The Cutting Edge, Vol 3, No 2 (2021)

Mitigation of Disease in Primary Schools: Implications for School Policies during the COVID-19 Pandemic

Caitlyn McGinley*
Stanford University

Minha Khan*
Stanford University

Lina Karamali
Stanford University

Avrum Noor
Stanford University

Introduction

Children in America spend an average of 1,000 hours in school each year. This accounts for one-sixth of their total waking hours (Department of Education 2008). Given this extensive time and exposure, school environments play a vital role in community health. This is especially true when it comes to the spread of infectious disease. Integral characteristics of traditional school environments, such as the high mixing rates of school children, general architectural environment, and the culture and hygiene of school-aged children, can lead to large outbreaks of viral diseases such as influenza (Gemmetto et al., 2014). This is particularly concerning because school-age children are an at-risk group to viral disease. Behavioral determinants affect children who are unaware of health risks around them and are typically unable to take actions to reduce their risk. Physiological determinants such as less developed immune systems and less capacity to resist vector-borne diseases and developmental determinants like immature organs make children more vulnerable to disease and damage in their early years (Gemmetto et al., 2014).

Past proposals for disease mitigation in schools include the closure of schools when an outbreak occurs. However, such measures come with high associated social and economic costs, making alternative, less disruptive interventions highly desirable (Gemmetto et al., 2014). Recently, disease modeling has allowed for an opportunity to design

**indicates equal authorship*

models of micro-interventions (Stehlé et al.) This is made possible through school-based viral surveillance, in particular, using high-resolution contact network data from school environments to model disease spread. Such surveillance can be an essential part of managing community health, as it provides early warnings for outbreaks (Gemmetto et al., 2014) and supports early action disease mitigation.

The aim of our project is to use high-resolution contact network data from a primary school and apply social network analysis techniques to understand the spread and mitigation of disease within it. In doing so, we hope that such methods will inform the way policymakers and healthcare officials model, understand, and address disease spread in schools to ensure that neither the education nor the health of students is compromised. In the sections to follow, we will introduce our dataset and share our findings using network visualizations and descriptors, community detection methods, and centrality measures. Finally, we will discuss how our findings can inform our understanding of school policy development during a pandemic and what implications that has for the COVID-19 crisis.

Data

Our data captures the temporal network of contacts between primary school students and teachers in a private catholic school in France. The data was collected over the span of two days: October 1st and October 2nd, 2009. Interactions between individuals were measured through sensors which were set up in the school. A total of 77,602 contact events were recorded between 242 individuals (232 children and 10 teachers). The data file identifies the unique IDs of students and teachers, their classes, and the active face-to-face contact interval period measured in seconds. Additionally, it identifies the grade that the individual studies in or teaches. There is no personal information about the students and teachers available.

Findings

Visualizing Disease Pathways

As can be seen in Figure 1A, visualizing the two-day data results in a highly dense and clustered network. The black nodes below indicate teachers, whereas the colored nodes each represent students from a particular classroom. The clustering of students showcases these classrooms, each with their own class teacher. Some preliminary observations that can be made are that the teachers play a central role in the network, there's distinct clustering that might lead to distinct communities with further analysis, and lastly, that while there are two separate classes per grade, the classes at the same grade level still operate in close connection to one another across all grades.

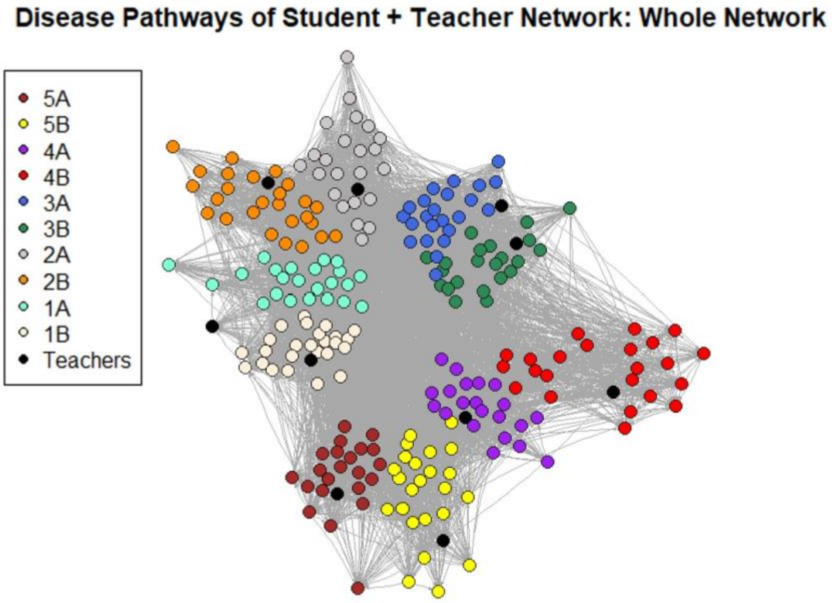
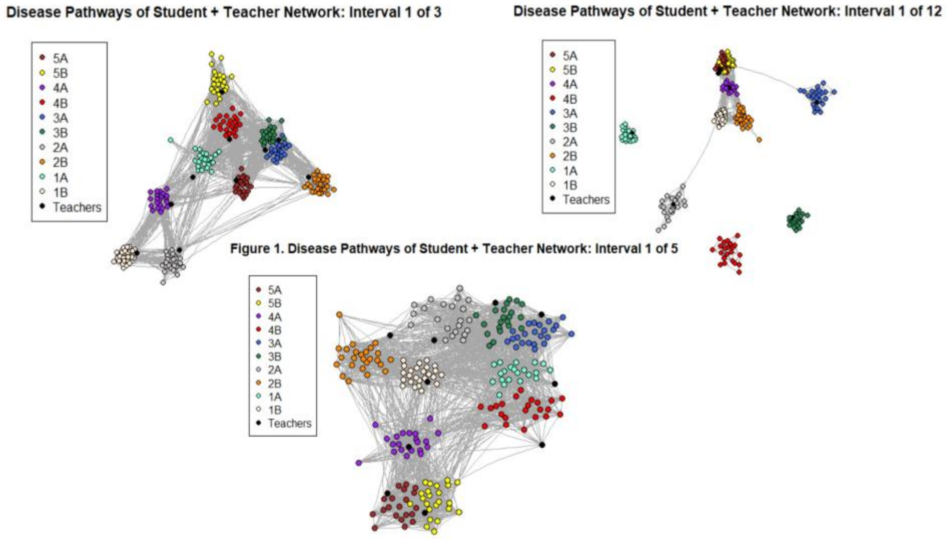

Disease Pathways of Student + Teacher Network: Whole Network

FIGURE 1A. Mapping Full Network Disease Pathways.

Given the density of the figure above, it is important to further transform the data to make meaning of the network above. Firstly, we experimentally divided the data at different time intervals—three, five, and twelve—in order to simplify the temporal dataset. This was done in an effort to conserve data, as there are concerns about deleting vertices with a small out degree. The initial groupings, which are just cutting the network into thirds, fifths, and twelfths, respectively, were chosen arbitrarily. However, after visualizing the different intervals, as seen in Figure 1B, the strengths of each grouping became apparent.


Disease Pathways of Student + Teacher Network: Interval 1 of 3

Disease Pathways of Student + Teacher Network: Interval 1 of 12

Figure 1. Disease Pathways of Student + Teacher Network: Interval 1 of 5

FIGURE 1B. Mapping Full Network Disease Pathways at Different Intervals.

Each interval gives a different visualization of the network. Interval 1 of 3 showcases the first third of the network, meaning that the network has been divided into thirds and we are viewing the first snapshot of the network. At Interval 1 of 3, the nodes in each cluster are densely packed together, not giving us any adequate information about the interactions in each class. Dividing the network into twelfths begins to break the temporal into distinct components, creating local bridges, and clustering in the network. Each class becomes separated from one another, which makes sense, as the temporal data is divided by twelve as it has a higher chance of capturing moments when students are isolated in their class together versus interactively mixing out at recess or in the halls. Lastly, we divide the network into fifths. At this interval, we can begin to visualize the dynamics between students a bit more clearly, both within their respective classes as well as within the school in entirety. We can see that some students also play a more peripheral role in their interactions with their peers. It also becomes clear that teachers from the same grade level interact with one another as well as their counterparts' students. Overall, we decided that dividing the data into fifths was the best way to visualize the data set.

In order to provide a quick snapshot of the relationship between nodes across the time interval, we created the visualizations of the network for four of the five time intervals in Figure 1C. These networks retain the initial clustering format of Figure 1A while also allowing us to see the interactions between these clusters in a narrower time window. Our visualizations suggest that while contacts occur mostly within each class, there is significant interaction between students of different classes across all time intervals.

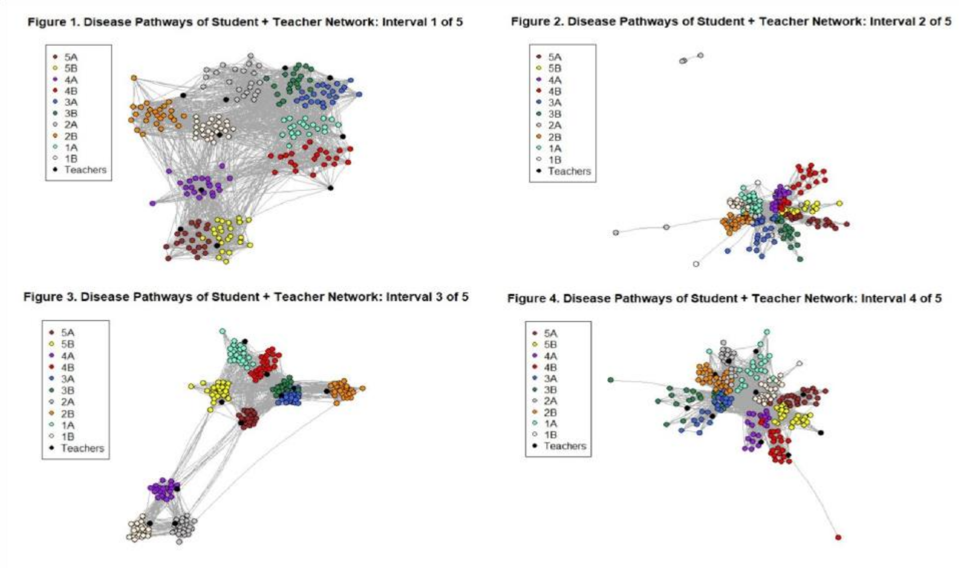


FIGURE 1C. Mapping Interval-Level Disease Pathways.

While we do not have metadata for the time of day, the temporal data was collected based on the interactions between students, hence, we can make educated predictions. For instance, it appears that Interval 1 of 5 might capture a highly-interactive activity that involves the entire school, perhaps a school-wide assembly, while intervals 2 and 4 possibly illustrate interactions at recess between students and teachers. This is hypothesized as we see more isolated components as well as local bridges, indicating students socializing with one another in their preferred friend groups. Furthermore, Interval 3 of 5 shows strong triadic closures between class 5A, 2A, and 1B.

Perhaps during this time, older students were mentoring younger students in the classroom in a "big buddy" like scenario. Again, it seems that Interval 3 of 5 is capturing something more similar to classroom time than recess given the tightly-knit clustering of the classes.

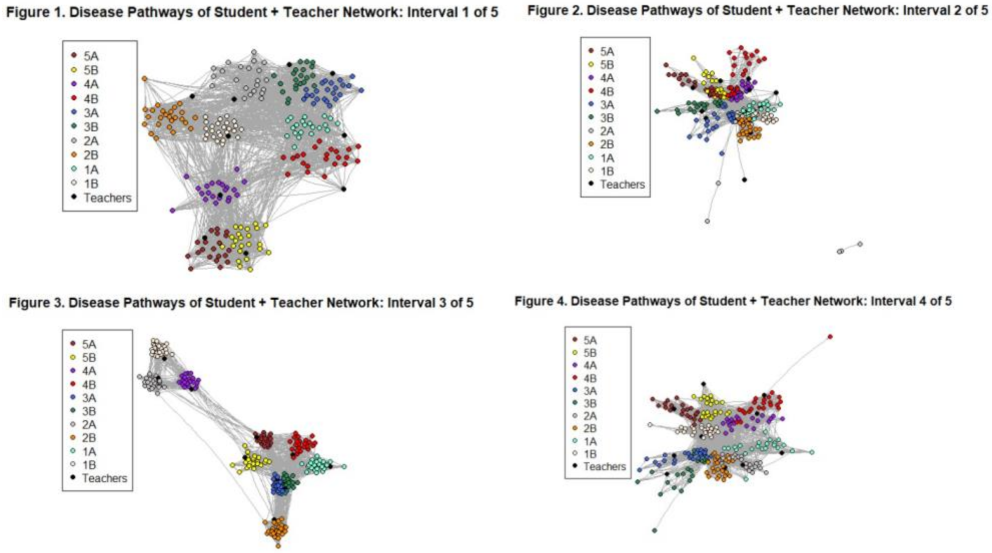


Figure 1. Disease Pathways of Student + Teacher Network: Interval 1 of 5

Figure 2. Disease Pathways of Student + Teacher Network: Interval 2 of 5

Figure 3. Disease Pathways of Student + Teacher Network: Interval 3 of 5

Figure 4. Disease Pathways of Student + Teacher Network: Interval 4 of 5

FIGURE 1D. Mapping Interval-Level Disease Pathways with Set Layout.

In addition, we visualized both the network both with a set layout, meaning each individual node stays in place across each interval. The initial layout is more dynamic and uses Fruchterman-Reingold (FR), which draws graphs in an aesthetically-pleasing way. FR graphs position the nodes of a graph in two-dimensional or three-dimensional space so that all the edges are of more or less equal length and there are as few crossing edges as possible, by assigning forces among the set of edges and the set of nodes, based on their relative positions, and then using these forces either to simulate the motion of the edges and nodes or to minimize their energy. In other words, FR sets each node in the optimal place for the specific interval. However, by creating a set layout, you can see how individual nodes change across time because they're always in the same spot, which is not the case for FR. However, since we do not have information about nodes at an individual level, this type of visualization is not as optimal as the FR layout, as if you look closely, the set layout is slightly denser and more amorphous looking than the FR version.

Network Descriptors

There were 242 nodes in the network. On average, each child has 323 contacts per day with 47 other children. The average daily interaction time of each child is about 176 minutes. Most of these contact periods are short, but longer periods of contact are also present in the data.

Figure 1 provides insight into the descriptors of the network to better understand the relationship between different nodes and the structure of the network. Given that our data was collected over a two-day period, some of the descriptors of the network do not capture the real-world interaction between nodes with precision. In Figure 1, we provide the descriptors for the entire network, alongside the descriptors for a single interval (1 of 5) to provide a more realistic description of the interactions between nodes. Given that the interval is 1 of 5, it showcases the relationship between nodes during the earlier hours of the day. This interval was chosen at random to showcase the network at any single point in time and provide a more realistic picture of how nodes interact and organize. Similar analyses can be run for other intervals to understand how relationships between nodes vary throughout the day.

Measure	**Full Network (Longitudinal)**	**Single Interval (Snapshot)**
Graph Density	2.16	0.06
Mean Degree	1039.45	29.75
Diameter (unweighted)	4	7
Path Length	1.73	2.15
Global Clustering Coefficient	0.48	0.45
Local Clustering Coefficient	Max: 0.04	Max: 0.96
Out-Degree Node	Max: 2121	Max: 45

In-Degree Node	Max: 2214	Max: 47

FIGURE 2. Network Descriptors for Whole Network vs. Interval.

The graph density was 2.16. Density of a social network examines how connected the nodes in a network are to one another. The density calculated for this network is more than 1.0, which means that of all possible edges among the nodes, most nodes interacted multiple times with one another. This is plausible because of the time frame of two days over which the data was collected. This also adds to the graph's overall visual density, as visualizing the entire network becomes amorphous. Hence, this is why we decided to visualize the graph at intervals. In interval 1 of 5, however, nodes interacted with 6% of all possible nodes. This suggests that in a short period of time, student interaction is much less than across a longer temporal period. However, the mean degree is very high amongst the two-day period- 1039.45- and, more surprisingly, also high as a snapshot. Within a single interval, on average, a student was interacting with 29 other students. The in-degree and out-degre measures support this notion of students being highly sociable. In this network, the diameter (without weights) is 4. This means that the maximum distance from one student or teacher to another is 4. The average path length was 1.73, suggesting that the average number of steps in the shortest path between any two nodes is 1.73. In interval 1 of 5, both the diameter (7) and the average path length (2.15) were longer given that nodes did not interact with each other as often in a single time period as they did over the course of two days. Overall, these measures suggest that there is frequent and physically close interaction between nodes in the networks. This complicates the mitigation of disease as such contacts suggest that spread of disease between nodes may be a highly contagious and timely phenomenon

The global clustering coefficient measures the extent to which nodes in a graph are likely to cluster together in the network. It is the probability that your neighbors know each other; the likelihood that your friend's friend is your friend. The likelihood of a closed triplet in this network is over 48%. Given that many of the students are classmates, or at the least, schoolmates, this figure is reasonable. The figure is similar when we look at a snapshot measure of the graph (44%). Meanwhile the local transitivity ranged from 0.00035 to 0.04301. This suggests that the students and teachers do not have very tight-knit groups of friends and rather are commonly interacting with multiple nodes, as opposed to being highly clustered with a single group of nodes across time. At the interval level, we observe tight-knit groups, while in contrast, the overview of the interactions reveals no tightly-knit clusters, but rather, an amorphous group. This can be explained by the nature of the intervals. The intervals are snapshots in time, and therefore, can capture moments of close-contact interaction. When the scope is larger, within the whole school-day, the brief moments of close-contact interaction are not as prominent.

Given that our only attribute information is whether the node is student or teacher, we cannot make a statement about homophily. It is interesting to note that when we do take a snapshot of the local transitivity, the figure significantly increases (with a maximum of 0.96), suggesting that in short periods of time, students do tend to interact within their cluster more commonly. An explanation for this is the difference in the type of interactions between nodes throughout different time periods (intervals) during the day. The activities students are engaged in and the amount of space available to them during those activities will indicate the local transitivity score. For example, when students are in class, we suspect higher local transitivity versus when they are dispersed out at recess. Given that our single interval is 1 of 5, indicating that it is a snapshot of student activity during the earlier hours of the day, we can conclude that it is capturing the local transitivity when students are all in a single classroom with limited space and high interaction between one another.

The conclusions we draw from these descriptors are that the nodes in this network are highly sociable and connected- whether we view it across a two-day period or a single interval. Disease mitigation efforts must take into account the interconnectedness of the students and policies on social distancing must consider the sociability of primary school students. Regarding vectors of disease, in terms of the identities (students vs teachers) of individuals who are “super-spreaders” such queries will be addressed in the Centrality section.

Community Detection

Factors to consider when it comes to analyzing diseases include how infectious the disease is with regards to proximity and how far the disease can spread. A disease is considered infectious if it can spread from one person to another who are in frequent contact and in the same proximity. Therefore, when it comes to studying how infectious a disease is, we are concerned with students and teachers that belong to the same community where close proximity is frequently maintained. When it comes to studying how far a disease can spread, we are concerned with studying students and teachers across different communities where we want to see whether distance plays a factor. Our analysis of the data thus far reveals that interactions between nodes are frequent and proximal. This section will further explore the patterns in node interactions.

Because no information is given on the types of communities present in the school, it is suitable to use community detection methods to identify the different communities within the school. We will examine the communities formed by four community detection methods: Edge-

Betweenness, Fastgreedy, Walktrap, and InfoMap. Generally, community detection methods allow the network to be partitioned into different components with the highest modularity Q, which means that students and teachers with similar connections are grouped together. This is useful to identify the hotspots of infections, since diseases tend to be more infectious and ever-present around people within the same community where they share similar connections. Hypothetically, the clustering of diseases would coincide with the presence of communities and therefore, coincide with the clustering of nodes in the network.

The following shows the number of communities/clusters arrived by each of the four methods (without performing any cut_at() operations on the community objects) at each interval and its modularity.

Interval	***Community Detection***	***No. of clusters***	***Modularity***
1	Edge-Betweenness	40	0.32
	Walktrap	8	0.56
	InfoMap	10	0.78
	Fastgreedy	8	0.55
2	Edge-Betweenness	71	0.09
	Walktrap	19	0.30
	InfoMap	24	0.57
	Fastgreedy	8	0.29
3	Edge-Betweenness	3	0.46
	Walktrap	9	0.66
	InfoMap	10	0.82
	Fastgreedy	9	0.66

4	Edge-Betweenness	92	0.19
	Walktrap	15	0.36
	InfoMap	18	0.59
	Fastgreedy	7	0.36
5	Edge-Betweenness	39	0.11
	Walktrap	12	0.43
	InfoMap	19	0.68
	Fastgreedy	9	0.42

FIGURE 3A. Table of Results for Community Detection

It is interesting to note that each of these four methods arrive at very different numbers when it comes to detecting the communities present in the network. This may be due to the difference in how these four methods detect communities. Judging by the number of communities present by eye, there are approximately 7-8 communities in each interval. Therefore, if this were the case, then the fastgreedy method is the most accurate in separating the network into communities.

The following shows how these four methods separate the network into the clusters:

1. Edge-Betweenness Method

Since Edge-betweenness models the number of shortest paths passing through an edge, it helps identify which interactions among the school population help spread the virus in the shortest path possible. Instead of trying to construct a measure that tells us which edges are the most central to communities, the Edge-betweenness algorithm focuses on edges that are most likely "between" communities

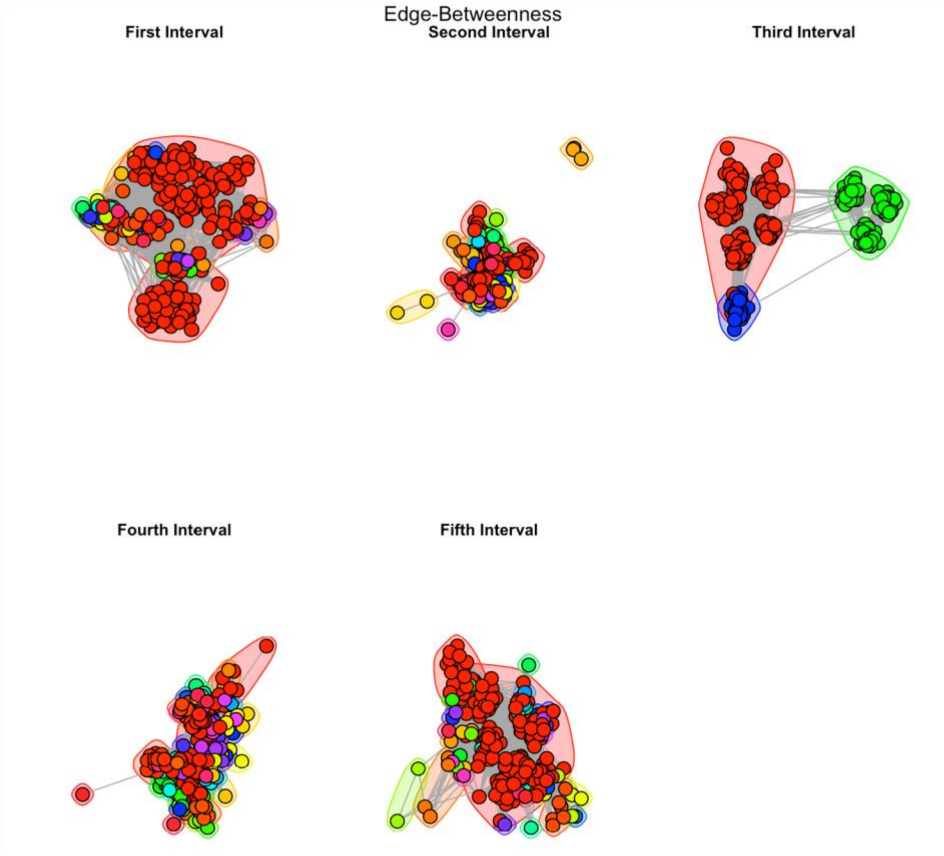


FIGURE 3B. Edge-Betweenness Method

From the figure above, we see the presence of one large main community throughout the five of intervals, which is highlighted in red. Edges going to and from the red community to the other communities have the highest edge-betweenness, meaning that in the event of an outbreak, the red community serves as the greatest gateway for the spread of virus among the network. This implies that the most effective method to infect the entire school network is by infecting a node in the red community or a community nearby. The large red community surrounded by smaller communities suggests a possible social construct at play—a social structure within the school. We hypothesize that the red community might represent the upper grades within the school, possibly 3rd through 6th while the smaller communities are the lower grades such as Kindergarten, 1st, and 2nd grade. In theory, it would make sense that the upper grades would share more "between" connections than the lower grades.

2. *WalkTrap Method*

Based on walktrap method community detection in the figure below, the sizes of communities change drastically throughout the time intervals.

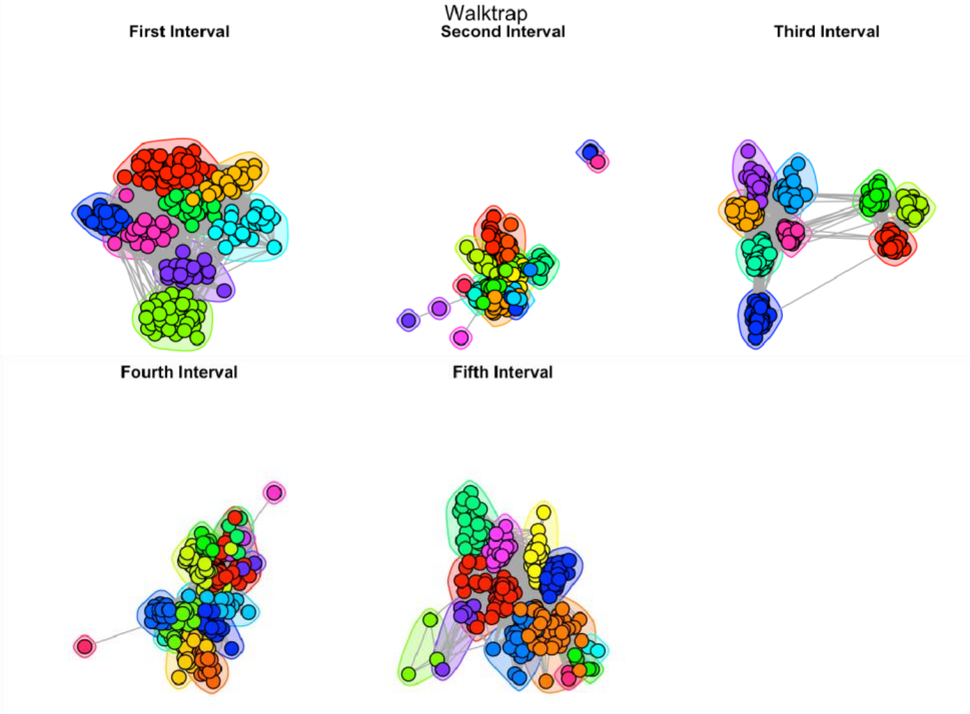


FIGURE 3C. Walktrap Method

If people in the school were to interact "randomly" within the school, their interactions would be analogous to the random walks used within the walktrap method. The number of communities formed by this method is between 9-19 communities. The general idea of the Walktrap Method is that if you perform random walks on the graph, then the walks are more likely to stay within the same community because there are only a few edges that lead outside a given community. Hence, this is why we see more distinct communities in the Walktrap method as compared to the Edge-Betweeness method. The fact that there is no one large main community implies that people in school actively participate in a variety of communities. The random walks are likely to stay within the same classroom cohorts as only a handful of edges lead to contacts outside the classroom, such as siblings or friends not within the same class or grade. This theory was also supported by the network's local-clustering coefficients. Due to how dynamic the communities formed are as a result of such participation, in the event of a virus outbreak within the school, tracing the source of the infection would be extremely hard.

3. InfoMap and FastGreedy

Based on the communities formed by InfoMap and the fast-greedy method, we see that nodes with similar interactions belong to the same community. Therefore, if an infection were to occur, InfoMap and the fastgreedy method would be the best methods to find other nodes that may be infected. This can be done by finding other people that are represented by the nodes in the same community as the person who is infected.

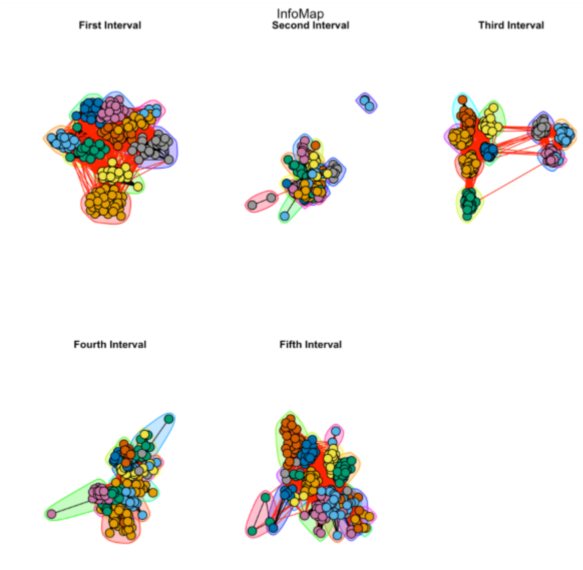


FIGURE 3D. InfoMap Method.

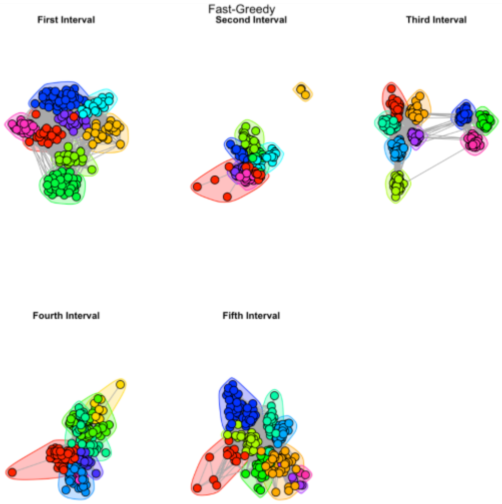


FIGURE 3E. Fast-greedy Method.

Centrality

In previous sections, we have adopted a structural (network- and community-level) analysis of the spread of disease in a primary school. In this section, through measures of centrality, we will conduct a node-level analysis to better understand the role a single individual plays within the spread of disease in a network. We will examine centrality through degree centrality, betweenness centrality, closeness centrality, Eigenvector centrality, and PageRank centrality. Each of these measures provides a unique analysis and perspective on how a disease would spread in our network, whether through having many connections, having lots of influence, etc. The nodes that have the highest centrality measures in each of these areas will provide us with valuable information about how specific nodes may perpetuate the spread of disease more than others, thus allowing us to develop micro-interventions for disease mitigation.

As we portioned our data into five intervals, we analyzed the centrality measures for each of the five intervals as well as for the complete data.

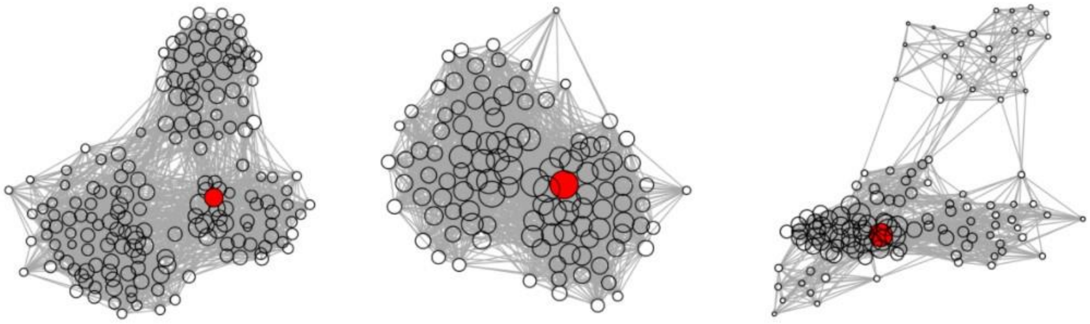
	Node: Degree Centrality	**Node: Betweenness Centrality**	**Node: Closeness Centrality**	**Node: Eigenvector Centrality**	**Node: PageRank Centrality**
Interval 1	1890: 51	1551: 1340.1	1426: 0.0003	1673: 1.000	1922: 0.0661
Interval 2	1673: 74	1700: 813.88	1426: 0.0001	1673: 1.000	1916: 0.0582
Interval 3	1551: 59	1653: 1861.9	1426: 0.0004	1551: 0.9824	1909: 0.0497
Interval 4	1851: 70	1761: 718.43	1426: 0.0002	1730: 1.000	1922: 0.0759
Interval 5	1907: 51	1594: 1340.1	1426: 0.0009	1822: 1.000	1922: 0.0668
Full Data	1695: 2532	1708: 401.24	1426: 0.0025	1665: 1.000	1922: 0.0849

FIGURE 4.

Figure 4 shows the nodes with the highest centrality measures for each interval and each respective measure of centrality (approximated for clarity). We will analyze this data by discussing each measure of centrality individually and understanding its implications for our research. The figures that follow will visualize the centrality for each interval as well as for the entire dataset.

1. *Degree Centrality*

Degree centrality measures how well connected a node is; i.e., the node with the greatest number of connections has the highest degree centrality. In our study, this node is the student or teacher with interactions with the greatest number of people. In Fig. 4, we can see the numbers of the nodes with the greatest degree centrality in each of the intervals and in the full data. Interestingly, the node with the highest degree centrality changes in each of the intervals and in the full data. However, the nodes are all students, which makes sense, as students are usually more likely to interact with more people than are their teachers (though they all interact during class, students likely also interact with one another outside of class). They are also almost all of different grades. The difference in the nodes with the greatest degree centrality tells us that those who have many interactions do not necessarily maintain the same degree of interactions consistently; that is, just because one node has the greatest degree centrality during one interval does not mean that that node will always maintain the highest degree centrality. In fact, the node with the highest degree centrality in the full data did not have the most in any one interval.



Degree Centrality Interval 4 **Degree Centrality Interval 5** **Degree Centrality**

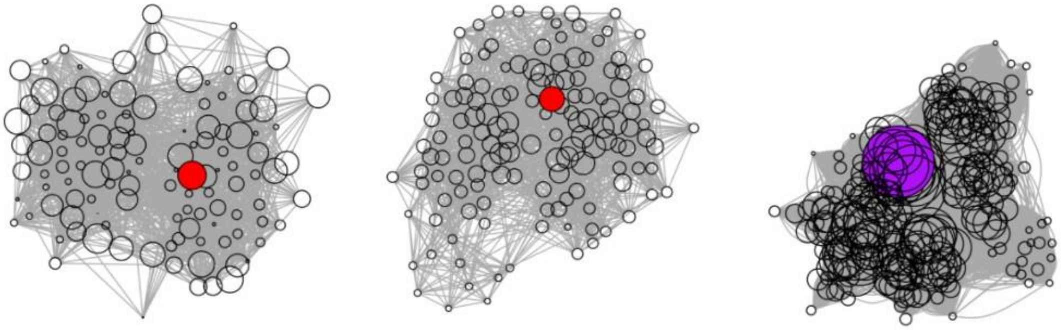


FIGURE 5A: Degree Centrality.

This data lets us identify hubs in the network through which disease might spread. The nodes with the greatest degree centrality measures are also significant ways in which the disease could spread. This follows logically, since the more people one knows, the more places there are for that person to spread a disease. So if any of the nodes with the highest degree centralities over the span of the data were to get infected with a disease, those nodes would likely be epicenters of disease spread to their numerous connections. These nodes are highlighted in red in the interval data plots, and in purple in the full data plot in Fig. 5A. The plots exclude nodes with degree centralities under 25 for the sake of visual clarity (though the full data plot does not). Unsurprisingly, there seems to be a trend among the nodes with the highest degrees to be densely connected with other nodes, including other nodes with high degrees. This shows us visually where the disease might spread, and is consistent with our centrality results.

2. Betweenness Centrality

Betweenness centrality measures how many times a node provides the shortest path between two other nodes. Just as a high betweenness in a node could signal high conveyance of information, so too can it warn of high risks of disease spread. For example, if a student is the link between two other students (especially if they are the only link between those students), the spread of a disease through that student could mean a substantial increase in the disease in the overall network. This is especially true when the node is a bridge or a local bridge and connects multiple communities. If two communities are connected only by a single node, it only takes the infection of that node to spread disease to both these separate communities.

Referring to Fig. 4, we see that again, different nodes have the highest betweenness centralities for each of the intervals and for the full data.

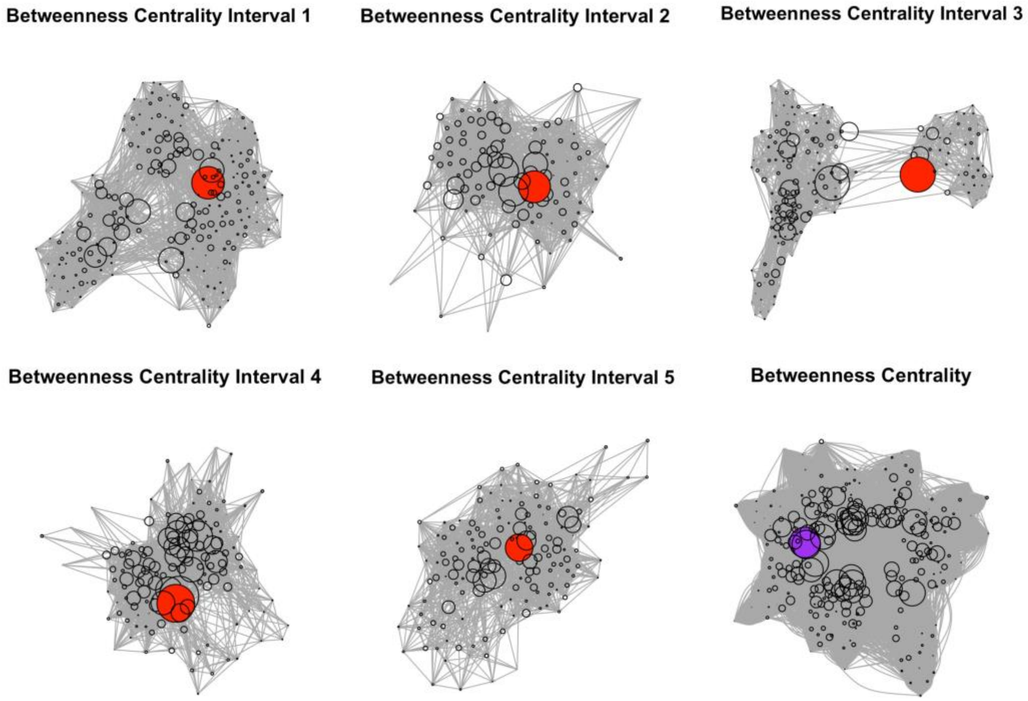


FIGURE 5B: Betweenness centrality.

In Fig. 5B, we can somewhat observe this phenomenon of betweenness. The plots again exclude some nodes in the lowest range of values for the sake of visual clarity (as do the following plots, save the full data plots). Some of the highlighted nodes are more clearly of high betweenness; for example, in Interval 3, the node with the highest betweenness seems to be a kind of bridge, thus making it valuable as a connection among the nodes and signaling its potential for infecting other nodes. These nodes with high betweenness often have the shortest paths between two other nodes, and in terms of disease this means that they could easily spread the disease to other nodes or communities. Interestingly, node 1551, the node with the highest betweenness in Interval 1, has the highest degree in Interval 3. This node is a 2A student, but because all the nodes of highest betweenness are of different grades, it is difficult to draw any conclusions about this node based solely on its attributes. Its prevalence among centrality measures, though, suggests that certain nodes are more connected overall in more ways than one (i.e., in degree centrality, betweenness, etc.), which in turn tells us that certain nodes are more pivotal in the spread of disease on multiple planes.

3. Closeness Centrality

Closeness centrality is another measure that gives us valuable information about our network. The higher a node's closeness centrality, the more easily that node can access other nodes. In our situation, greater ease of access can also mean greater ease of infection. *Fig. 4* tells us that there is one node with consistently the highest closeness centrality throughout the intervals and full data. Node 1426 is a student in grade 5A, shown in *Fig. 5C* as the highlighted nodes.

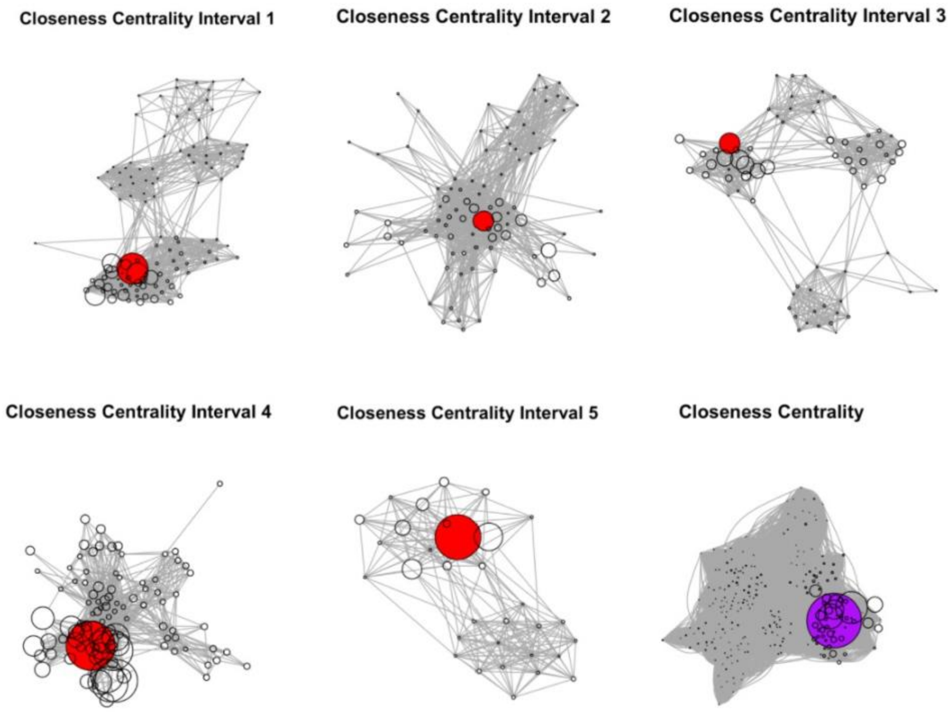


FIGURE 5C: Closeness centrality.

The plots are shown at slightly different scales to best represent the data, but we can see that in most cases, the nodes with the highest closeness centralities are clustered close to one another (represented by the bigger nodes in the plot). Since nodes with high closeness centralities have relatively short distances to other nodes and can easily spread information or disease, it makes sense that they might be more densely connected or clustered. Because node 1426 has consistently high closeness measures, it is an essential node to the spread of disease. If one node has consistently the highest ease of access to other nodes, it will be a consistent player in the spread of disease in our network.

4. Eigenvector Centrality

Eigenvector centrality is another interesting measure with regards to potential infection in this network. Eigenvector centrality tells us how important a node is based on whom it knows. Interestingly, according to *Fig. 4*, there are several nodes with high Eigenvector centralities that also have had high levels of other centrality measures. For example, nodes 1673 and 1551 both had the highest degree centralities during some interval. There did not seem to be any trend among these central nodes regarding grade level; however, this does tell us that not only are these nodes well connected, but they are also important according to their connections.

Eigenvector centrality can be very significant in the spread of disease through a network for social reasons, in addition to any simple connectivity reasons. A node with a high

Eigenvector centrality can use its importance or influence to affect other nodes or influence the nodes with more power. A high Eigenvector centrality shows that the node has important connections and is important in itself, and both these aspects may be leveraged to change the network. In the example of COVID-19, a node with more influence could tell others to wear masks (or not wear masks) and thus influence the spread of disease for the better or worse.

Similarly, that node could decide to host a party or enforce social distancing, and because of its importance and influence, these decisions could greatly affect the rate of infection.

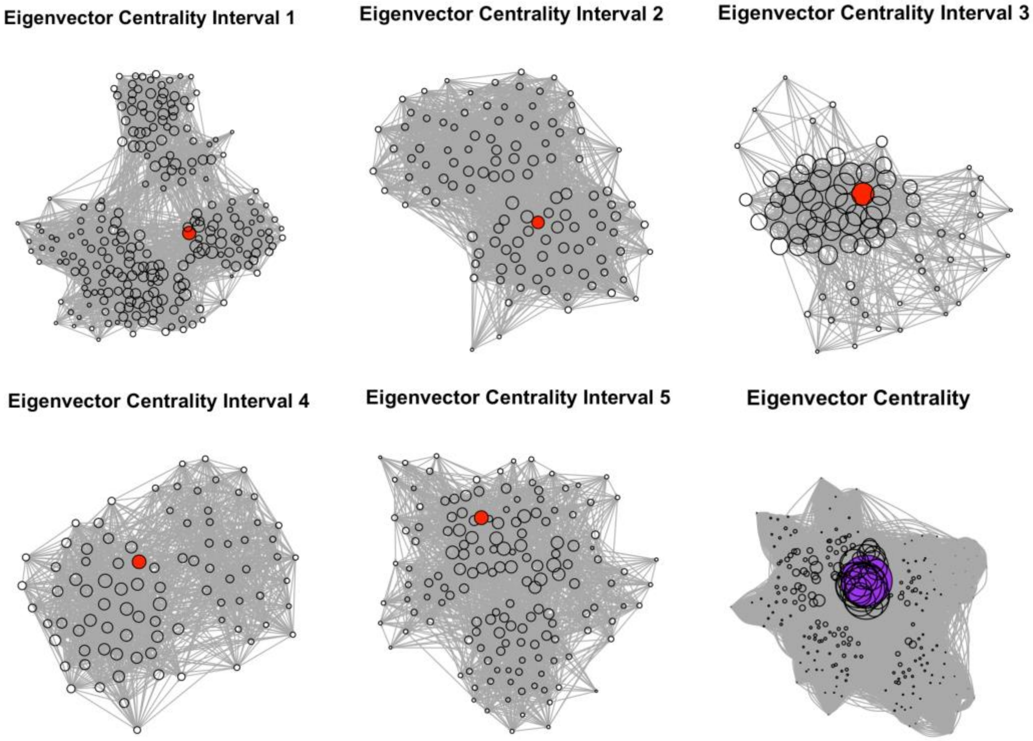


FIGURE 5D: Eigenvector centrality.

As we can see from *Fig. 5D,* nodes with high Eigenvector centrality often cluster with other nodes of high Eigenvector centrality. This is unsurprising, since the importance of a node's connections add to its own importance, and the important nodes must therefore be more or less interconnected. Thus, the clusters of nodes in Eigenvector centrality plots have important influence in how the rate of infection continues in a network.

5. PageRank Centrality

Very similar to Eigenvector centrality, PageRank centrality tells us that the more important a node's connections are, the more important that node is. PageRank is a variant of the Eigenvector centrality which applies to directed networks. Interestingly, node 1922 has the highest PageRank centrality for four of the six categories for which we measured centrality.

Node 1922 is a 3A student that evidently has consistently important connections.

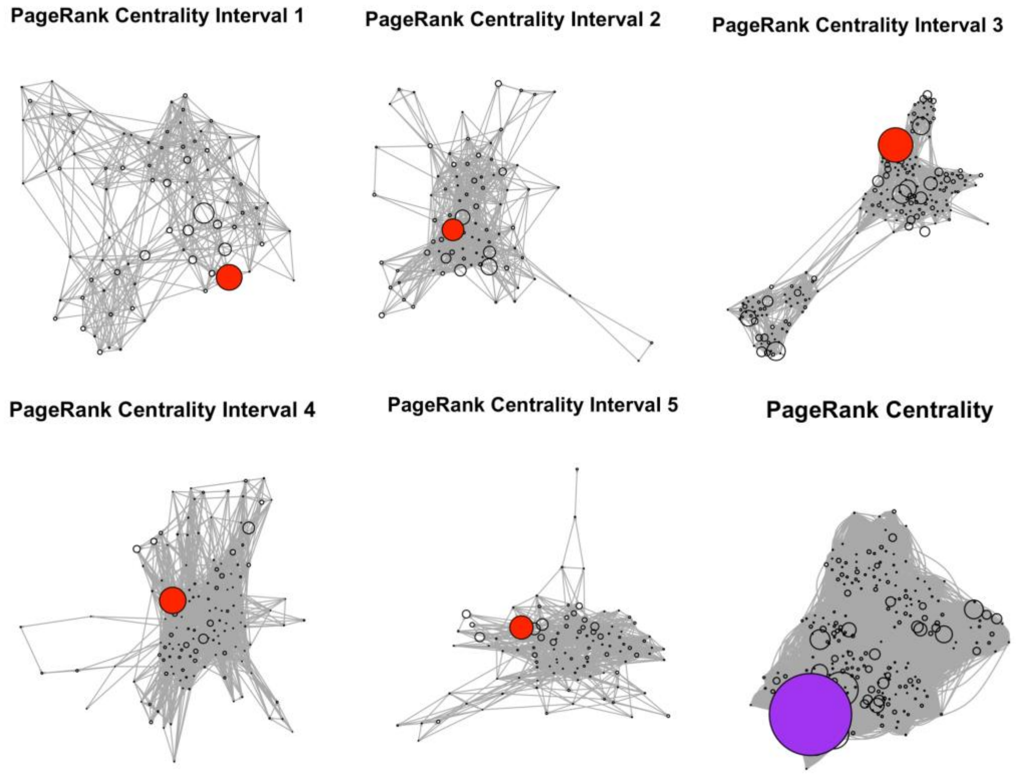


FIGURE 5E.

The plots for PageRank centrality are interesting in that most of them contain nodes that are barely connected to the network and may be only connected by local bridges. These nodes tend to generally be ones with low PageRank centrality. This makes sense, because the fewer the important connections that a given node has, the lower its PageRank centrality and the more isolated it will be in the plot. Overall, PageRank centrality tells us that node 1922 is substantially important in the spread of infection among nodes, since it is almost always the node with the highest PageRank centrality and thus the most importance. Similarly with Eigenvector centrality, a node like this one might have the power to influence others or use its importance to make change, whether beneficial or harmful.

We can see by looking at various centrality measures that each of the nodes with the highest respective centrality measures have different risk factors for spread of disease. By looking at centrality measures, we can understand how certain nodes might be more influential or might be a common vessel for the spread of disease. For example, when a node that knows the most people gets sick, that node is likely to infect more people than other nodes, and so on. These measures also give us a look into nodes' popularities or roles within social groups, and this can also help us understand the spread of disease within specific networks better. Though there was no observable pattern of centrality (with regards to attributes, etc.), we observed several nodes that were important in different ways, such as those nodes that appear more than once as having the highest centrality measure. This suggests that it is not just a matter of quarantining students of a particular class or grade, or identifying a single node and removing it. Instead, we see that multiple nodes are central in this network, which means that social distancing as a whole is likely required to mitigate disease spread.

Implications for COVID-19

During the Covid-19 pandemic, much of normal life has come to a standstill. Nationwide, school closures have become commonplace as researchers warn that school environments play an important role in the community spread of infectious diseases. This is attributed to high mixing rates of school children. Our findings support this claim; through social network analysis, we identify the primary school networks to be highly dense and sociable. This understanding of patterns of movement in behavior and interaction between school-going children is the first step in developing models which can mitigate disease within their communities. Our dataset provides high-resolution contact network data from school environments that allows epidemiologists to implement disease modeling, design models of micro-interventions to mitigate disease, and compare the outcomes of alternative mitigation measures.

Currently, in the United States, the closure of schools has become an efficient mitigation strategy, but has been shown to have never negative outcomes for children both socially and academically (Gemmetto et al. 2014). In addition, the physical closure of schools and the uptake of digital schooling has proven to be associated with high social and economic costs, making alternative, less disruptive interventions highly desirable. Interventions that don't involve closing schools often implement both behavioral mitigation strategies and physical mitigation strategies in the form of redesigning building layouts. For example, The Brooklyn Lab Charter School has partnered up with various architecture firms to formulate a public plan that schools around America can use and implement that comply with Covid Health Guidelines. Moreover, they show an ICT (Information and Communication Technology) classroom before and after Covid, as seen in Figures 6A and 6B. It is clear that they have divided the classrooms into two, allowing a similar ratio of space to students in a safe and healthy way. Furthermore, they are limiting each classroom to 15 students and everyone, including teachers and staff, has to wear masks and social distance at all times.

FIGURE 6A AND 6B: The Brooklyn Lab Charter School.

However, our findings suggest that the patterns of interaction between primary school children will complicate these physical distancing measures. The dense temporal network pattern suggests that primary school children are inclined to be highly-interactive, particularly with their classmates. Students and teachers commonly interact with multiple nodes, as opposed to being highly clustered with a single group of nodes, thus making the entire school community susceptible to disease if even one node is sick. Moreover, given the size of the groups that need to be formed to accommodate the communities in primary schools, it is almost impossible to continue schooling while maintaining the given number of groups. This makes social distancing within schools very difficult. Given that different nodes rate highest across centrality measures, we cannot simply remove a single node from schools to mitigate disease spread. Multiple nodes, in other words, too many students, are situated as central nodes in a primary school network.

These findings collectively suggest that given the levels of interaction between students and teachers as shown in our findings, strict policies surrounding interactions between students, teachers, and classes must be followed to limit the interaction students are accustomed to. Based on this understanding of the networks of students, we propose several options that educators and policy makers can implement within primary schools to limit the disease spread (Gemmetto et al. 2014):

(i) close the student's specific class for a fixed duration ("targeted class closure" strategy);
(ii) close the class and the other class of the same grade for a fixed duration ("targeted grade closure" strategy);
(iii) the entire school is closed for a fixed duration ("whole school closure" strategy).

This will forcibly reduce the contact between students and allow for disease to be mitigated. This is supported by the fact that the characteristic common among schools that are doing well is that they are all operating under capacity. Such schools have opened with arrangements designed to minimize crowding, such as grouping students to come to school on different days and allowing students to attend only from home (Boyle 2020). This would mean that students remain within their classroom communities and could be a possible solution to safely allow students to return to school. However, this too contains risks, given that unsupervised students are likely to interact across classrooms, as our data suggests.

In addition, as schools begin to open up, the newest strategy deployed in order to avoid a spike in the rate of infections (assuming that masks are being mandatorily enforced on school campuses) has been that of architectural and resource interventions. The New York Times (2021) published an interactive assessment of how air circulates in a closed classroom compared to an open window, and a HEPA (high-efficiency particulate air) filter and box fan in a window.

FIGURE 7A: Infected Air in Closed Classroom.

Figure 7A illustrates air circulation inside a classroom with closed windows and absence of a HEPA filter and box fans. Figure 7A shows the air of an infected student (highlighted in blue and outlined with a red textbox) circulating within a classroom without proper modifications. Within the schematic, the darker the line, the higher viral load shedded within the surrounding air.

FIGURE 7B: Infected Air in Open Window Classroom.

In contrast, Figure 7B depicts the difference of the infected air emitted from the infected students once a window is opened in the classroom. With such mitigation effort, there is less contaminated air circulating inside the classroom, but there is risk for infection for those seated directly adjacent to the student.

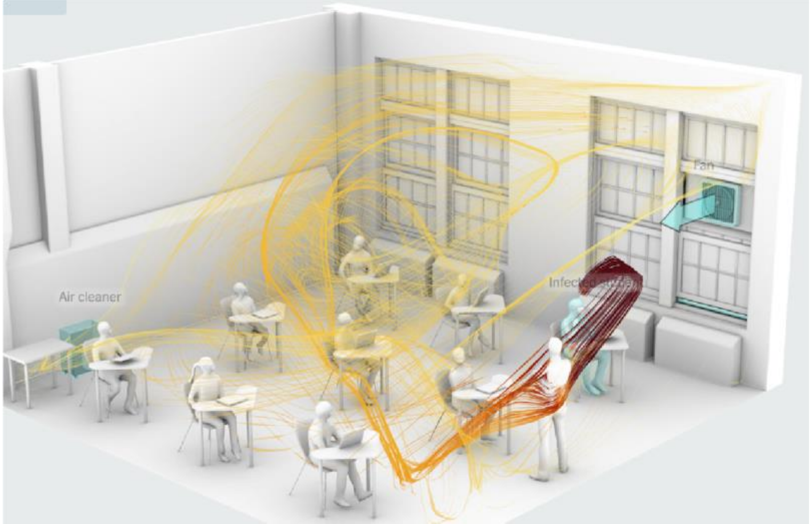


FIGURE 7C: Infected Air with Box Fan Classroom.

Figure 7C illustrates the difference of the air emitted from the infected students once a box fan blows sterilized air into the classroom. In the left corner, an air cleaner (blue) can also be seen.

Together, these two tools mitigates the circulation of air contaminated with viral load.

FIGURE 7D: Infected Air in Recirculated Classroom.

Figure 7D depicts the difference of the air emitted from the infected students once a box fan blows infected air outside of the classroom and recirculates sterilized air via an air cleaner. In this schematic, the air cleaner is in the center of the room. This depiction illustrates the safest classroom, as infected air is redirected outside and an air cleaner sterilizers the room.

Conclusively, a strategic ventilation system is needed to reduce the spread of COVID-19.

Conclusion

The high levels of interaction between primary students within the classroom and between classes makes implementing appropriate mitigation strategies extremely difficult.

Effective contact tracing, testing, and isolation are needed to control the spread of COVID within the population. Without strictly set protocol, managing and mitigating the spread of disease while allowing students to meet in person is a costly and difficult task because of the sociability of the networks of primary school students. As Benjamin Linas, associate professor of medicine and epidemiology at Boston University School of Medicine stated, “You can only open your school safely if you have COVID under control in your community”, a primary school is one large community with high interaction between students and teachers. At the present time, given our understanding of primary school networks, re-opening of schools is a high-risk decision. If classrooms must open, intentional strategies must be implemented to reduce the contact between students. This includes reducing class size, social distancing in the classroom, utilizing tools for mitigation with a particular focus on ventilation systems, and adopting a hybrid approach to reduce the number of students attending school on any given day. This must be paired with educational interventions with students and teachers about the risks of COVID-19 and how easily it can be spread in the school environment. Students and teachers must be collaborators in the effort to mitigate disease in their classrooms. Disease mitigation is a communal effort and primary schools remain amongst our most valuable and most vulnerable communities.

References

Bartzokas, N., M. Gröndahl, K. Patanjali, M. Peyton, B. Saget, & M. Syam (2021). Why Opening Windows Is a Key to Reopening Schools. *The New York Times*.

Boyle, P. (2020). Kids, school, and COVID-19: What we know — and what we don't. AAMC. https://www.aamc.org/news-insights/kids-school-and-covid-19-what-we-know-and-whatwe-don-t

Brooklyn Lab Charter School. "Back to School Instructional Program Scheduling Map."

Department of Education (2008). *Schools and Staffing.* https://nces.ed.gov/surveys/sass/tables/sass0708_035_s1s.asp

Gemmetto, V., A. Barrat, & C. Cattuto (2014). Mitigation of infectious disease at school: targeted class closure vs school closure. *BMC Infect Dis*, 14, 695. https://doi.org/10.1186/s12879-014-0695-9

Stehlé J., N.Voirin, A. Barrat, C. Cattuto, V. Colizza, L. Isella, C. Régis, J.-F.Pinton, N. Khanafer, W. Van den Broeck, & P. Vanhems (2011). Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees. *BMC Medicine*, 9 (1): 87-10.1186/1741-7015-9-87.

Note: the code used for the social network analysis can be made available on request.