The Small World Network of College Classes: Implications for Epidemic Spread on a University Campus*

Version 4/11/2020

Under review; comments welcome

Kim A. Weeden

Cornell University

Benjamin Cornwell

Cornell University

* Corresponding author: Kim Weeden, kw74@cornell.edu. Both authors contributed equally to this project. The authors thank Marin Clarkberg (Institutional Research and Planning) and Cassie Dembosky (former Registrar) at Cornell University for providing access to anonymized data, Lauren Griffin for research assistance, and Jake Burchard, Erin York Cornwell, and John Schneider for advice. Cornell University had no role in the study design, data analysis, or preparation of the report.

Abstract

Beginning in March 2020, many universities shifted to on-line instruction to slow the spread of the novel coronavirus, and many now face the difficult decision of whether and how to resume in-person instruction. This article uses complete transcript data from a medium-sized residential American university to map the two-node network that connects students and classes through course enrollments. We show that the enrollment networks of the university and its liberal arts college are "small-world" networks, characterized by high clustering and short average path lengths. In both networks, at least 98% of students are in the main component, and most students can reach each other in two steps. Removing very large courses slightly elongates path lengths, but does not disconnect these networks or eliminate all alternative paths between students. Although students from different majors tend to be clustered together, gateway courses and distributional requirements create cross-major integration. We close by discussing the implications of course networks for understanding potential epidemic spread of infection on university campuses.

On March 6, 2020, the University of Washington became the first university in the United States to shift to on-line instruction in response to the novel coronavirus. A number of institutions soon followed suit and by late March, most universities had either suspended classes or shifted to online instruction. By mid-April, many institutions had cancelled face-to-face summer sessions and were considering if face-to-face instruction could resume in fall.

The epidemiological justification for canceling face-to-face instruction is that infected students can spread the virus to other students in the classroom, who can then infect students in their other classes. In network terms, face-to-face instruction creates a two-mode network (see Breiger 1974; Borgatti and Everett 1997) in which students are connected through their classes and classes are connected through their students. Although enrollment in the same class does not capture all possible sources of connections between students (see Discussion), it is a major source of social structure in college students' day-to-day lives.

The growing field of social epidemiology shows that social networks contribute to distribution and diffusion of disease as well as risky health behaviors (Dickison, Havlin, and Stanley 2012; Hill et al. 2010; Luke and Harris 2007; Morris 2004; Valente 2010). Interpersonal contacts in everyday settings can contribute to propagated outbreaks of disease through direct contact – for example, when an infected person sheds the virus through respiratory droplets or aerosols in their breath. Social distancing can mitigate such transmission (Caley, Philip, and McCracken 2008; Cooley et al. 2016; Kelso, Milne, and Kelly 2009; Poletti, Ajelli, and Merler 2012; Valdez, Macri, and Braunstein

2012). Community spread can also occur via indirect contact – for example, when an infected person sheds the virus onto a surface that becomes contaminated.

In this context, the structure of the social network prior to social distancing can provide insight into factors that increase the likelihood of both direct and indirect exposure to a virus. These factors include the number of contacts people have, how frequently they are exposed to social contexts, whether people are clustered in a dense subgroup or connected only through bridges or hubs that connect otherwise poorly connected subgroups, whether people with similar attributes are concentrated in clusters or interspersed throughout the network, and the length of contact chains that connect people to each other indirectly. A two-node network can also provide insight into indirect contact that occurs via exposure to common forums such as stores, restaurants, clubs, organized group events, and, as we examine here, classrooms (Binson et al. 2001; Cornwell and Schneider 2017; Frost 2007; Laumann et al. 2004; Niekamp et al 2013; Oster et al. 2013).

Prior research on the social networks of students tends to focus on questions about how a particular network structure comes into being (e.g., racially homophilous friendship networks, or particular faculty collaborations), or how a particular network structure causally affects an outcome such as student grades or college enrollment decisions (see Biancani and McFarland 2013 for a review). Far less work describes the overall structure of networks. In a notable exception, Israel and colleagues (Israel, Koester, and McKay 2020) used transcript data from the University of Michigan to identify specific courses and students that have high degree centrality, meaning they act as "connectors," or hubs, across the network. Our analysis complements this important effort, but uses a different university context, explores how the organization of universities and classes into majors affects network structure, and emphasizes the attributes of the overall network most relevant to social epidemiological and policy questions.

We use complete transcript data from Cornell University to describe and visualize the structure of the two-node co-enrollment network during a typical semester. First, we describe the structure of the entire enrollment network, differentiating students by their major field of study. Second, we describe the enrollment network of the subset of students and courses in Cornell's liberal arts college.

Our goal is to provide a starting point for epidemiologists who want to study the diffusion of a virus through campus networks, university leaders who are faced with the difficult decision of whether to end or resume face-to-face instruction, and members of the higher education community (faculty, staff, parents, and students) who are trying to understand these decisions.

[Note: The discussion of Data and Methods appears after the Results and Discussion, in keeping with the structure of papers at the journal where this paper is under review.]

Results

University-wide Network

The course enrollment network for Cornell University in Spring 2015 is depicted in Figure 1. This network includes 17,391 nodes (3,797 courses and 13,594 students) and 71,795 edges (i.e., course enrollments). Light gray squares represent courses, circles represent students, and course enrollments are indicated with light gray lines that link students to their courses. Larger courses (≥ 100 students enrolled) are depicted as slightly larger squares with red borders. Nodes are arranged in the two-dimensional space using a spring-embedding algorithm (Borgatti 2002) such that students are positioned close to: (1) the courses in which they are enrolled; and, by extension, (2) other students who are enrolled in the same courses.

-- Figure 1 about here --

The two-mode network resembles a stadium, with students arrayed around a set of common courses in the middle cluster, which includes the larger courses. The common courses tend to be introductory courses, which are often used by students in various majors to meet distribution requirements. (These courses play an important role in the cohesion of the larger network, a point to which we return below; see also Israel et al 2020.) More advanced and major-specific courses appear all around the periphery of the network.

The enrollment network is highly structured by discipline. Students in the social sciences (blue circles) and STEM (orange circles) occupy separate regions in the network, which shade into each other. Multidisciplinary and dual-degree students (red) are scattered throughout the network, but tend to occupy space between the social sciences and STEM. Humanities, arts, and design students (yellow) are interspersed throughout but are most often found near the social sciences region, especially out toward the periphery. Students who have not declared a major (green) generally occupy the space on the left side of the network, which is disproportionately populated by lower-level courses.

Table 1. Characteristics of University and Liberal Arts College Course Enrollment Networks

Social Network Measures	University-Wide Network	Liberal Arts College Network
2-Mode (Student-to-Course) Network		
Number of students (*n*)	13,594	3,806
Number of courses (*m*)	3,797	1,467
Number of edges (*l*)	71,795	13,885
Network density	.001	.002
Proportion of nodes in largest component	.992	.997
Component ratio	.002	.001
Proportion of nodes in largest bi-component	.989	.942
Projected 1-Mode (Student-to-Student) Network		
Number of unique edges (*l*)	3,435,489	287,327
Network density	.037	.040
Clustering coefficient (transitivity/closure)	.437	.480
Average geodesic distance	2.069	2.233
Network diameter (largest observed distance)	6	5
Average *k*-step reach centrality		
$k = 1$	.037	.040
$k = 2$	.874	.725
$k = 3$	.981	.994
$k = 4$	.982	.998

The network contains a large number of students and has a low overall density (0.037). It is nevertheless characterized by considerable clustering. The weighted clustering coefficient for the projected one-mode student-to-student network is .437 (see Column 1, Table 1). This means that when student A has at least one class with student

B, and student B has at least one class with student C, it is also frequently the case that student A has a class with student C. Simulations show that in a randomly wired Erdős-Rényi network of the same size with the same density as this network, one would expect to see a clustering coefficient of only about .04. The level of clustering in this network is an order of magnitude greater. The relatively high level of clustering in this network is not surprising, as students commonly move through their major and distribution requirements with sets (cohorts) of fellow students.

Despite its disciplinary regionality and high level of clustering, the entire network constitutes close to a single component. Only 146 notes are outside this component, and not shown in Figure 1. Including these nodes, 99.2% of the student and course nodes are situated in the one main component.[1] This means that almost all of the students on campus eventually will be exposed to each other, at least indirectly, by virtue of their regular course schedules.

The density of the student-to-student projection of the two-mode network is .037, meaning that a given student shares a class with 3.7% of the other students, on average. There are 3,435,489 unique co-enrollment connections between pairs of students via their classes in this dataset. With 13,594 students in the network, this means that, on average, a given student will have been in a classroom with more than 500 *different* classmates by the time they have completed one round classes on their schedule. Students have, on

[1] The component ratio in this network is .002 (see Table 1). In a network in which all nodes are isolates (i.e., separate components), the component ratio will be 1. In a network in which all nodes are in the same component, the component ratio will be 0. This suggests that this network is highly concentrated in a single massive component.

average, 5.3 classes in a semester, including Physical Education and other 1-credit courses.

Most students in the enrollment network can 'reach' each other through very few steps. The average student-to-student geodesic distance in the projected one-mode network from the main component of the two-node network is 2.1. (Geodesics are the shortest paths between a given pair of students.) This means that most students are, on average, about one course removed from each other. If one chose a given student at random, that student is likely to attend class with a student who, in turn, attends class with any other randomly chosen student. Put differently, although it is unlikely that any two randomly chosen students would be enrolled in the same course, it is highly likely that they would be enrolled in different courses that both include the same third party.

In the co-enrollment network, students are typically connected to each other via multiple pathways. Fully 98.9% of the student and course nodes are in the largest bi-component, a substructure within a network that would require the removal of at least two nodes (of any type, course or student) before it breaks apart. This indicates that nearly all students on this University's campus are interconnected via multiple paths that are *independent* of each other, consisting of entirely separate students and separate courses. In other words, there is no single course, and no single student, that if removed would eliminate the potential for mutual indirect exposure between any pair of students.

Another indicator of the high level of cohesion in the network is *K*-step reach centrality, which captures the number of other students that a given node can access, even if only indirectly, within a certain number of steps. As shown in the first column of Table

1, the average student can reach 3.7% of the other students in just one step via a shared course. Going one step beyond, students can reach the vast majority of the network – 87.4% of the other students – via their connections to shared classmates. This number increases to 98.1% at three steps out.

The combination of the high overall clustering and short average geodesic path lengths between pairs of students indicates that the student-to-student projection of the enrollment network constitutes a small-world network (Watts and Strogatz 1998). Few students are far removed from each other, even though each student takes on average just over five classes. Even advanced students who are highly specialized and focusing on upper division courses in their major are only a few steps away from any other randomly chosen student on campus. Indeed, the diameter of the network – which is the length of the path between the most distant pair of students – is 6. One might think of this as the distance between extreme opposite sides of the structure shown in Figure 1.

Large courses can serve as hubs in the network, tying students together from different majors and network regions. At Cornell, 126 courses enrolled at least 100 students; 38 enrolled 200 or more students; and five courses enrolled over 400 students. The largest class enrolled 626 undergraduate students. It only takes one large class on a student's schedule to increase their direct exposure to the student body exponentially. Given that many large courses are distribution courses, they tend to draw students who are: (1) from different disciplinary backgrounds; and (2) at different stages of advancement through their degrees.

Although large courses are hubs in this network, they do not fully account for the "small-world" nature of the network. To demonstrate this, we removed the 126 courses with more than 99 students; together, these courses enrolled 24,323 students, or over one-third of all students. Even without these courses, the campus-wide network remains highly connected. The average geodesic distance increases from 2.1 to 2.6, but the percent of nodes in the main connected component remains high, at 99.1%. The largest bi-component in this trimmed network contains 89.6% of the nodes in the network, meaning that the vast majority of students remain connected via multiple independent pathways, via multiple courses.

Liberal Arts Network

We also examine the course enrollment network for students and courses in the liberal arts college within Cornell University. This subnetwork includes 1,467 courses and 3,806 students, and is depicted in Fig. 2 with the same color scheme as in Fig. 1. Large course nodes (large squares with red borders; n=18) represent courses with ≥ 100 students enrolled. The courses in the liberal arts subnetwork are generally smaller than those in the full network, both because the courses themselves have lower enrollments and because the students who matriculated in other colleges but took courses in the liberal arts college have been removed from the analytic sample.

-- Figure 2 about here --

The liberal arts network structure (see Figure 2) does not include as prominent a common core of courses in the middle of the diagram, although some larger courses do appear and more advanced and major-specific courses appear around the periphery. As in the University-wide network, majors structure the enrollment network. Students in the social sciences (blue) and STEM (orange) occupy different regions in the network, with social science students appearing toward the lower right and STEM students tending more toward the top. Humanities (yellow) students appear on the right side of the diagram, interspersed among social science students. Students who have not declared majors (green) occupy the left side of the diagram adjacent to larger, more general, and lower-level courses; by definition, these are first and second-year students, since all students must declare a major by the end of their sophomore year. Multidisciplinary and mixed majors are scattered throughout the network, especially within the social science and STEM regions.

The network nearly constitutes a single component. Figure 2 does not show the 16 nodes (9 students, 7 courses) that appear in separate components. Taking these into account, 99.7% of the student and course nodes are situated in the main component (component ratio = .001).

The level of clustering in the liberal arts network is also high. The weighted clustering coefficient for the projected one-mode student-to-student network is .480. Simulations show that in a randomly wired network of the same size with the same density as the network that we examine here, we would expect to see a clustering coefficient of only about .03. The level of clustering in this subnetwork is thus more than an order of magnitude greater than would be expected at random.

The density of the student-to-student projection of the two-mode network is .040. There are 287,327 unique direct connections between pairs of students through classes.

With 3,806 students in the network displayed here, this means that, on average, a given student will have been in a classroom with about 150 other students by the time they have completed one round classes on their schedule.

In the liberal arts college network, most students can reach each other through very few steps. The average student-to-student distance via courses in the main component of the two-mode network is 2.2. As in the University-wide network, it is unlikely that any two liberal arts students would be enrolled in the same course, but highly likely that they would be enrolled in different courses that includes the same third party. The average student can reach 4.0% of the other students in just one step (i.e., via a shared course); 72.5% of the other students in two steps; and 99.4% in three steps. The diameter of the network is 5.

As in the full university network, many alternative geodesic pathways link pairs of liberal arts students together. A total of 94.2% of the student and course nodes are in the largest bi-component of this subnetwork, meaning that they are interconnected via multiple paths that are independent of each other and therefore operate through entirely separate students/courses. As in the University-wide network, the liberal arts college network is highly integrated.

This high level of connection is not dependent on the 18 courses that enroll 100 or more liberal arts students. If we exclude these courses, the proportion of students in the main connected component remains high at 98.2%. The average geodesic distance in the projected student-to-student network increases slightly from 2.2 to 2.5, while the average number of geodesics linking the students declines to 19.6. The share of students who can reach each other in just two steps declines from 72.5% to 47.1% without these large courses, but the number of students who can reach each other in three steps does not change appreciably: 96.9% without the large courses, compared to 99.4% with these courses.

Discussion

Initial reactions to universities' decisions to shift to on-line instruction were mixed, with some observers lauding the shift as a necessary step in the fight against the diffusion of COVID-19 and others criticizing it as an overreaction. Our results show that the course enrollment structure provides short chains of potential infection through which diseases can spread through college campuses via direct contact. Critically, course networks remain highly connected even in a liberal arts college, where classes tend to be smaller, and even if one excludes the largest courses from a face-to-face enrollment network. This implies that a hybrid model of instruction, wherein large courses are taught online and smaller courses are taught face-to-face, cannot resolve the challenge of course co-enrollment as a potential means of transmission.

There are several limitations of this study. First, we do not have transcript data for professional or graduate students, a small proportion of whom enroll in undergraduate courses or in courses that co-meet with undergraduate courses. Because of this data limitation, we likely underestimate the number of students with whom an average student comes in contact through their classes. Similarly, the network graphs in Figures 1 and 2 do not depict separate components consisting of graduate students who only enroll in courses with other graduate students, nor the bridges between populations created by courses that enroll both undergraduate and graduate students.

Second, we cannot assess whether our results generalize to other universities. Even universities of similar size differ in the number and type of majors that are offered, the extent to which majors and courses are concentrated in a single college or multiple colleges, and the degree of college autonomy in setting curricular requirements. We suspect that the "stadium" structure of the full enrollment network in Figure 1 characterizes most large American public universities (see also Israel et al 2020), and that the highly integrated subnetwork in Figure 2 is a reasonably good approximation of co-enrollment networks in smaller liberal arts colleges and universities. But, we cannot test these claims with data from one university.

Third, course enrollment networks will understate the social and physical connections among courses and students, particularly on residential campuses. They do not capture connections between courses that are created by sharing a physical space (e.g., two classes using the same room in succession), which – depending on the nature of the virus – could affect transmission through contaminated surfaces. They also do not capture the incidental contacts that occur between students in hallways, on quadrangles between class periods, or in the commercial areas that surround most colleges or universities. And, perhaps most important, course enrollment networks do not capture the many ways that students are connected outside of the classroom through advisors, friends, parties, athletics and other extra-curricular activities, or living situations. Given the multiplex ways that students come in contact with each other outside the classroom, the results of this paper are a conservative depiction of the extent to which college life facilitates contact, direct or indirect, among students.

At the same time, course enrollment networks may overstate the density of the networks through which a virus is likely to be transmitted. Most obviously, two students who are co-enrolled in a large lecture course may never come in close physical proximity to each other. Similarly, classes, particularly large ones, rarely achieve full attendance. Future work should consider factors such as physical space within a classroom or attendance rates to fine-tune estimates of how course enrollment networks may pattern the diffusion of a virus, a rumor, an idea, or anything else that can be transmitted through direct or indirect social contact on a college campus.

Finally, the network we have analyzed here is a static, aggregate representation of the overall course enrollment record. It does not capture the dynamics of movement between courses, and it ignores the sequence in which students attend each of their classes. This temporal sequencing may be consequential for epidemic spread. For example, if larger classes tend to meet earlier in the week and smaller classes and discussion sections later in the week, an infection that arrives with a student after a weekend away from campus may spread more quickly than if the reverse were true. There are other important subtleties here as well, such as the timing of repeated exposures and the order in which different courses meet in the same classrooms. The temporal and spatial sequencing of class meetings likely affects how many weeks of class are needed before maximum direct contact exposure among students is achieved. Future research is needed to undertake a more detailed analysis of these dynamics and their implications for the potential for an epidemic spread.

The college campus environment provides conditions that increase the risk of a propagated outbreak of highly infectious diseases. Our results show that the course enrollment structure provides multiple short chains of infection that potentiate the spread of viruses. They also show that a hybrid model in which large courses are taught on-line and smaller seminars are taught through face-to-face instruction would not appreciably reduce the interconnectedness of students in the full course enrollment network. In a smaller college, canceling the face-to-face sessions of large courses elongates indirect chains of contact slightly, but does not eliminate them. These results suggest caution in reopening colleges and universities for face-to-face instruction in response to the COVID-19 pandemic.

Methods

Our data come from Registrar's records of Cornell University, which enrolls approximately 15,000 undergraduate students, 6,200 graduate students, and 2,700 professional students. Each of the seven undergraduate colleges at Cornell admits its own students and sets its own graduation requirements. Although there is no common University-wide curriculum, colleges allow students to take courses outside their college, some of the smaller colleges outsource their introductory or required gateway courses, and many courses are cross-listed across colleges, meaning that a course taught in its originating department (the "parent" course) also has course numbers in other departments or colleges ("child" courses). This reduces structural fragmentation within the enrollment network.

The transcript data cover all undergraduate students who enrolled in courses in Spring of 2015, excluding those who spent the semester in a study abroad program or on academic hiatus. Given the upheaval caused by the COVID-19 response, we were unable to secure data for this paper for Spring 2019, the most comparable semester to Spring 2020. The University did not change appreciably in size or structure over these 4 years, so basic network patterns should be unaffected. However, our 2015 data may slightly underestimate the connectedness of current students through co-enrollment, given secular trends in undergraduate major selection toward computer science, where classes tend to be large, and away from the humanities, where classes tend to be smaller.

We used a course-level data file covering the liberal arts college and Cornell's on-line course roster to identify courses that are taught in locations other than the main campus (e.g., Cornell's Tech campus in New York City, Rome, Washington DC) or that are cross-listed. We exclude all courses that are not taught on the main campus and, analogously, all students who are taking their courses at an off-campus location. We assign all enrollments in cross-listed courses to the parent course. The analysis of the liberal arts college network only includes students who enrolled in this college, courses for which this college is the parent, and biology courses in the cross-college undergraduate program in biology. Table 2 provides descriptive statistics at the student and course level for both the University-wide and the liberal arts college analyses.

Table 2. Characteristics of Students and Courses in University and Liberal Arts College

	University		Liberal Arts College	
Students				
Number of students (*n*)	13,594		3,806	
Majors (%)				
Humanities, Arts, and Design	6.5		9.8	
Social Sciences	30.1		18.8	
STEM	43.2		23.8	
Multidisciplinary or Mixed	9.8		11.5	
Undeclared	10.4		36.1	
Mean (stdev) courses per student	5.3	(1.2)	5.1	(1.1)
Mean (stdev) co-enrolled students	571.2	(347.6)	192.9	(141.8)
*Courses**				
Number of courses (*m*)	3,797		1,467	
Mean (stdev) size	18.9	(41.1)	9.5	(18.5)
Number with 100-199 students	88		18	
Number with 200+ students	38		0	

NOTE: Data are from student transcripts, Spring 2015. Courses include Physical Education and other 1 or 2-credit course. Course enrollments in liberal arts college only include students from this college.

We also differentiate the students by their field of study. To keep the network graphs visually manageable, we collapse Cornell's more than 100 majors into five categories: humanities, fine arts, performing arts, and design; social sciences; STEM; multidisciplinary and mixed; and undeclared. The humanities, social sciences, and STEM categories include multi-major students whose majors all fall in the same field: for example, a student double-majoring in Biology and Chemistry would be coded as STEM. The "multidisciplinary and mixed" category includes students in "design-your-own" majors and in majors that explicitly require courses across two or more of the three major fields of STEM, social sciences, and humanities (e.g., [example suppressed for anonymity]. This category also includes multi-major students whose majors do not fall in the same field: for example, a dual-major in Biology and English would be coded as "mixed." Most of the students in the "undeclared" category are first- and second-year students in the liberal arts college and one other small college; students in the other colleges are typically admitted directly into a major or, in the case of engineering students, can safely be coded as STEM even if they haven't declared a specialty.

We use the transcript data to construct an affiliation matrix, **A**, which is a binary, two-mode matrix (see Borgatti and Everett 1997; Wasserman and Faust 1994). **A** contains information about a set $\{n_1, n_2,...,n_g\}$ of g = 13,594 students, who are arrayed down the rows of the matrix, and their ties to a set $\{m_1, m_2,...,m_h\}$ of h = 3,797 courses/sections, which are arrayed across the columns of the matrix. **A** contains g x h = 51,616,418 total cells. The cells in **A** contain a set $\{e1, e2,...,el\}$ of l lines between these two classes of nodes. A given cell in **A** – for example, $<n_1, m_1>$ – indicates whether student n_1 was enrolled in course m_1 (0 = "no," 1 = "yes").

Numerous structural features of network **A** have implications for direct or indirect epidemic spread of disease. One is the network's overall level of cohesion, which can be measured in several ways. We calculate two-mode network density (the number of ties present divided by the total possible); the number of ties in the largest component of **A** (that is, students/courses that are at least indirectly connected to each other through at least one path); as well as the component ratio of the overall network, which equals 1 if all nodes are in separate components and 0 if all nodes are connected in a single component.

We are also interested in the extent to which students themselves are interconnected, and how far removed from each other they are in the course network. To examine this, we create a one-mode projection of **A** (derived by multiplying **A** by its transpose, $\mathbf{A}^{\mathbf{T}}$), $\mathbf{A}_{\mathbf{P}}$, which yields a valued matrix that reflects the number of courses in which each student is co-enrolled with each other student (Breiger 1974). We dichotomize $\mathbf{A}_{\mathbf{P}}$, then use the resulting binary network to analyze the structure of (potential) student-to-student contact via shared courses.

We derive several measures from the projected network. The clustering coefficient, which measures transitivity (Holland and Leinhardt 1971; Watts and Strogatz 1998), indicates the extent to which students who are enrolled in courses with a common third party also tend to take courses with each other. Such clustering provides opportunities for reinforced (direct and indirect) contact transmission. We also examine the average geodesic distance between each pair of students, via their courses. A distance of 2, for example, means that while two students in are not enrolled in any courses together, they are both enrolled in different courses with some third student in common. Finally, we analyze *k*-step reach centrality (see Borgatti, Everett, and Freeman 2002), which reflects the proportion of other students in the network that a given student can reach in a given number, *k*, of steps.

Data and Code Availability

The data for this article were used with non-transferable permission from Cornell University and are not publicly available. Code for the final version of the manuscript will be posted on SocArxIv.

Ethical Compliance

We have complied with all relevant ethical regulations. The study is exempt from IRB review, per the guidelines posted by Cornell's IRB office.

References

Biancani, S., & McFarland, D. A. Social networks in higher education. *Higher Education: Handbook of Theory and Research* (ed. M.B. Paulsen) 151-214 (Springer 2013). DOI 10.1007/978-94-007-5836-0_4

Binson, D., Woods, W. J., Pollack, L., Paul, J., Stall, R., & Catania, J. A. Differential HIV risk in bathhouses and public cruising areas. *Am. J. of Public Health* **91,** 1482-1486 (2001).

Borgatti, S. P. 2002. NetDraw: graph visualization software. (Analytic Technologies, 2002).

Borgatti, S. P., & Everett, M. G. Network analysis of 2-mode data. *Soc Network* **19,** 243-270 (1997).

Borgatti, S. P., Everett, M. G., & Freeman, L. C. Ucinet for windows: software for social network analysis. (Analytic Technologies, 2002).

Breiger, R. L. The duality of persons and groups. *Social Forces* **53**, 181-190 (1974).

Caley, P., Philip, D. J., & McCracken, K. Quantifying social distancing arising from pandemic influenza. *J. R. Soc. Interface* **5**, 631-639 (2008).

Cooley, P. C., Bartsch, S. M., Brown, S. T., Wheaton, W. D., Wagener, D. K., & Lee, B. Y. Weekends as social distancing and their effect on the spread of influenza. *Comput. and Math. Org. Theory* **22,** 71-87 (2016).

Cornwell, B., & Schneider, J. A. Social venue range and referral chain impact: implications for the sampling of hidden communities. *PloS One* **12,** e0181494 (2017).

Dickison, M., Havlin, S., & Stanley, H. E. Epidemics on interconnected networks. *Phys. Rev. E,* **85,** 066109 (2012).

Frost, S. D.W. Using sexual affiliation networks to describe the sexual structure of a population. *Sex. Transm. Infect.* **83**, i37-i42 (2007).

Hill, A. L., Rand, D. G., Nowak, M. A., & Christakis, N. A. Infectious disease modeling of social contagion in networks. *PLoS Comput. Bio.*, **6**, e1000968 (2010).

Holland, P. W., & Leinhardt, S. Transitivity in structural models of small groups. *Comp. Group Stud.* **2**, 107-124 (1971).

Israel, U., Koester, B. P., & McKay, T. A. Campus connections: student and course networks in higher education. *Innovat. High. Ed.* (2020). https://doi.org/10.1007/s10755-019-09497-3

Kelso, J. K., Milne, G. J., & Kelly, H. Simulation suggests that rapid activation of social distancing can arrest epidemic development due to a novel strain of influenza. *BMC Publ. Health* **9**, 117 (2009). https://doi.org/10.1186/1471-2458-9-117

Laumann, E. O., Ellingson, S., Mahay, J., Paik, A., & Youm, Y. (eds). *The Sexual Organization of the City* (Univ. Chicago Press, 2004).

Luke, D. A., & Harris, J. K. Network analysis in public health: history, methods, and applications. *Annu. Rev. of Publ. Health* **28**, 69-93 (2007).

Morris, M. (ed). *Network Epidemiology: A Handbook for Survey Design and Data Collection* (Oxford Univ. Press, 2004).

Niekamp, A. M., Mercken, L. A. G., Hoebe, C. J. P. A., & Dukers-Muijers, N. H. T. M. A sexual affiliation network of swingers, heterosexuals practicing risk behaviors that potentiate the spread of sexually transmitted infections: a two-mode approach. *Soc. Network, **35,*** 223-236 (2013).

Oster, A. M., Wejnert, C., Mena, L. A., Elmore, K., Fisher, H., & Heffelfinger, J. D. Network analysis among HIV-infected young black men who have sex with men demonstrates high connectedness around few venues. *Sex. Transm. Dis.* **40**, 206–212 (2013). https://doi.org/10.1097/OLQ.0b013e3182840373

Piontti, A. P. y., Perra, N., Rossi, L., Samay, N., & Vespignani, A. *Charting the Next Pandemic: Modeling Infectious Disease Spreading in the Data Science Age*. (Springer, 2018).

Poletti, P., Ajelli, M., and Merler, S. Risk perception and effectiveness of uncoordinated behavioral responses in an emerging epidemic. *Math. Biosci.* **238**, 80-89 (2012).

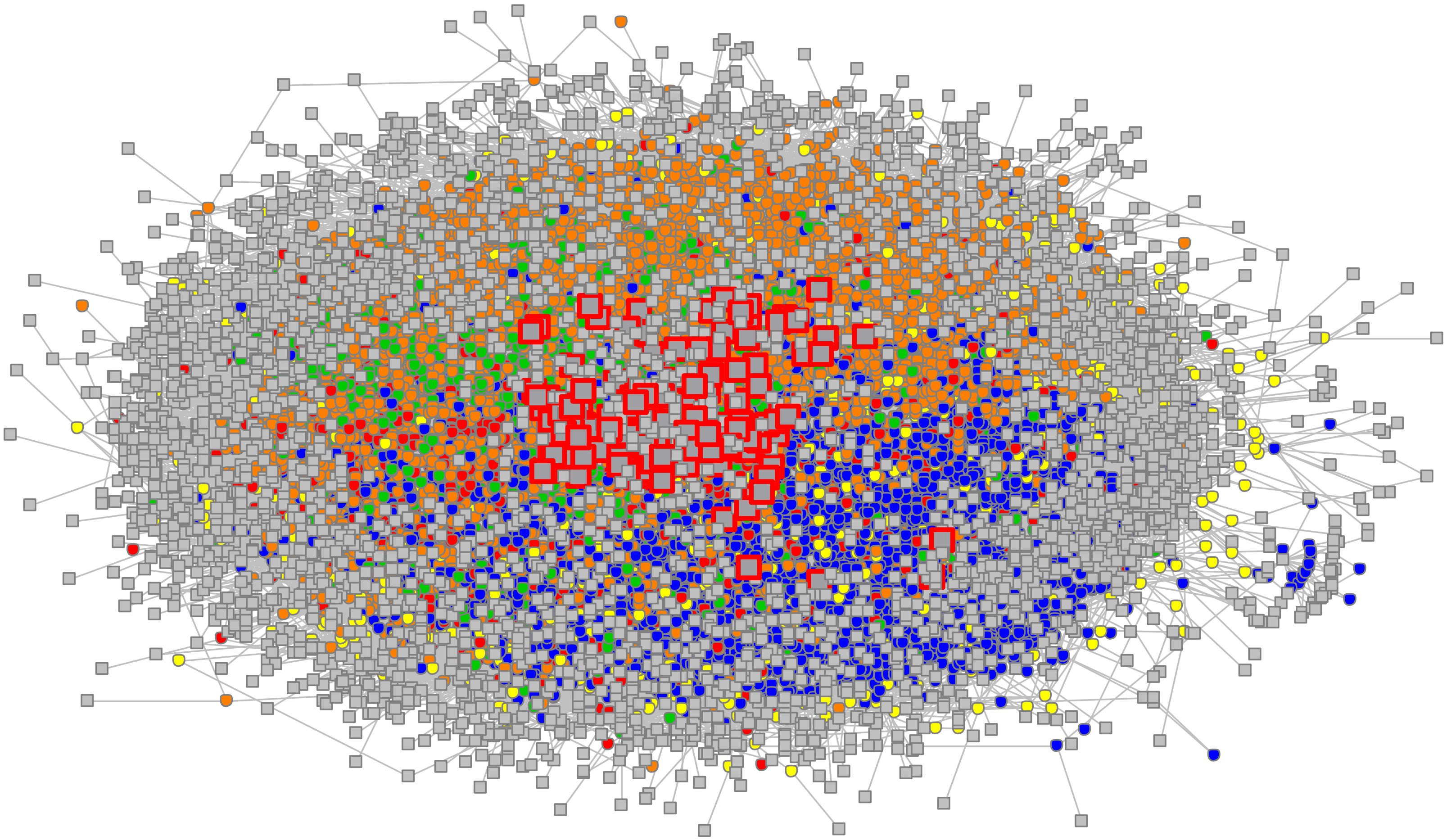
Valdez, L.D., Macri, P.A., & Braunstein, L.A. Intermittent social distancing strategy for epidemic control. *Phys. Rev. E* **85,** 036108 (2012).

Valente, T. W. 2010. *Social Networks and Health: Models, Methods, and Applications.* (Oxford Univ. Press, 2010).

Wasserman, S., & Faust, K. *Social Network Analysis: Methods and Applications* (Cambridge Univ. Press, 1994).

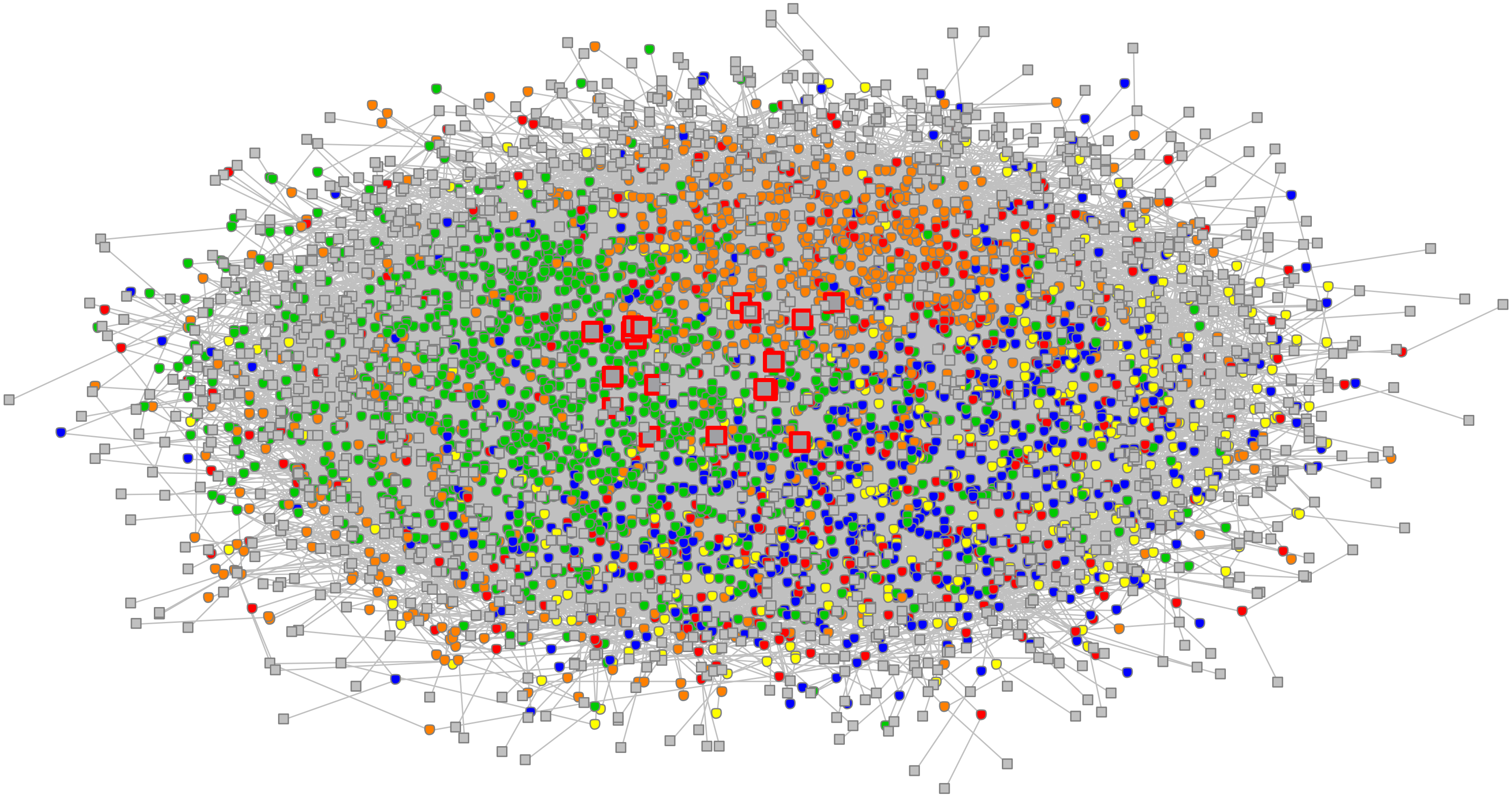
Watts, D. J., & Strogatz, S. Collective dynamics of "small-world" networks. *Nature* **393**, 440-442 (1998).

Figure 1. Main Component of the Course Enrollment Network for a Medium-Sized University (N = 17, 391 Nodes, Including 3,797 Courses and 13,594 Students).



Note: This network diagram was produced using the graph-theoretic layout algorithm in Netdraw (Borgatti 2002). Light gray squares represent courses; larger gray squares with red border indicate courses of ≥ 100 students. Students are represented by small circles with colors identifying their major(s): Yellow = humanities; Blue = social sciences; Orange = STEM; Red = multidisciplinary/mixed; Green = undeclared. Students' enrollment in particular courses is indicated with straight light gray lines. Note that this diagram excludes 146 nodes (38 courses and 108 students) that were not connected to this main component.

Figure 2. Main Component of the Liberal Arts College Course Enrollment Network for a Medium-Sized University (N = 5,273 Nodes, Including 1,467 Courses and 3,806 Students).



Note: This network diagram was produced using the graph-theoretic layout algorithm in Netdraw (Borgatti 2002). Light gray squares represent courses; larger gray squares with red border indicate courses of ≥ 100 students. Students are represented by small circles with colors identifying their major(s): Yellow = humanities; Blue = social sciences; Orange = STEM; Red = multidisciplinary/mixed; Green = undeclared. Students' enrollment in particular courses is indicated with straight light gray lines. Note that this diagram excludes 16 nodes (7 courses and 9 students) that were not connected to this main component.