Campus Connections: Student and Course Networks in Higher Education

Uriah Israel¹ • Benjamin P. Koester¹ • Timothy A. McKay¹

Published online: 13 January 2020

© The Author(s) 2020

Abstract

Residential higher education brings thousands of students together for multiple years and offers them an array of shared intellectual experiences and a network of social interactions. Many of these intellectual and social connections are formed during courses. Students are connected to students through courses they take together, and courses are connected to one another by students who take both. These courses and the students who take them form a bipartite network that encodes information about campus structures and student experiences. Because all institutions of higher education collect and maintain precise records of what courses students take, it is possible to assemble a student-course network that quantitatively describes the interactions among students and courses. We provide an example that demonstrates the identification of courses effective at creating unique connections among students and reveals how students and majors can be strongly connected or dispersed. We show how social network analysis can be used to improve our understanding of the learning environment at the University

Uriah Israel received a B.S. in Physics and a B.S.E in Computer Engineering from the University of North Carolina – Charlotte and an M.S. in Applied Physics from the University of Michigan – Ann Arbor. He is currently a Ph.D. candidate in the Applied Physics program at the University of Michigan – Ann Arbor.

Benjamin P. Koester received a B.S.E in Chemical Engineering, M.S. in Human Genetics, and a Ph.D. in Physics from the University of Michigan – Ann Arbor. He is currently a Research Scientist in Physics at the University of Michigan – Ann Arbor.

Timothy A. McKay received a B.S in Physics at Temple University and a Ph.D. in Physics at the University of Chicago. He is currently the Arthur F. Thurnau Professor of Physics, Astronomy, Education and the Associate Dean for Undergraduate Education in the College of Literature, Science, and the Arts at the University of Michigan - Ann Arbor.

✉ Uriah Israel
ulisrael@umich.edu

¹ Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA

of Michigan, and we hope that this kind of analysis is of interest to persons at other institutions.

Keywords Network analysis · Curriculum studies · Undergraduate education

The courses that students take during their college studies arise from a combination of institutional requirements and the choices the students make to fulfill them, along with their other interests. Each choice creates an opportunity for connection to faculty members, other students, ideas, and experiences. The courses chosen are subsequently recorded on a transcript that focuses on the activities of a particular individual: major, courses and grades received, grade point average, and perhaps honors. However, the experience of the student can be better understood when placed in context and understood in relation to other students, which provides insights into interactions with peers and connections to the broader intellectual environment.

The interactions that occur among college students on the same campus have consequences for their experiences as learners in single courses (Biancani & McFarland, 2013; Carolan, 2013) and reflect the eventual academic outcomes at the conclusion of their studies (Biancani & McFarland, 2013; Gurin, Dey, Hurtado, & Gurin, 2002). To varying degrees these connections may be categorized as physical, behavioral, or associative (Carolan, 2013). Physical connections occur when individuals occupy the same space, behavioral connections include the active exchange of information, and associative connections emerge from the shared intellectual experiences and activities of individuals in the same course or major. Collectively these connections constitute a social network (Wasserman & Faust, 1994). Using network science and analysis to study this social network enhances our understanding of the learning environment of modern higher education.

A network where the nodes represent people or groups of people and the edges represent the interactions or connections among them is known as a social network. Because of the various ways in which people and groups interact or connect, the edges in a social network can take many forms. For example, edges can be formed from communication (Java, Song, Finin, & Tseng, 2007; Newman, Forrest, & Balthrop, 2002) or relationships (Ellison, Steinfield, & Lampe, 2007; Ugander, Karrer, Backstrom, & Marlow, 2011). A collaboration between and among groups of scientists is an example of a social network in higher education (Dobson, Carreras, Lynch, & Newman, 2007; Newman, 2001). In a network of that type the nodes represent individual scientists, and two scientists are connected (i.e., there is an edge between them) if they share co-authorship on a paper. Carolan (2013) refined the network description to four levels of analysis. The first is the “ego” level, referring to an analysis of one node, the ego, and all of its connections, which are the “alters.” The next two levels grow to collections of nodes, either dyads (pairs) or triads (triplets). The fourth level, which is the focus of this manuscript, is the analysis of the complete network, which encompasses sets of actors and ties among them in a bounded sample.

The study we report in this article responds to the call of Biancani and McFarland (2013), who argued that social network research in higher education “lacks a rich body of descriptive work portraying the student experience of college from a network perspective.” We first provide useful definitions with a literature review. Next we describe the study. Discussion, practical applications, suggestions for future research, and conclusions follow.

Definitions

The Term *Network*

Within the domain of *network science* a network is generally understood as a collection of two different types of objects called nodes and edges that depict a system. The nodes in the network represent subcomponents of the system while the edges, which connect the nodes and can also be called connections, represent interactions between and among these subcomponents. The resultant structure of the network can describe the connectedness of the system, the stability of the system, the information flow, and much more. This network structure then allows for the identification of components and interactions of interest. Two examples of large systems that have been represented as networks and that might be familiar to many are power grids (Watts & Strogatz, 1998) and railway networks (Sen et al., 2003). There is a set of commonly measured network attributes and statistics that help in understanding network structure and composition.

Measurement Terms

We now define two of the common measurement terms that are found in the literature related to network science. Both degree centrality and local clustering coefficient were used in our study.

Degree Centrality This measure simply counts the total number of nodes to which each node is connected within the network. In NetworkX, which is a software package, this value is normalized by dividing the number of connections (or edges) a particular node has by the total number of possible connections the node could have (i.e., a network with n nodes is normalized by $n-1$). In this sense it reports the fraction of all the network nodes to which each node is connected. Degree centrality is often considered in relation to information flow through a network because it provides an estimate of the probability that this node will play a role in transmitting ideas (or information) within the network.

Local Clustering Coefficient The clustering coefficient relies heavily on the idea of network transitivity. *Transitivity* is a term that reflects relationships: if node u is connected to node v and v is connected to w , then u is also connected to w . “The friend of my friend is also my friend” (Newman, 2018, p. 183). This leads us to an idea of *partial transitivity*, which is a concept in social networks where the fact that u knows v and v knows w does not guarantee that u knows w , but makes it more likely (Newman, 2018). The clustering coefficient is an attempt to measure the level of partial transitivity of the network, whereas the local clustering coefficient is the clustering coefficient for a specific node. For each node, it probes how often this node provides a unique connection between two other nodes. The value of clustering coefficient indicates whether a node is embedded in a tightly clustered region of the network. Conversely, it tells us something about how powerful a node is for connecting otherwise unrelated nodes. The lower the value of the clustering coefficient, the more powerful it is at providing unique connections (Newman, 2018).

Related Research

Due to the benefits of social network analysis, interest in applying it to education research has arisen. Biancani and McFarland (2013) reviewed multiple settings in which networks have been applied in education research, including student relationships and faculty research collaborations within and among departments. They identified threads of research as “networks as dependent variables,” “networks as independent variables,” and “descriptive.” As dependent variables, common themes have involved the roles of race and ethnicity in the formation of friendships (D’Augelli & Hershberger, 1993; Defour & Hirsch, 1990; Kenny & Stryker, 1994) by using surveys and later using Facebook data matched with institutional administrative data (Mayer & Puller, 2008). Lewis, Kaufman, Gonzalez, Wimmer, and Christakis (2008) also explored additional demographic traits as predictors of network behavior on Facebook. As independent variables, networks have been notably tested as predictors of GPA (Antrobus, Dobbelaer, & Salzinger, 1988; Zimmerman, 2003), student integration and persistence (Thomas, 2000), and health outcomes (Duncan, Boisjoly, Kremer, Levy, & Eccles, 2005).

The earliest example of a descriptive study of student networks is found in (Festinger, Schachter, & Back, 1950), who studied the friendships and communities formed among married veterans in housing at MIT. That work along with later descriptive studies (Newcomb, 1961; Salzinger, 1982) initiated a field that has since been lacking in activity (Biancani & McFarland, 2013). With our work we hope to rekindle this field by using large student enrollment datasets, which all colleges have, as well as modern computational techniques that reshape data into a network setting for further evaluation.

Social networks as they exist in learning communities have been considered at the classroom level (Dawson, 2008; Grunspan, Wiggins, & Goodreau, 2014) and more extensively in massive open online courses (MOOCs) and E-learning (e.g., Cela, Sicilia, & Sánchez, 2014; Fincham, Gašević, & Pardo, 2018; Veletsianos, Collier, & Schneider, 2015). Additional examples can be found in education research. For instance, Dawson (2008) compared communication logs among students in an online forum to draw connections between them. In that study the student network measures of closeness, degree, and betweenness were used to assess each student’s sense of community. Betweenness is defined as the extent to which a node lies on the paths between other nodes (Newman, 2018). Dawson (2008) and Calvó-Armengol, Patacchini, and Zenou (2009) asked how a student’s centrality within a school friendship network affects student performance. Watanabe and Falci (2016) looked at friendship network size among faculty members and its relation to perception of work-family culture. Diramio, Theroux, and Guarino (2009) used networks to look at faculty hiring patterns at top universities.

These studies offer a few examples of ways to characterize and analyze networks in higher education. They highlight how networks play a role in shaping higher education. Much of this research took advantage of only a few network measures to draw powerful conclusions: among them degree, closeness, and betweenness are most common. These studies focused on small networks, that is, no more than a few thousand connections and a small sample of individuals. The limited nature of the available data may therefore reflect a biased subset of the whole.

One of the central challenges of social network studies that still remains is how to measure connection (i.e., how edges are defined in social networks). Many forms of connection which we might like to study (e.g., friendships, collaborations, inspirations, conflicts, mutual support, mentoring) are not comprehensively recorded. As a result, social network studies need to rely on methods of estimating these based on context, survey data, or inference.

For this study we took advantage of one area of higher education for which data about substantive connection, that is, extensive physical proximity, shared intellectual experience, and a suite of activities, is carefully recorded. Every college and university has maintained careful records of the courses taken by students. These records provide an opportunity to study networks of substantive campus connection in ways which can be replicated across the landscape of higher education.

The Study

Purpose

The purpose of this study was to demonstrate the application of network analysis to a large administrative dataset in order to gain insights into the connections formed among students and courses in higher education.

Data

We used data from a large, selective, public, state university with over 40,000 undergraduate students and several colleges. In an explicit effort to make student information data more easily accessible to researchers, the University of Michigan Information Technology Services staff created the Learning Analytics Data Architecture, or LARC (Office of Enrollment Management, University of Michigan, 2019). This dataset, which is curated and distributed by the Office of the Registrar, has enabled a wide range of learning analytics research efforts, including this project. It is available to researchers for a project if their IRB is approved by Michigan and if there is a signed MOU. It provides an authoritative and complete view of the student data present in the data warehouse; and it is updated every semester, similar to the public data releases of “big science” projects such as the Sloan Digital Sky Survey (York, Adelmann, & Anderson Jr., 2000) or the GAIA Satellite (Brown et al., 2016).

The LARC currently includes four main tables. They contain over 400 columns describing more than 200,000 students who enrolled in roughly 15,000 courses since the year 2000. The content of the tables is reviewed and updated every semester to accommodate researchers’ needs and to allow for redefinition of fields as appropriate. All of the characteristics of students and courses used in this study were either drawn directly or derived from the LARC dataset. The data dictionary for LARC is publicly available online. For clarity and ease of understanding, we used more familiar names for data elements in this work rather than the official names used in LARC.

In addition to defining the connective structure of our bipartite network, data from LARC provided insight into labeling both student and course nodes with a variety of metadata. The labeling may include descriptive metadata such as the name and number of the course, offering department(s) and college, credit hours, time and location of meetings, structure (lecture, lab, discussion, seminar), enrollment, history of offerings, prerequisites, and categorization in terms of college requirements. For the metadata about students who take each course, LARC provides insights about students’ backgrounds, including information from campus admission, prior courses taken, and previous academic performance. Demographic information includes records of age, gender, ethnicity, country and state of origin, first or continuing generation status, and intended major at time of enrollment. Also, there is a complete record of subsequent courses taken and of honors and degrees ultimately earned by students.

Fig. 1 Example of a bipartite (two-mode) network. The top set of nodes are courses labeled C1, C2, ...C6, and the bottom set nodes represent the students enrolled in those courses S1, S2, ... S6. This is an example of enrollment in one semester and the connections can be interpreted as follows: student_1 (S1) is connected to student_5 (S5) because they were both enrolled in course_2 (C2) in that semester. Courses are connected when one student enrolls in both courses during the same semester. Course_1 (C1) is connected to course_4 (C4) because student_2 (S2) is enrolled in both courses

For this study we used NetworkX (Hagberg, Swart, & Chult, 2008) to model and analyze the connections among students and courses. NetworkX is a Python package constructed specifically for network analysis. We used it to extract from the networks a set of traits which characterize the role of nodes (students or courses) in the overall networks.

Method

In this section we begin by documenting the construction of our student-student and course-course networks, and we then describe the extraction of network parameters characterizing the nodes in each of the networks.

Building and Partitioning the Network The full student-course network is bipartite, meaning it consists of two types of nodes (i.e., student nodes and course nodes) (Newman, 2018). A student is connected to another student through courses that they both take, and courses are connected to courses through students who take both (Fig. 1). In practice this network is quite complex. Each student may be connected to another student through a variety of courses they take together. The list of shared courses surely colors the nature of that “connection.” Likewise, each course may be connected to another by a few or many different students, and the quality of this course-course connection is flavored by the composition of these students. The network is rich in information and can be used to address many questions about the student experience. The co-enrollment (bipartite) network is then flattened into two separate networks, a student-student network (Fig. 3) and a course-course network. In flattening the network into two separate networks there can be some information loss; however, there are still useful insights to gain from the individual networks. Once constructed, we treated these student-student and course-course networks separately for the study.

The construction of this network required consideration of boundaries. In any given term, students enrolled in courses may have academic careers that began long before a particular course enrollment and the term being examined. As a student’s career progresses, they may be

connected to students whose academic careers will continue after they graduate. Likewise, a course offered only once, before or after the term of interest, may be connected to a current course by a student who spends multiple years on campus. In this sense, there were no essential boundaries in time for our student-student or course-course networks. Every student on campus was connected backward in time through a “friends-of-friends” network, which extends back to the founding of the institution and forward in time to students not yet born.

In our construction of the network, we simplified the interpretation and eased the computational burden created by these boundary effects by restricting the study to a well-defined set of students who entered and exited the university at approximately the same time. We focused on the ego networks of a cohort of 6738 students. The ego network refers to the analysis of the individual student and their connections. These students entered University of Michigan as undergraduate students for the first time in the fall of 2011 (including transfer students), and they graduated or were still enrolled by winter 2016. By the end of this five year period, 90.9% of those who entered in fall 2011 had completed a degree. This five-year graduation rate is typical for Michigan undergraduates. Most cohort students (90%) were in their first term of college attendance, and these students typically completed 6–10 terms of coursework during this five year period (the statistical mode is 8). The remainder of the cohort was almost entirely transfer students, who typically took between 2 and 6 terms of courses. The most frequent number of courses completed by the cohort of students was 33, with an average of about 4.1 courses per term.

The complete student-course network of these students included the full complement of their classmates from fall 2011 to fall 2016, a total of 68,946 students who enrolled in a total of 6152 courses. The student network itself was an aggregation of 6738 ego networks, one for each individual student within the cohort. This aggregation resulted in a network containing 68,946 nodes: 6738 egos and their 62,208 alters or classmates. Thus, the thrust of the analysis of the student network concerned this cohort: we observed only some parts of the network for students (alters) who entered before fall 2011 or who graduated after winter 2016. This made information about network structure “fuzzy” at the boundaries. We did not know whether, for example, two “older” students might be connected by a course taken prior to fall 2011 or two “younger” students who entered in fall 2016 might later become connected by a course. Only connections observed within the cohort of students who entered in fall 2011 and were still enrolled or graduated by winter 2016 are complete. For this reason, we studied and now report only the networks produced by cohort students in what follows.

Results

We now describe characteristics of the course network and the student network, exploring along the way how these measurements might be used by various stakeholders in the higher education system.

The Course Networks

We begin by examining the courses with highest degree centrality, which is a measure counting the number of connections a node has. Network statistics for the top 10 courses by degree centrality are shown in Table 1. For each node we report course name and number, clustering coefficient, degree centrality, total number of students enrolled from fall 2011 to winter 2016, and course format.

Table 1 Course Network Statistics

Course	CC	DC	N	Format
STATS-250	0.0776	0.523	20,231	LEC
ASTRO-106	0.104	0.422	4065	LEC
UC-280	0.0993	0.4	12,632	REC
ECON-101	0.102	0.411	14,065	LEC
PSYCH-111	0.104	0.403	12,516	DIS
DANCE-100	0.11	0.392	2761	LAB
ANTHRCUL-101	0.117	0.379	8190	LEC
ENGLISH-223	0.122	0.361	3184	REC
SPANISH-232	0.124	0.342	7206	REC
PSYCH-240	0.134	0.337	6621	LEC

Top 10 courses by degree centrality

CC clustering coefficient, DC degree centrality, N total enrollments over a 6 year span

Several of these classes (e.g., STATS 250 – Introduction to Statistics and Data Analysis, ECON 101 – Principles of Economics, PSYCH 111 – Introduction to Psychology) have especially large enrollments, which naturally increases their degree centrality and gives them a prominent role in the course network. STATS 250 in particular is the largest course on campus, taken by well over half of all Michigan undergraduates at some point in their careers: 75% of the students in the course have sophomore or junior standing. STATS 250 continues to provide a basic foundation in statistical thinking to students from many disciplines: social scientists in Psychology, Sociology, and Economics; natural scientists in Biology, Chemistry, and Astronomy; and humanities majors in History, English, and Linguistics. The low local clustering coefficient (0.078) reflects the fact that it is especially likely to form the *only* connection between students whose academic experiences are otherwise remote.

ASTRO 106 – Aliens, provides an interesting and non-intuitive counterexample to the obvious large enrollment trend. While it enrolls only a fifth as many students as STATS 250, it has nearly the same level of degree centrality. ASTRO 106 is a one credit course on extraterrestrial life that is often taken by individuals in the College of Literature, Science, and Arts, either out of interest or as part of fulfilling a quantitative reasoning requirement. As such, it is often taken by students in the latter half of their studies (60% have junior or senior standing) and draws from a very wide variety of majors. ASTRO 106 has a higher local clustering coefficient, indicating that it forms fewer unique connections between students than does STATS 250. However, the quality of those connections is likely different because ASTRO 106 is not taught in a large lecture hall.

Smaller enrollment is also characteristic of DANCE 100 – Introduction to Dance, which is housed in a separate, and smaller college: the School of Music, Theater, and Dance. The enrollments in DANCE 100 are smaller than any of the other courses in Table 1. Designed specifically to introduce non-dance majors to the subject, it is offered in many small “lab” sections, in which students drawn from all over campus are brought into extended, close contact. Such a course plays an outsize social and intellectual role on campus, a role which might be invisible to both students and campus leadership without this network analysis.

UC 280 – Undergraduate Research, provides another dramatically different example. This course, which meets once a week in groups of 40, delivers support for students engaging in undergraduate research in faculty labs during their first and second years on campus. As such, it engages a wide variety of students in the shared experience of working closely with a faculty-led research group. The high degree centrality of UC 280 shows that it connects a large

number of students, and the very low clustering coefficient shows that the connections it forms among students are often unique.

The role of these courses as especially powerful connectors should be more widely known, both to students and campus leaders, because they are adept at leveraging existing diversity on campus by exposing students to peers with whom they might not otherwise interact. The variety of enrollments and formats also serve as important reminders that the course structure and format of STATS 250 includes several hours a week in large lectures, while ASTRO 106 is offered in smaller, more intimate discussion sections, which classify the connections differently. While they share the same space and engage with the same content, STATS 250 students are likely to engage with only a small fraction of their classmates, while those in ASTRO 106 likely enjoy a setting that may allow for more meaningful interaction.

Other large enrollment introductory courses like ECON 101 – Principles of Economics, PSYCH 111 – Intro to Psychology, and ANTHRCUL 101 – Introduction to Anthropology connect to many other classes; but they provide *unique* connections among these courses less often than does STATS 250. They are less likely to connect to courses across more substantial disciplinary divides on campus. This distinction becomes stronger for more advanced courses on the list, like PSYCH 240 – Introduction to Cognitive Psychology. This course is taken primarily by students who will major in psychology or one of the several forms of neuroscience, so it is less likely to connect to more distant subject areas. A course does not necessarily facilitate unique connections among other courses by virtue of its size.

Fig. 2 The Degree Centrality vs. Clustering Coefficient for 1878 University of Michigan courses with enrollments $N > 100$. Each point represents a course and total course enrollment over the time period considered; the 10 courses in Table 1 occupy the high degree centrality low cluster coefficient. Point sizes indicate the relative course size. Marginal histograms show the one-dimensional distributions of degree centrality (top) and clustering coefficient (right)

Examining the structure of the course network more generally, we see that the relationships among total enrollment, degree centrality, and the clustering coefficient are strong (Fig. 2), with the largest, most highly connected courses also more likely to provide unique connections. However, there is complexity here too.

Some large enrollment courses, for example, PHYSICS 240 – General Physics II remain relatively isolated and do not connect to many other courses; and they rarely connect otherwise unconnected pairs of students. Conversely, some relatively small enrollment courses have strikingly high degree centrality and low clustering coefficients, which shows that they are especially effective at providing connections among otherwise remote courses. A handful of courses live in very tightly clustered environments so that almost every pair of courses they are connected to is *also* connected to one another. Examples of such courses include studio music courses, along with advanced undergraduate courses in Pharmacy, Classics, and Naval Architecture.

These examples help to illustrate some of the ways in which course network information might inform various audiences on campus. Students could use this information to seek out courses which will connect them to a more diverse array of other students. Faculty members

Student-Student sub-network

Fig. 3 Flattened student-student network for large majors. Shaded circles represent individual students (nodes), connected by edges whose length is inversely proportional to the strength of the connection between students. Students co-enrolled in many courses (as those in same major often are) cluster tightly within a major, and the majors themselves cluster according to the co-enrollment of students in the two majors

might use these networks of connection to better understand where their students are coming from and where they might go. Administrators might use these networks to better understand the student experience and perhaps to draw together the community of instructional teams working on the most connected courses, supporting them more openly in their efforts to create especially inclusive and equitable experiences for the diverse students whom they serve.

The Student Networks

In Fig. 3 a sample of the student network is shown for a selection of highly populated majors, both egos and alters. As expected, students cluster by major through the courses in which they co-enroll; and some majors are more tightly clustered in the network than others, reflecting both the relative flexibility of curricular requirements and the choices made by each individual in course selection. Neuroscience majors, which include many pre-medical students, are tightly clustered, due in part to a large set of prerequisite biology and chemistry courses, while Political Science majors show reduced clustering that reflects the relatively greater freedom they have in fulfilling their degree requirements. The clustering also does not obey hard boundaries. The English majors found among the Economics cluster (and vice versa) are individuals who used freedom in the curriculum to enroll in several courses commonly taken by students in the other major, and perhaps even to double major.

In Fig. 4 we display the relationship between the full student network degree centrality and clustering coefficient for all of the students in the fall 2011 cohort. The correlation between the two is similar to that seen in the course network, though less pronounced, largely because the highest

Fig. 4 Degree Centrality vs. Clustering Coefficient for University of Michigan students. Each point represents a student in the cohort defined in the text. Marginal histograms show the relative distributions of cluster coefficient (right) and degree centrality (top) among students

degree centrality seen in the student network is only about half that of the course network. Some basic features of the student network merit further mention.

The students with the highest degree centrality in the fall 2011 cohort are connected through course co-enrollment to almost 25% of all the 6738 Michigan undergraduates who entered in that term; they each took courses with about 1500 unique individuals. Those with the top five degree centralities all had different majors, though all appeared to be pursuing a track toward medical school. Four had completed their undergraduate degrees: a Biopsychology, Cognitive Science and Neuroscience B.A.; an Asian Studies B.S.; a Biomolecular Science B.S.; and a Biology, Health, and Society B.S. One of the five students had not yet graduated in winter 2016. All of these students were enrolled for the full five years, completing ten terms and taking anywhere from 41 to 47 courses. Degree centrality can only rise as one adds additional classes, and it does so especially rapidly with large STEM courses, like these which premedical students regularly take.

There are also students with very low degree centrality; they are connected to almost no other students in the fall 2011 cohort. Many are transfer students, who entered in fall 2011 as juniors or seniors; took courses primarily with older, non-cohort students; and graduated without forming extensive connections in this cohort. The most extreme cases, occupying the upper left of Fig. 4, are students who returned to school to receive, for example, a second-career B.S. in Nursing. By embedding enrollment in the network setting and without explicitly labeling transfer students, the resultant network statistics easily identify them as outliers. These measures reflect an important reality of their university experience, that is, they interacted with a much smaller and less diverse group of peers.

At each level of degree centrality, we find students whose clustering coefficients cover a broad range. Some are embedded in relatively dense parts of the network, like large majors in the College of Engineering. Such students rarely create unique connections between pairs of other students: they are almost always already connected. Others are vital connectors, regularly providing the only connections between students in deeply connected but otherwise separate neighborhoods. These students are unusual, often majoring in two or more fields, sometimes studying in two separate colleges such as the School of Education and the College of Literature, Science, and the Arts.

Discussion

Practical Applications of Network Analysis

A complete student-course bipartite network can provide answers to questions raised at many levels within higher education. For example, presidents, provosts, deans, department chairs, and faculty members may seek to better understand how to design their curricula or to allocate resources to courses that might provide opportunities for a greater number of students to experience the benefits of diversity. Which courses are especially important for creating interdisciplinary, cross-campus connections? Which courses provide especially rich opportunities for connections among students with differing backgrounds, interests, and goals? Where do first year and senior students or traditional and non-traditional, students interact? Campus leaders may also use these same analyses to gain a deeper understanding of the student experience, identifying groups of students who are especially well-connected or especially isolated, exploring the relationships between curricular requirements and student connections, and designing new courses which enhance desirable connections where they are lacking.

There are also practical applications for students, who may now probe the breadth and depth of their academic experience with greater clarity. They will have the opportunity to see what kinds of courses are likely to build their network of connections and to examine and evaluate the extent to which they have contributed to the network of connections across the campus. They can query how similar or diverse they are to other students in the network and then make informed decisions. Students can see how they are connecting otherwise disparate parts of campus (e.g., they could be connecting two departments that do not normally have students interacting). They can also see the reverse and see how isolated they are relative to the possibility of connections on campus. While answers to some of these may seem intuitive, network analysis helps us quantify these ideas.

Once constructed, student-student and course-course networks can be analyzed to identify community structures using any of the variety of community finding algorithms developed by the network science community over the last few decades (Newman, 2018). Communities finding algorithms partition networks into groups of nodes having high within group connectivity and low between connectivity. Community finding provides insights into the hierarchical structure of networks. If used on communities of students or courses, community finding algorithms will likely identify obvious communities, that is, majors and colleges, while also quantifying similarity and difference in new ways, showing, for example, that the courses taken by economics students more closely resemble that of natural scientists than social scientists. These types of algorithms also probe which elements of the curriculum are important primarily for a discipline and which elements are cross-disciplinary. Community finding will identify the communities that emerge due to the network structure of connections and will allow for a direct comparison to department requirements.

As these results are presented to groups of students, faculty, administrators, and our peers in higher education, future work will be undertaken to understand how they are actually used in practice. Moreover, measures such as these will be of broad interest to members of the research community who may use them to address larger research questions that use network statistics as either independent variables predictive of certain outcomes or in the mode of dependent variables, where the network statistics are the outcomes themselves (Biancani & McFarland, 2013).

Refinements and Extensions

As others have pointed out, networks may not measure precisely what was intended. The process of validation for a measure or instrument is an integral part of social science; for example, Fincham et al. (2018) and Messick (1994) recently treated the matter in detail in the context of social ties formed in MOOC forums. In the framework of Carolan (2013), co-enrollment in a course certainly creates a physical and associative connection, but what behavioral connections are formed among students is not known. Is it necessary that both students take the course at the same time, or can a meaningful “shared” experience emerge when they both complete the same course in different terms? Should the strength of connection be weighted by credit-hours, time in class, course structure, or measures of difficulty? Is the connection created by a small seminar stronger than that created by a large lecture course? Are two courses equally “connected” when a student takes them both in the same term or takes one in the first year and one in the fourth year? Sensible answers to these questions depend upon the specific outcomes or phenomena of interest. This work does not directly address all of these questions, but we suggest that it does lay the groundwork for future studies.

Important refinements and extensions of this work remain to be explored. Each choice made in the construction of our bipartite student-course network is open for reconsideration. For example, an immediate task is the exploration of weighting of the connections produced by co-enrollment. This may be especially important if we want our student-student networks to model social connections reasonably well.

Another significant extension will allow for course connections to form when the same students take a course in different terms. This will change the course network in substantial ways, emphasizing course sequences associated with major requirements which are currently absent. In an analogous way we might create a student network meant to reflect only shared intellectual experiences and to connect students who take the same course in different terms. Such a network would describe shared intellectual experiences, social ones less so.

Other opportunities exist to gather richer, more precise measures of campus connections. Perhaps the most important involve the input from the students themselves. A number of learning analytics studies (Zwolak, Dou, Williams, & Brewé, 2017) have been conducted by relying on student self-reports of networks of connection, ranging from networks on social media like Facebook to self-reports of who studies with whom.

Conclusion

In this article we have demonstrated the power of the network framework in the context of large institutional datasets. Networks are a direct means to quantify elements of the course-course and student-student interactions for *all* students and *all* courses. The features of the network presented thus far only scratch the surface of the rich description this framework affords. Recent literature (Biancani & McFarland, 2013) has called for reinvigorating social network analysis in higher education as a means of description of higher education. We believe that our analysis responds to this challenge by assembling a bipartite student-course network, which consists of over 68,000 students connected by 6152 courses. The flattened student and course networks are then turned to describe the relationships of courses to one another through the students that take them and relationships of students to one another through the courses in which they co-enroll. We reached the following conclusions.

- The intuited belief that high enrollment courses uniquely connect students from different academic backgrounds is confirmed for some courses, but is not the rule.
- Specific low enrollment courses can also serve as equally powerful unique connectors of students.
- Students cluster by major, as expected; and unusual students, such as transfer students, appear with low degree centrality due to the smaller number of courses they have taken on campus.
- Some majors, particularly pre-medical, exhibit higher degree centrality due in part to enrollment in more large courses.
- High degree centrality in students is also not a guarantee that they form unique connections (low local clustering coefficients) among different communities.

These results confirm that network measures accurately recover intuitive connections and uncover those that are less than apparent. At this point it is important to recall that, despite the ability of a course to facilitate unique connections among students, simply occupying a space in a large lecture at the same time is no guarantee of a meaningful exchange. Proper interpretation of these results requires an acknowledgement that course formats can be more or less conducive to this kind of engagement, and lack of course format should be borne in

mind by those using these tools to draw conclusions about particular courses and their students.

Sometimes a lack of connection may be especially troubling, as it might be if students working on algorithmic data science have little chance to encounter ethicists. Other examples of troubling forms of isolation may take place along lines of social class, ethnicity, gender, or nationality. We suggest that institutions could have the opportunity to minimize the systems that perpetuate inequality in higher education by systematically examining networks of campus connections.

While limited in scope, this work provides a first look at the ways in which network measures of student connection through course co-enrollment can provide new insights into how students connect with students and courses connect with courses on the campus of a large public research university. Because the data used to build these networks is available at essentially every university and college, this kind of analysis could be replicated. Doing so would help everyone concerned with higher education better understand how campus connections form and where they do not form.

Acknowledgements The authors gratefully acknowledge the efforts of the University of Michigan Information Technology teams whose hard work created the Learning Analytics Data Architecture, which enabled this project. We also acknowledge financial support from the Andrew W. Mellon foundation through the College and Beyond II project and from the University of Michigan.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Antrobus, J. S., Dobbelaer, R., & Salzinger, S. (1988). Social networks and college success, or grade point average and the friendly connection. In S. Salzinger, J. S. Antrobus, & M. Hammer, *Social Networks of Children, Adolescents, and College Students* (pp. 227–246). Hillsdale, NJ: Erlbaum associates. <https://doi.org/10.4324/9781315792729>
- Bianciani, S., & McFarland, D. A. (2013). Social networks research in higher education. In M.B. Paulsen (Ed.), *Higher Education: Handbook of Theory and Research*, 28, 151–215. https://doi.org/10.1007/978-94-007-5836-0_4
- Brown, A. G., Vallenari, A., Prusti, T., Bruijine, J. H., Mignard, F., Drimmel, R., et al. (2016). Gaia data release 1: Summary of the astrometric, photometric, and survey properties. *Astronomy & Astrophysics*, 595, A2. <https://doi.org/10.1051/0004-6361/201629512>
- Calvó-Armengol, A., Patacchini, E., & Zenou, Y. (2009). Peer effects and social networks in education. *Review of Economic Studies*, 76, 1239–1267. <https://doi.org/10.1111/j.1467-937x.2009.00550.x>
- Carolan, B. V. (2013). *Social network analysis and education: Theory, methods & applications*. Thousand Oaks, CA: SAGE. <https://doi.org/10.4135/9781452270104>
- Cela, K. L., Sicilia, M. Á., & Sánchez, S. (2014). Social network analysis in e-learning environments: A preliminary systematic review. *Educational Psychology Review*, 27, 219–246. <https://doi.org/10.1007/s10648-014-9276-0>
- D'Augelli, A. R., & Hersherberger, S. L. (1993). African American undergraduates on a predominantly white campus: Academic factors, social networks, and campus climate. *The Journal of Negro Education*, 62, 67–81. <https://doi.org/10.2307/2295400>
- Dawson, S. (2008). A study of the relationship between student social networks and sense of community. *Educational Technology & Society*, 11, 224–238.
- Defour, D. C., & Hirsch, B. J. (1990). The adaptation of black graduate students: A social network approach. *American Journal of Community Psychology*, 18, 487–503. <https://doi.org/10.1007/bf00938119>

- Diramio, D., Theroux, R., & Guarino, A. J. (2009). Faculty hiring at top-ranked higher education administration programs: An examination using social network analysis. *Innovative Higher Education*, 34, 149–159. <https://doi.org/10.1007/s10755-009-9104-5>
- Dobson, I., Carreras, B. A., Lynch, V. E., & Newman, D. E. (2007). Complex systems analysis of series of blackouts: Cascading failure, critical points, and self-organization. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 17, 026103. <https://doi.org/10.1063/1.2737822>
- Duncan, G. J., Boisjoly, J., Kremer, M., Levy, D. M., & Eccles, J. (2005). Peer effects in drug use and sex among college students. *Journal of Abnormal Child Psychology*, 33, 375–385. <https://doi.org/10.1007/s10802-005-3576-2>
- Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook “friends”: social capital and college students’ use of online social network sites. *Journal of Computer-Mediated Communication*, 12, 1143–1168. <https://doi.org/10.1111/j.1083-6101.2007.00367.x>
- Festinger, L., Schachter, S., & Back, K. (1950). *Social pressures in informal groups: A study of human factors in housing*. New York, NY: Harper.
- Fincham, E., Gašević, D., & Pardo, A. (2018). From social ties to network processes: Do tie definitions matter. *Journal of Learning Analytics*, 5(2), 9–28. <https://doi.org/10.18608/jla.2018.52.2>
- Grunspan, D. Z., Wiggins, B. L., & Goodreau, S. M. (2014). Understanding classrooms through social network analysis: A primer for social network analysis in education research. *CBE—Life Sciences Education*, 13, 167–178. <https://doi.org/10.1187/cbe.13-08-0162>
- Gurin, P., Dey, E., Hurtado, S., & Gurin, G. (2002). Diversity and higher education: Theory and impact on educational outcomes. *Harvard Educational Review*, 72, 330–367. <https://doi.org/10.17763/haer.72.3.01151786u134n051>
- Hagberg, A., Swart, P., & Chult, D. (2008). Exploring network structure, dynamics, and function using NetworkX. *Proceedings of the 7th Python in Science Conference* (pp. 11–16). Retrieved from <http://conference.scipy.org/proceedings/SciPy2008/>
- Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we twitter. *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis - WebKDD/SNA-KDD 07*. <https://doi.org/10.1145/1348549.1348556>
- Kenny, M. E., & Stryker, S. (1994). *Social network characteristics of white, African-American, Asian and Latino/a college students and college adjustment: A longitudinal study*. Retrieved from <https://eric.ed.gov/?id=ED379537>
- Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., & Christakis, N. (2008). Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks*, 30, 330–342. <https://doi.org/10.1016/j.socnet.2008.07.002>
- Mayer, A., & Puller, S. L. (2008). The old boy (and girl) network: Social network formation on university campuses. *Journal of Public Economics*, 92, 329–347. <https://doi.org/10.1016/j.jpubeco.2007.09.001>
- Messick, S. (1994). Validity of psychological assessment: Validation of inferences from persons’ responses and performances as scientific inquiry into score meaning. *ETS Research Report Series*, (2), i–28. <https://doi.org/10.1002/j.2333-8504.1994.tb01618.x>
- Newcomb, T. M. (1961). *The acquaintance process*. New York, NY: Holt, Rinehart and Winston.
- Newman, M. E. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98, 404–409. <https://doi.org/10.1073/pnas.98.2.404>
- Newman, M. E. J. (2018). *Networks*. Oxford, United Kingdom: Oxford University Press.
- Newman, M. E. J., Forrest, S., & Balthrop, J. (2002). Email networks and the spread of computer viruses. *Physical Review E*, 66, 035101. <https://doi.org/10.1103/physreve.66.035101>
- Office of Enrollment Management, University of Michigan (2019). *Learning Analytics Data Architecture*. Retrieved from <https://enrollment.umich.edu/data-research/learning-analytics-data-architecture-larc>
- Salzinger, L. L. (1982). The ties that bind: The effect of clustering on dyadic relationships. *Social Networks*, 4, 117–145. [https://doi.org/10.1016/0378-8733\(82\)90029-6](https://doi.org/10.1016/0378-8733(82)90029-6)
- Sen, P., Dasgupta, S., Chatterjee, A., Sreeram, P. A., Mukherjee, G., & Manna, S. S. (2003). Small-world properties of the Indian railway network. *Physical Review E*, 67, 036106. <https://doi.org/10.1103/physreve.67.036106>
- Thomas, S. L. (2000). Ties that bind. *The Journal of Higher Education*, 71, 591–615. <https://doi.org/10.1080/00221546.2000.11778854>
- Ugander, J., Karrer, B., Backstrom, L., & Marlow, C. (2011). *The anatomy of the Facebook social graph*. Retrieved from <http://arxiv.org/abs/1111.4503>
- Veletsianos, G., Collier, A., & Schneider, E. (2015). Digging deeper into learners experiences in MOOCs: Participation in social networks outside of MOOCs, notetaking and contexts surrounding content consumption. *British Journal of Educational Technology*, 46, 570–587. <https://doi.org/10.1111/bjet.12297>
- Wasserman, S., & Faust, K. L. (1994). *Social network analysis: Methods and applications*. Cambridge, England: Cambridge University Press.

- Watanabe, M., & Falci, C. (2016). Workplace faculty friendships and work-family culture. *Innovative Higher Education*, 42, 113–125. <https://doi.org/10.1007/s10755-016-9373-8>
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of “small-world” networks. *Nature*, 393(6684), 440–442. <https://doi.org/10.1038/30918>
- York, D. G., Adelman, J., & Anderson, J. E., Jr. (2000). The Sloan digital sky survey: Technical summary. *The astronomical journal*. Retrieved from <https://doi.org/10.1086/301513/meta>
- Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics*, 85, 9–23. <https://doi.org/10.1162/003465303762687677>
- Zwolak, J. P., Dou, R., Williams, E. A., & Brewster, E. (2017). Students’ network integration as a predictor of persistence in introductory physics courses. *Physical Review Physics Education Research*, 13. <https://doi.org/10.1103/physrevphyseducres.13.010113>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.