representing students who posted a single thread that received no replies; in all cases these were introduction posts rather than physics content or questions. These nodes are removed from subsequent calculations on the actor network.

The remaining network is very densely connected, with a clustering coefficient of 0.64. This measure indicates that if students A and B have posted in a common thread, as have students B and C, then 64% of the time students A and C will also share a common thread. Much of this high level of connectivity was provided by a small number of high-volume threads; for example, seven threads in the data received more than 20 replies. The content of these “hub” threads varied: one was about the impending end of the lab portion of the course, two were for post count boosting (to receive more extra credit for forum participation), two were discussion/commiseration about exams, one was about a physics question that had been posed in class, and one was about forming study groups.

Figure 1 shows the actor network, with nodes sized by degree (number of connections) and colored by grade range. Grades used here and in correlations exclude extra credit points from forum posting. There are several nodes without grade data: these are the instructor, a CN staff member, and students who withdrew from the course. The thickness of lines in the figure indicates the weight of the connection (number of shared threads) between two students. Nodes at the periphery of the object tend to have one or two weight-1 edges, while the center of the network becomes quickly obscured by the many overlapping lines.

Figure 2 shows the cumulative distribution function of edge weights, which approximately follows a power law distribution. More than 50% of all edges are weight 1 or 2, and fewer than 10% have a weight higher than 5. Taken with the high mean number of posts per student, this indicates that most students participated with a large sampling of their peers rather than sticking to small cliques of people.

Bruun and Brewe [6] found that PageRank correlated with both final course grades ($\rho = 0.27$) and future grades ($\rho = 0.33$). Target entropy correlated with future grades ($\rho = 0.38 - 0.45$), and hide correlated negatively with final course grade ($\rho = -0.35$) and future grades ($\rho = -0.32$). Though target entropy was not correlated with current-semester grades in their work, it was sufficiently prominent in the results that we include it here for comparison. Table I shows the resulting correlation coefficients and p -values from our permutation tests. In the CN network, PageRank was not well correlated with final grade, $\rho_{PR} = 0.18$ compared to 0.27 in Bruun and Brewe’s work. Target entropy correlated with future but not current-semester grades in the earlier work, while here there was a correlation of 0.29 with the current semester grade, a small to medium effect size. Finally, both studies found a negative correlation between hide and final grade, $\rho_H = -0.27$ here compared to -0.35 in the prior study.

FIG. 1. Actor projection of the forum network. An edge indicates that two students posted in the same thread, and thicker lines indicate more threads in common. Nodes are sized by degree and colored by course grade, with blue, yellow, and red corresponding to high, medium, and low grade ranges (grade percentage also indicated on bar). White indicates that grade information is unavailable.

FIG. 2. Cumulative distribution function of edge weights on log-log axes. The y-axis shows the probability that an edge in the network has a weight of at least k , where k is the x-axis value. Edge weights follow approximately a power law distribution, with low weights much more frequent.

TABLE I. Correlation coefficients (ρ) of three centrality measures in the CourseNetworking forum with course grade. p -values were calculated from permutation tests ($n = 10000$ iterations).

Measure	ρ	p -value
PageRank (PR)	0.18	0.037
Target entropy (TE)	0.29	0.001
Hide (H)	-0.27	0.002

V. DISCUSSION AND FUTURE WORK

The question “how does student position in the forum network correlate with final grade?” seems straightforward on the surface. In practice, it requires non-trivial decisions about how to construct links between students and what measures of centrality may be most appropriate, based on comparisons to previously published models and the context of forum use in this course. Many classroom networks are constructed from surveys that directly query students about their interactions at only one or two points in the semester [5, 7, 9, 13, 14]. The CN data provide a much richer record of interactions, which may more closely resemble the time-aggregate and weighted data analyzed by Bruun and Brewe [6]. However, we see from our generally low correlations between network position and final grade that this first-order approximation is missing important network dynamics.

The more detailed study ahead will explore two branches: refinement in connecting the forum network, and attention to the content of posts. For the first case, some additional models for connecting forum networks do exist [11], as well as

further tools for analyzing bipartite networks [16]. We will also highlight network interactions with the instructor to see if these links were particularly beneficial. For the second branch, a significant gain in predictive power is expected from adding post content to the analysis. Categorizing forum interactions even at a rough level can split the interaction network into several distinct layers. This type of classification scheme more closely matches the approach of Bruun and Brewe [6] where interactions were divided into problem solving, concept discussion, and in-class social networks. In their study, for example, PageRank correlated with final course grade on the “problem solving” layer but not on the “concept discussion” layer. Thus, separating out post content types can disentangle many different student interactions, and clarify which forum exchanges are most likely to be beneficial. To this point, content analysis of forum posts is substantially more advanced than network construction in the literature, and a number of theory-based instruments exist for classifying this material [19]. Incorporating content data will require time-intensive analysis of the semester-long record, but promises much additional insight.

When a more detailed network (or set of network layers) has been developed, community detection tools will allow additional probing of student positions relative to their peers. We may find that student groups form and condense over the semester [8, 9], or alternately that the CN setting encourages more varied but less tight-knit interactions between students than an in-class network. For instructors using online forums to support or create classroom community, knowing what social structures tend to evolve on their own can inform decisions about what kind of instructor presence will best support the course learning goals.

-
- [1] J. R. Howard, *Discussion in the college classroom: Getting your students engaged and participating in person and online* (Jossey-Bass, San Francisco, CA, 2015).
 - [2] S. Gilardi and C. Guglielmetti, *The Journal of Higher Education* **82**, 33 (2011).
 - [3] A. Gavrin and R. Lindell, *Students' social media participation is positively correlated with homework effort* (2016), contributed poster at 2016 Physics Education Research Conference.
 - [4] E. Williams, E. Brewe, J. P. Zwolak, and R. Dou, in *2015 Physics Education Research Conference*, edited by A. D. Churukian, D. L. Jones, and L. Ding (American Association of Physics Teachers, 2015), pp. 375–378.
 - [5] Y. Yang, B. Nainabasti, D. T. Brookes, and E. Brewe, in *2014 Physics Education Research Conference*, edited by P. V. Engelhardt, A. D. Churukian, and D. L. Jones (Minneapolis, MN, 2015), pp. 283–286.
 - [6] J. Bruun and E. Brewe, *Physical Review Special Topics - Physics Education Research* **9**, 020109 (2013).
 - [7] E. Brewe, L. H. Kramer, and G. E. O'Brien, in *AIP Conference Proceedings* (2010), vol. 1289, pp. 85–88.
 - [8] J. Bruun and I. G. Bearden, *PLoS ONE* **9**, e112775 (2014).
 - [9] A. Traxler, in *2015 Physics Education Research Conference*, edited by A. D. Churukian, D. L. Jones, and L. Ding (College Park, MD, 2015), pp. 331–334.
 - [10] S. Dawson, *Educational Technology & Society* **11**, 224 (2008).
 - [11] R. Aviv, Z. Erlich, G. Ravid, and A. Geva, *Journal of Asynchronous Learning Networks* **7**, 1 (2003).
 - [12] V. Tinto, *The Journal of Higher Education* **68**, 599 (1997).
 - [13] H. Cho, G. Gay, B. Davidson, and A. Ingrassia, *Computers & Education* **49**, 309 (2007).
 - [14] H.-L. Yang and J.-H. Tang, *Journal of Asynchronous Learning Networks* **7**, 93 (2003).
 - [15] M. E. J. Newman, *Physical Review E* **64**, 016132 (2001).
 - [16] S. P. Borgatti and D. S. Halgin, in *The SAGE Handbook of Social Network Analysis*, edited by J. Scott and P. J. Carrington (SAGE Publications, 2011), pp. 417–433.
 - [17] S. Brin and L. Page, in *Seventh International World-Wide Web Conference (WWW 1998)* (Brisbane, Australia, 1998), URL <http://ilpubs.stanford.edu:8090/361/>.
 - [18] D. Z. Grunspan, B. L. Wiggins, and S. M. Goodreau, *CBE—Life Sciences Education* **13**, 167 (2014).
 - [19] B. De Wever, T. Schellens, M. Valcke, and H. Van Keer, *Computers & Education* **46**, 6 (2006).