Analysis and Methodology of Inhibiting COVID-19 Spread on a University Campus

Minhao Zhang[a], Geng Zhang[a], Shengxi Zhang[a], Zian Liu[a], Vitaly Ford[a] and Victoria Turygina[b]

[a]*Arcadia University, 450 S Easton Rd, Glenside, PA 19038, U.S.A.*

[b]*Ural Federal University, Prospekt Lenina, 51, Yekaterinburg, Sverdlovskaya oblast', Russia, 620075*

Abstract

Most of the states in the U.S. are slowly transitioning back to "normal", and educational institutions must weigh in the decision of maintaining the quality of the courses while protecting the health of students in the academic years ahead. We are interested in investigating the circumstances that would help schools stay open during COVID-19, creating safe educational conditions under such a severe situation. Our goal is to move a certain number of courses online to achieve a satisfactory infection rate most efficiently. At the same time, we attempt to maximize the number of face-to-face classroom experiences as most students prefer attending courses on campus over attending them online. In our model, we introduce three parameters to evaluate the risk of every course and determine the most suitable set of courses to be converted online. The parameters include Degree Centrality, Closeness Centrality, and Betweenness Centrality. Those parameters are aggregated in a rectified value. We describe the methodology of our approach and future work, in which we will conduct simulation and sensitivity analyses.

Keywords

COVID-19, university campus, centrality parameter, social network

1. Introduction

As COVID-19 has been quickly spreading across the world, public and private businesses, restaurants, and other gathering places have been forced to close down for a few months. Now that most of the states in the U.S. are slowly transitioning back to "normal", educational institutions must weigh in the decision of maintaining the quality of the courses while protecting the health of students in fall 2021. Many schools have announced that all courses will be converted to a synchronous/asynchronous online format to create social distances between students so that they are less likely to be infected by COVID-19, whereas others decided to select a hybrid/flexible approach.

Arcadia University has around 2,500 students (including approximately 700 graduate students) and 340 faculty members. On-campus, it has 15 main buildings housing different departments, dining halls, and other places. Even though Arcadia University is a small liberal arts university, like many other universities, it also faces the problems of whether the campus

IVUS2021: Information Society and University Studies 2021, April 23, 2021, Kaunas, Lithuania

mzhang_01@arcadia.edu (M. Zhang); gzhang_02@arcadia.edu (G. Zhang); szhang_01@arcadia.edu (S. Zhang); zliu@arcadia.edu (Z. Liu); fordv@arcadia.edu (V. Ford); v.f.volodina@urfu.ru (V. Turygina)

should reopen and how to protect the safety of students, faculty, and staff. We are interested in investigating the circumstances that would allow schools to open again in the academic years ahead, creating safe educational conditions under such a severe situation.

2. Background Information

One of the major safety concerns is raised by the fact that students regularly encounter many of their peers on campus throughout the day and gather together in groups outside of classrooms. Consequently, we decided to evaluate the degree of interconnectedness/centrality of students with respect to a potential viral outbreak. Our research goal is to develop data-driven policies and simulate virus spread on the Arcadia University campus to provide a safer environment for students to take face-to-face classes by converting a targeted number of the courses to an online format.

Given that schools would follow the guidelines and procedures defined on federal, state, and local levels, we assume that students would follow safety practices and only some courses need to be transferred online. However, based on the Laurence Steinberg's article "Expecting Students to Play It Safe if Colleges Reopen Is a Fantasy" [1], we take into consideration that the students' age group could be defined as "high risk-takers" and, therefore, students are prone to allow themselves a leeway, especially in informal group settings, meaning that there is a possibility of students not wearing masks and not following the safety protocols when they should be doing so.

We realize that it is not possible to completely suppress the virus if it is already on campus, even though it would be feasible to contain it, assuming that everyone follows the safety guidelines. However, we hope that this research can provide a data-driven approach that could be used as a reference for school decision-makers to make teaching plans.

By studying Cornell University's research findings [2], we found that the Cornell University team used transcript data describing 3 enrollment networks that connect students via classes, creating certain conditions that simulate the virus spread. According to epidemiological studies, they assumed that the cancellation of face-to-face courses would reduce the spread of the virus in which case a bipartite network needed to be created. Referring to their data acquisition and implementation methodology, we built bipartite social network graphs based on students' schedules from Arcadia University's fall semester data to observe the degree of correlation among students and classes. Also, by analyzing statistical data, we simulated the spread of the virus and removed the highly concentrated courses to observe whether moving some courses and certain students online will significantly change the virus's transmission results.

In this study, we added extra parameters to simulate different situations that may occur on campus. For example, we determined the virus spread with or without masks and the impact of opening the dining hall.

The core motivation for this research is to answer the following questions:

1. Would schools be able to stay open during COVID-19 by moving high sensitivity classes and particular students online?
2. How many students would be infected compared to the model without converting classes to an online format?

Our simulation (presented in the follow-up article [3]) is based on the assumption that the school has reduced the infection rate as much as possible and can frequently test students on campus. We started our simulation based on the policies presented in [2], specifically focusing on the initial value of R_0 [4].

2.1. Notable characteristics of our approach

1. We have students' actual course information meaning that our data are real rather than hypothetical, leading to more realistic predictions during the simulation analysis.
2. The parameter R_0 [4], defined as the metric of how many other people an infected individual could infect, directly reflects the impact of our measures, such as "moving certain courses online." After obtaining an R_0 adjusted during our simulation, we can quickly retrieve updated results, including the number of infected people, and allow for running experiments numerous times with different parameters.
3. In order to evaluate and determine a satisfactory result, we compare the infection rate of students attending classes face-to-face with the general infection rate in the local region. The general infection rate in Pennsylvania represented by R_0 is 3 [5] based on CDC [6] at the time of writing this paper. Such a comparison allows us to estimate the risk of getting infected on campus versus staying home and taking classes online. If the infection rate of students in the school is lower, we consider it as a satisfactory result.

3. Data Source

To analyze the interconnectedness among students on campus, we used their course schedule information for fall 2020. We designed a program in Python available at [7] to obtain all the course registration information and clean the data. After cleaning the data (e.g., removing students who have not registered at all) from the raw data we obtained, we created our data source based on the information of 1,816 undergraduate students including the registered courses, time period of each course, and the classroom locations. An example of the finalized data format is presented in figure 1.

Our dataset contains 1,816 students and 689 courses. We anonymized student IDs representing each student because we focus on reducing the spread of the virus rather than analyzing specific personal information. As shown in figure 1, we extracted the course codes and time periods of the students, allowing us to cluster them accordingly. Based on these data, we can determine which courses have the highest degree of centrality to decide if they need to be moved online. Also, since we keep the data in a *JSON* format, it is straightforward to add other appropriate parameters during the simulation. A more detailed analysis of the collected data is available in the follow-up paper [3].

4. Methodology

We present a mathematical modeling network for COVID-19 at Arcadia University. Besides the data related to the students who have not registered for the fall semester and unavailable data

```
"1": {
    "Classes": ["AN120\/Lecture\/1", "CS203\/Lecture\/1", "CS203L\/No fee lab\/1", "MA203\/Lecture\/1", "MA221\/Lecture\/2"]
    "Times": [
        ["MWF", "11:00 AM - 12:05 PM"],
        ["MWF", "9:45 AM - 10:50 AM"],
        ["T", "12:20 PM - 2:00 PM"],
        ["TR", "2:10 PM - 3:50 PM"],
        ["MWF", "1:30 PM - 2:35 PM"]
    ],
    "Rooms": ["Unknown", "Arcadia Univ, Boyer Hall, Room 014", "Arcadia Univ, Boyer Hall, Room 014", "Unknown", "Unknown"]
},
"2": {
    "Classes": ["AN120\/Lecture\/1", "MA203\/Lecture\/1", "MA221\/Lecture\/2", "MA341\/Lecture\/2"],
    "Times": [
        ["MWF", "11:00 AM - 12:05 PM"],
        ["TR", "2:10 PM - 3:50 PM"],
        ["MWF", "1:30 PM - 2:35 PM"],
        ["MWF", "12:15 PM - 1:20 PM"]
    ],
    "Rooms": ["Unknown", "Unknown", "Unknown", "Unknown"]
},
"3": {
    "Classes": ["BI332\/Lecture\/1", "BI332L\/Lab\/1", "HN393L\/Lecture\/1", "PY221\/Lecture\/1", "PY221L\/No fee lab\/3"],
    "Times": [
        ["MWF", "12:15 PM - 1:20 PM"],
        ["T", "1:00 PM - 3:50 PM"],
        ["MWF", "9:45 AM - 10:50 AM"],
        ["MWF", "11:00 AM - 12:05 PM"],
        ["T", "9:00 AM - 10:50 AM"]
```


Figure 1: A sample of the finalized data format.

of graduate students, we gathered data of 1,816 undergraduate students out of about 2,500 undergraduate and graduate students to support our model. Most graduate programs at Arcadia are online so graduate students do not affect our simulation.

In comparison with the universities having a large student population, Arcadia has a relatively small class size that is typically less than 20. This fact greatly reduces contact among students. Also, Arcadia has enough classroom resources to avoid a possible infection caused by classes happening one after another in the same room. Based on these conditions, our model is designed to provide suggestions and help university leadership with the decision-making process when they consider whether or how to bring students back for a residential fall semester.

We are aware that an accurate prediction is not an attainable goal. So, we established a model with R_0, the most commonly used parameter in the epidemic transmission, and determined the initial R_0 for schools under sufficient protection obeying the recommendation from CDC [8]. Our simulation goal is to reduce R_0 as much as possible by decreasing the interconnectedness among students to achieve a reasonable situation, looking for the possibility of the majority of students to take classes and keep from being infected.

Our approach focuses on finding courses with the highest sensitivity coefficient through the social network and shifting as few courses as possible into an online format to reduce students' direct contact and achieve an acceptable infection rate. We also use the method of controlling variables to find the parameters that have the most impact on the epidemic model and formulate relevant policies for it. Moreover, we remove large courses with students from different majors to reduce the possibility of inter-departmental infection spread. It is worth noting that our model is based on the scenario that students reduce the number of other forms of contact as much as possible besides extracurricular activities; that is, students comply with the CDC policy [9].

The simulation outcome is optimistic. Even under the very pessimistic initial R_0 value set to 3.2, we can still get a satisfactory R_0 after moving only top-10 high sensitivity courses online, resulting in less than 316 students being infected before the end of the semester meaning that very few students would potentially need to quarantine. We chose to base the simulation results on moving 0, 5, and 10 courses online because it fits in with the approach we take to evaluate our proposed policy measures. Based on our calculations in the simulation analysis, removing 0, 5, 10 classes from the network corresponds to the number of infected students in the face-to-face semester, which is higher, almost similar, and lower respectively than that of a fully online semester. In the simulation section of our follow-up article [3], this will be clearly demonstrated.

Public places like cafeterias with heavy traffic should be closed down, otherwise, the spread of the epidemic would never be controlled. We have provided an alternative for these necessary closures and it will be presented in [3].

4.1. Bipartite social network graphs of collected data

We collected the course information of all students and established the network diagram between "students and courses" and between "students and students." In figure 2, we identified 18 components including one big component and 17 small components. We focus our attention on the one big component.

We made two diagrams with the network of "students and students" and the network of "students and courses" respectively. First of all, in the figure of students only, there are numerous connections among students represented by the red lines, meaning that if the school opens without taking any effective measures, the spread of the epidemic will be significant to close it down within a week.

In the second graph of "students and courses" (figure 3, we use asterisks to represent classes, and the more students there are in the class, the larger the symbols. Most of the courses are interconnected via students' connections.

4.2. Virus spread parameters and assumptions

We used three different R_0s to simulate the spread of the virus in optimistic, nominal, and pessimistic scenarios. The main difference in those scenarios is the number of days the students can potentially infect others before they are quarantined. On a relatively small campus like Arcadia, there is a possibility to conduct regular virus testing, so that the latent or asymptomatic patients can be quarantined as soon as possible, thus the number of days the infected students could transmit the virus would be significantly reduced. In the paper [2], the researchers estimated the R_0 in three different situations as well, namely optimistic, nominal, and pessimistic, defining the seriousness of the virus spread accordingly. However, due to the large number of graphs per scenario, we focused our attention on the pessimistic R_0 value as it describes the worst infection spread.

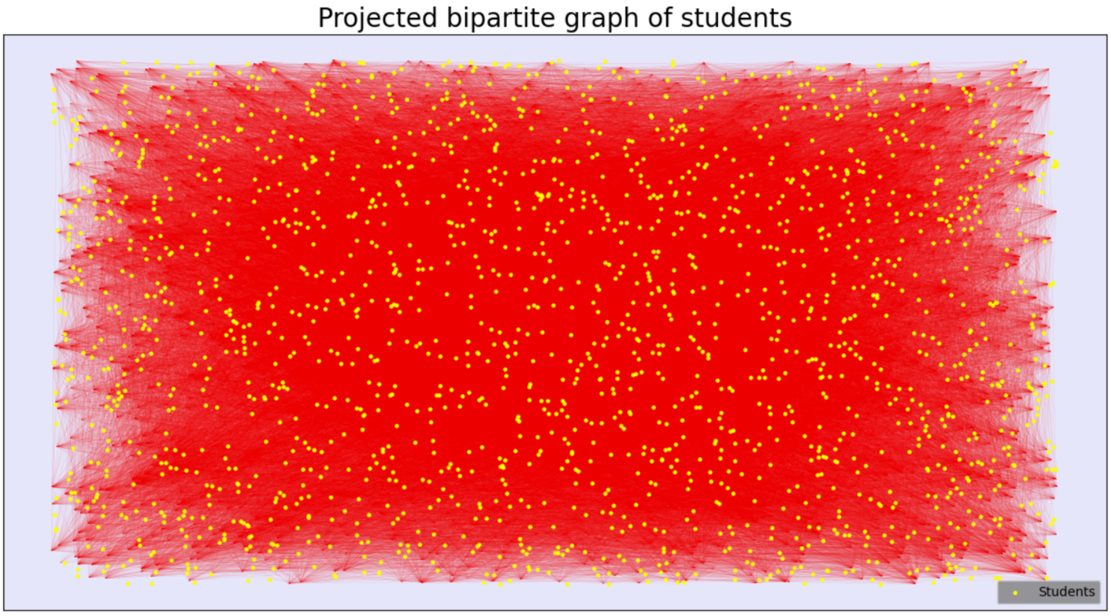

Figure 2: Students-students social network graph.

We relied on the definitions and parameter settings established in the paper [2] and defined as the following.

The expression for R_0 is defined as:

$$\begin{aligned}(Days\ infectious\ presymptoms+ \\ Expected\ days\ free\ postsymptoms) \times Contacts/day \times \\ Probability(infection\ transmissions|contact),\end{aligned} \tag{1}$$

where

$$\begin{aligned}Expected\ days\ free\ postsymptoms = \\ percent\ asymptomatic\ \times duration\ of\ symptoms + \\ percent\ symptomatic \times \\ 1 \div daily\ selfreporting\ probability.\end{aligned} \tag{2}$$

For the parameters' settings, we calculate R_0 according to the procedure above:

Optimistic: $R_0 = (2.5 + (27.3\% \cdot 10 + 72.7\% \cdot 1/18\%)) \cdot 8.3 \cdot 2.6\% = 2$
Nominal: $R_0 = (3 + (47.8\% \cdot 12 + 52.2\% \cdot 1/18\%)) \cdot 8.3 \cdot 2.6\% = 2.5$
Pessimistic: $R_0 = (3.5 + (68.3\% \cdot 14 + 31.7\% \cdot 1/18\%)) \cdot 8.3 \cdot 2.6\% = 3.2$

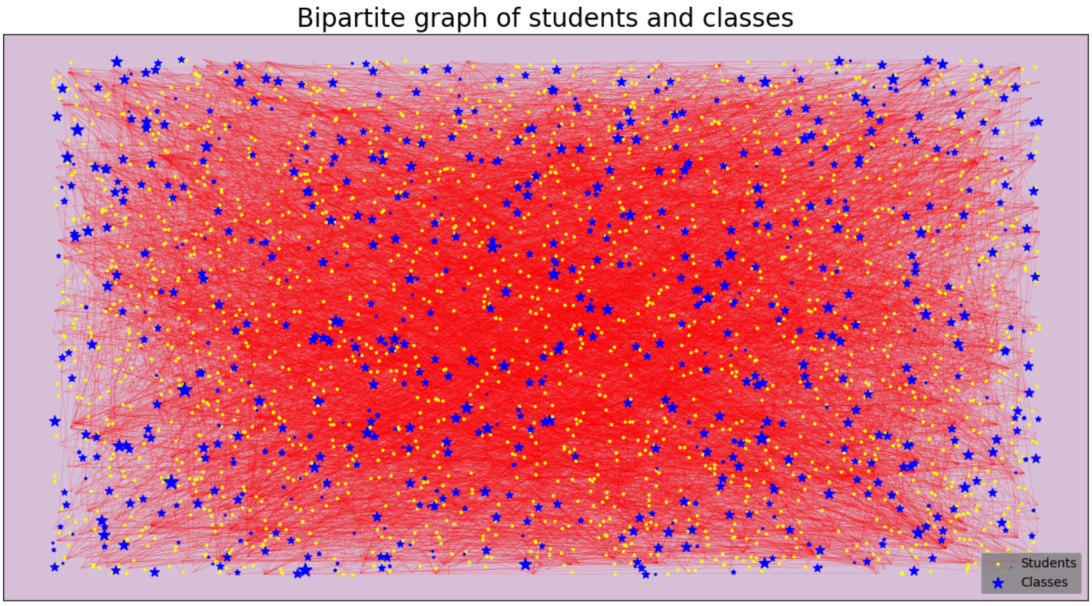

Figure 3: Students-courses social network graph.

These data were used to simulate the spread of the virus at Cornell University [2].

Note: As Arcadia has fewer students (no more than 20) in one class, the above-mentioned numbers are slightly higher than Arcadia-related numbers.

We assume that everyone will comply with the policy and wear masks. Arcadia's administration has issued a document explaining the requirements for the face-to-face meetings, including a requirement to wear a mask in public places with a note that failure to comply will constitute a violation of the code of conduct and could result in disciplinary actions [10]. Classrooms and other public places will be effectively and regularly cleaned. In the paper [2], the researchers have also simulated a similar scenario.

4.3. Goals of the proposed policies

We will turn a certain number of courses into an online format to achieve a satisfactory infection rate most efficiently. At the same time, we attempt to maximize the number of face-to-face classroom experiences as most students prefer attending courses on campus over attending them online.

4.4. Data analysis

In our model, we introduce three parameters to evaluate the risk of every course and determine the most suitable set of courses to be converted online. The parameters include Degree Centrality, Closeness Centrality, and Betweenness Centrality.

To calculate these three parameters, we need to understand what story they tell in our specific scenario:

- *Degree Centrality* identifies the important nodes with many connections.
- *Closeness Centrality* identifies the important nodes that are close to each other.
- *Betweenness Centrality* identifies the important nodes that are located on the shortest paths between other nodes on the network.

In our model, we assume that the most important nodes have many connections and the shortest path between any two nodes is short enough for a fast virus spread.

Under the above assumptions, the evolution of the parameters mentioned above is modeled by the following equations [11].

Degree Centrality:

$$C_{dge}(v) = \frac{dv}{|N| - 1} \tag{3}$$

Closeness Centrality:

$$C_{close}(v) = \frac{|N| - 1}{\sum_{u \in N\{v\}} d(u, v)} \tag{4}$$

Betweenness Centrality:

$$C_{btw} = \sum_{s,t \in N} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}} \tag{5}$$

where:

N is the set of the nodes in the network.
dv is the degree of node v.
$d(v, u)$ is the distance between nodes v and u.
σ is the number of the shortest paths between nodes s and t.
$\sigma(v)$ is the number of shortest paths between nodes s and t that pass through node v.
Note: all the nodes belong to the set of courses.

In our social network, Degree Centrality describes the number of classes connected to a specific student as well as how many students are connected to a particular course. Closeness Centrality explains how a course connects to other courses in the school and represents the sum of the shortest paths (the start points are the students in the class) to all students. Betweenness Centrality is similar to Closeness Centrality, except that it represents the importance of courses that are located on the path for the other courses to become connected. When a course has a high degree of Betweenness Centrality, it becomes the Suez Canal in our network: it greatly reduces the shortest path for other courses to become connected.

After obtaining the three centrality parameters of all nodes, we can get the rectified value of each node. The node with a high rectified value usually has more connections, closer distance, and stronger connectivity with other nodes.

We define the function of rectification as the following:

Table 1
Courses in sorted order according to the rectified values.

Courses	Degree Centrality	Closeness Centrality	Betweenness Centrality	Rectified Centrality
FT100	0.021065	0.255487	3.108477e-02	5.520307
FA102L	0.020668	0.252835	2.801977e-02	5.426778
BI242	0.018680	0.238833	1.126589e-02	4.955504
BI204	0.016296	0.239755	1.264442e-02	4.476854
CH201	0.014706	0.235259	9.679976e-03	4.132704
ID101	0.014308	0.243659	1.677117e-02	4.093348
BI323	0.014308	0.241198	1.191070e-02	4.081287
BI201	0.014308	0.239155	8.843107e-03	4.071271
BI329	0.014308	0.228339	8.360022e-03	4.018256
BI204	0.013911	0.234859	6.089622e-03	3.969685
ID101	0.013514	0.231316	1.048219e-02	3.871791
BI102	0.012717	0.247199	1.418215e-02	3.788587

$$C_{ret}(v) = \frac{C_{dge}(v)}{\bar{C}_{dge}(v)} + \frac{C_{close}(v)}{\bar{C}_{close}(v)} + \frac{C_{btw}(v)}{\bar{C}_{btw}(v)} \tag{6}$$

Through the rectification method, we remove the influence of weight on the pure numerical values and look at the aggregate result across all centrality parameters.

The outcome of arranging the courses from high to low according to their rectified values is presented in table 1.

5. Conclusion

In this paper, we presented our initial analysis of the academic course data, demonstrating that the rectified centrality value (based on the degree centrality, closeness centrality, and betweenness centrality) can identify courses that would be beneficial to switch to the online modality. We described the methodology and algorithms taking into account such information as course location, time, and the number of enrolled students. In the paper that will follow this article [3], we will expand on the knowledge obtained in this research and conduct simulations with sensitivity analysis and recommendations for building a safer environment on campus.

Acknowledgements

We would like to thank the Computer Science and Math Department, Arcadia University, for providing support and help during this research.

References

[1] L. Steinberg, Expecting Students to Play It Safe if Colleges Reopen Is a Fantasy, 2020. URL: https://www.nytimes.com/2020/06/15/opinion/coronavirus-college-safe.html.

[2] J. M. Cashore, N. Duan, A. Janmohamed, J. Wan, Y. Zhang, S. Henderson, D. Shmoys, P. Frazier, COVID-19 Mathematical Modeling for Cornell's Fall Semester (2020), URL https://people.orie.cornell.edu/pfrazier/COVID 19 (2020).

[3] M. Zhang, G. Zhang, S. Zhang, Z. Liu, V. Ford, V. Turygina, Application of Data-Driven Measures for Impeding COVID-19 Spread at an Academic Institution, in: accepted, 26th International Conference Information Society and University Studies - IVUS, 2021.

[4] B. Ridenhour, J. M. Kowalik, D. K. Shay, Unraveling R_0: Considerations for public health applications, American journal of public health 108 (2018) S445–S454.

[5] L. Matrajt, T. Leung, Evaluating the effectiveness of social distancing interventions against COVID-19, MedRxiv (2020).

[6] Centers for Disease Control and Prevention, Searching CDC site, 2021. URL: https://search.cdc.gov/search/index.html.

[7] Zhang, Minhao, Coronavirus: A Case Study at Arcadia University, 2020. URL: https://github.com/zhangminhao00/Coronavirus-A-Case-Study-At-Arcadia-University.

[8] Centers for Disease Control and Prevention, Considerations for Institutions of Higher Education, 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/community/colleges-universities/considerations.html.

[9] Centers for Disease Control and Prevention, How to Protect Yourself & Others, 2021. URL: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick.

[10] Arcadia University, Conduct and Compliance, 2020. URL: https://www.arcadia.edu/covid-19-health-and-safety-plan/conduct-and-compliance.

[11] P. Hage, F. Harary, Eccentricity and centrality in networks, Social networks 17 (1995) 57–63.