

MODELING THE SPREAD OF COVID-19 OVER VARIED CONTACT  
NETWORKS

A Thesis  
presented to  
the Faculty of California Polytechnic State University,  
San Luis Obispo

In Partial Fulfillment  
of the Requirements for the Degree  
Master of Science in Computer Science

by  
Ryan Lee Solórzano

June 2021

© 2021  
Ryan Lee Solórzano  
ALL RIGHTS RESERVED

## COMMITTEE MEMBERSHIP

TITLE: Modeling the Spread of Covid-19 Over Var-  
ied Contact Networks

AUTHOR: Ryan Lee Solórzano

DATE SUBMITTED: June 2021

COMMITTEE CHAIR: Theresa Migler, Ph.D.  
Assistant Professor of Computer Science

COMMITTEE MEMBER: John Seng, Ph.D.  
Professor of Computer Science

COMMITTEE MEMBER: Paul Anderson, Ph.D.  
Associate Professor of Computer Science

## ABSTRACT

### Modeling the Spread of Covid-19 Over Varied Contact Networks

Ryan Lee Solórzano

When attempting to mitigate the spread of an epidemic without the use of a vaccine, many measures may be made to dampen the spread of the disease such as physically distancing and wearing masks. The implementation of an effective test and quarantine strategy on a population has the potential to make a large impact on the spread of the disease as well. Testing and quarantining strategies become difficult when a portion of the population are asymptomatic spreaders of the disease. Additionally, a study has shown that randomly testing a portion of a population for asymptomatic individuals makes a small impact on the spread of a disease [9].

This thesis simulates the transmission of the virus that causes COVID-19, SARS-CoV-2, in contact networks gathered from real world interactions in five different environments. In these simulations, several testing and quarantining strategies are implemented with a varying number of tests per day. These strategies include a random testing strategy and several uniform testing strategies, based on knowledge of the underlying network. By modeling the population interactions as a graph, we are able to extract properties of the graph and test based on those metrics, namely the degree of the network. This thesis found many of the strategies had a similar performance to randomly testing the population, save for testing by degree and testing the cliques of the graph, which was found to consistently outperform other strategies, especially on networks that are more dense. Additionally, we found that any testing and quarantining of a population could significantly reduce the peak number of infections in a community.

## ACKNOWLEDGMENTS

Thanks to:

- Dr. Theresa Migler: you have helped guide me throughout this past year of unknowns and provided me constant support. Even though we could not meet in person, you made me never feel alone. Thank you so much for leading me through my academic journey and I look forward to seeing where our work, and any of your future work will go.
- Dr. John Seng: I've had you for so many classes now (I believe the official count is four quarters) and helped shape my academic experience. Your teachings helped me get one of my first jobs, and I was very appreciative to have you on my committee.
- Dr. Paul Anderson: You taught this computer engineer how to do the basics of data science. Without your class, I probably would still be working on the simulation and processing the results.
- Logan, Daniel, Will, and Mike: Thanks for all the laughs, games, and encouragement. College would not have been the same without you all. Here's to many more memories.
- Joey and Kevin: Thanks for pretending my grad school memes are good, and for making this year inside the best year it can be.
- Emily: For reading and revising my first draft, but more importantly providing so much love and support. I'm looking forward to more adventures.
- To my parents, Lalo and Elsa, and my sister, Bri: You all provided such a strong support system and provided me such strong role models that guided me

to where I am today. For the support and love and for making me the person I am today, thank you.

## TABLE OF CONTENTS

	Page
LIST OF TABLES . . . . .	ix
LIST OF FIGURES . . . . .	x
CHAPTER	
1 Introduction . . . . .	1
2 Background . . . . .	4
2.1 Graphs . . . . .	4
2.1.1 Graph Metrics and Graph Algorithms . . . . .	4
2.1.2 Contact Networks . . . . .	6
2.2 Epidemiological Models . . . . .	7
2.2.1 General Terminology . . . . .	8
2.2.2 Compartmental Models . . . . .	10
2.3 RT-PCR Based Tests . . . . .	10
3 Related Works . . . . .	11
4 Network Descriptions . . . . .	13
5 System Design . . . . .	15
5.1 Bucket Model . . . . .	15
5.2 Simulation . . . . .	17
6 Intervention Strategies . . . . .	19
6.1 No Testing . . . . .	20
6.2 Random Batch Testing . . . . .	21
6.3 Individual Testing . . . . .	21
6.4 Testing using Degree . . . . .	22

6.5	Testing using Graph Properties . . . . .	22
7	Results . . . . .	24
7.1	Single Simulation Results . . . . .	24
7.2	Maximum Number Infected . . . . .	29
7.3	R-value . . . . .	35
8	Evaluation . . . . .	42
8.1	Contact Network Factors . . . . .	42
8.2	No Testing . . . . .	43
8.3	Random Batch Testing . . . . .	44
8.4	Individual Testing . . . . .	45
8.5	Testing Using Graph Properties . . . . .	45
8.5.1	Degree . . . . .	45
8.5.2	Clique . . . . .	46
8.5.3	K-Core . . . . .	46
8.5.4	Betweenness Centrality . . . . .	47
8.5.5	PageRank . . . . .	47
8.6	Evaluation by R-Value . . . . .	48
9	Conclusion . . . . .	49
10	Future Work . . . . .	51
	BIBLIOGRAPHY . . . . .	53



## LIST OF TABLES

Table		Page
4.1	A description of each contact network used for this research [4]. . .	13
4.2	Properties of each contact network. Values are averaged over all the days of the study [4]. . . . .	14
6.1	The False Negative rates of RT-PCR based testing, where Day is days since exposure, and FNR is the False Negative Rate, or percent chance someone will get a false negative result . . . . .	20
7.1	A summary of the score for each strategy across each contact network when considering the maximum number infected . . . . .	35
7.2	A summary of the score for each strategy across each contact network when considering the R value . . . . .	36

## LIST OF FIGURES

Figure	Page
2.1 An example contact network graph . . . . .	7
2.2 Histogram showcasing the power law nature of the number of interactions per person in the Conference contact network . . . . .	8
5.1 Proposed SEQIR Model . . . . .	15
7.1 Results of a single simulation without any intervention . . . . .	25
7.2 Results of a single simulation with testing 90 students per day with the random batch strategy . . . . .	25
7.3 Results of a single simulation with testing 90 students per day with the individual testing strategy . . . . .	26
7.4 Results of a single simulation with testing 90 students per day prioritizing those with a high degree . . . . .	26
7.5 Results of a single simulation with testing 90 students per day prioritizing those with high clique numbers . . . . .	27
7.6 Results of a single simulation with testing 90 students per day prioritizing those with high k-core numbers . . . . .	27
7.7 Results of a single simulation with testing 90 students per day with a high betweenness centrality . . . . .	28
7.8 Results of a single simulation with testing 90 students per day with a high PageRank . . . . .	28
7.9 Averaged results of the maximum number infected for each strategy on the Workplace 1 contact network . . . . .	30
7.10 Averaged results of the maximum number infected for each strategy on the Workplace 2 contact network . . . . .	31
7.11 Averaged results of the maximum number infected for each strategy on the High School contact network . . . . .	32

7.12	Averaged results of the maximum number infected for each strategy on the Conference contact network . . . . .	33
7.13	Averaged results of the maximum number infected for each strategy on the Pre-School contact network . . . . .	34
7.14	Averaged results of the R-value for each strategy on the Workplace 1 contact network . . . . .	37
7.15	Averaged results of the R-value for each strategy on the Workplace 2 contact network . . . . .	38
7.16	Averaged results of the R-value for each strategy on the High School contact network . . . . .	39
7.17	Averaged results of the R-value for each strategy on the Conference contact network . . . . .	40
7.18	Averaged results of the R-value for each strategy on the Pre-School contact network . . . . .	41

## Chapter 1

### INTRODUCTION

Mankind was not prepared for the SARS-Cov-2 pandemic, and unless we learn from it, we will not be ready for the next pandemic. There are many behavior changes we can implement as a society which have been proven to slow the spread of a disease, such as wearing masks and social distancing. Yet, while these behavioral changes lower how likely it is to transmit the disease, they do not involve the isolation of individuals we think are sick. Furthermore, the testing/ quarantining plan in the United States as a whole is mostly voluntary and relies on anecdotal information of being able to list the people one was in contact with before being confirmed to have COVID-19. Duke University has showcased the effectiveness of a well implemented testing plan among a community, with their population showing much smaller numbers relative to their surrounding community [3]. The efficacy of their response was credited in large part due to their “aggressive testing” strategy, when combined with their pushes for social distancing, mask mandates, hand washing, and more.

So we may, hypothetically, take advantage of the robust network of smartphones and record the interactions between two individuals via their Bluetooth interactions. This thesis aims to leverage five such contact networks and make a testing intervention strategy using random testing, uniform testing, and graph theory metrics from the contact networks of individuals. The research done in this paper aims to answer several questions:

- RQ1: What percentage of the population would need to be randomly tested every day to minimize the spread of a disease?

- RQ2: Is testing random batches of a population more or less effective than every individual randomly testing twice a week (Similar to Cal Poly’s testing strategy)?
- RQ3: Can we decrease the number of tests required to effectively minimize the spread of a disease by using a testing strategy that uses graph theory?

For this thesis, we make several educated hypothesis for these research questions. However, the hypotheses one may create depends largely on what is defined as “effectively minimizing the spread of a disease.” In the case of the following hypotheses, we will consider “effectively minimizing” as minimizing the peak number of infectious individuals. This is influenced by the idea of “flattening the curve” in order to minimize the number of hospitalizations of a population. The question of evaluation is explored further in Chapter 8. The hypotheses are as follows:

- HP1: In order to minimize the spread of an infectious disease, at least half of a population will have to be tested each day.
- HP2: With a low number of available tests, having each individual regularly test will likely be more effective. However, as the quantity of tests becomes large, the random batches will likely be more effective (albeit a bit impractical).
- HP3: Assuming equal amounts of tests, a testing strategy with graph theory will perform better given a small quantity of tests. However, random testing will eventually dominate as the quantity of tests increases.

There have been several methods developed for testing if an individual has COVID-19. This thesis focuses on the Reverse Transcriptase Polymerase Chain Reaction (RT-PCR) based testing due to availability of research [10]. With these tests, you

have the potential to correctly or incorrectly test positive, or correctly or incorrectly test negative for COVID-19. When someone is falsely identified to have COVID-19, it is called a false positive, and when someone falsely tests negative it is called a false negative. Tests have what is called a false positive rate and false negative rate. In the simulation used by this thesis, we either declare a person 'positive' or 'negative' based on the false positive and negative rates, and appropriately quarantine. The values for these rates is discussed further in Chapter 5.

## Chapter 2

### BACKGROUND

#### 2.1 Graphs

Graph theory is a broad topic which has applications from topology to computer networks, yet has humble beginnings. In 1735, Swiss mathematician Leonhard Euler wondered if it was possible to cross every bridge of Königsberg, without crossing every bridge twice [1]. Euler proved that it was not possible, and furthermore is only possible if the city had at most 2 landmasses with an odd number of bridges attached to them. By solving this problem, Euler proved the first theorem of modern graph theory. The Königsberg bridge problem can be abstracted to a graph, where the landmasses are referred to as nodes and the bridges are edges. For this thesis, we will focus on *simple graphs* where an edge goes in both directions (an undirected graph), two nodes can only be connected by at most one edge, and a node cannot be connected to itself.

##### 2.1.1 Graph Metrics and Graph Algorithms

There are several graph metrics and algorithms which this thesis will utilize, which include the *degree* of vertices, *clique* subgraphs, *k-Cores*, *betweenness centrality*, and *page rank*. This section will define each metric or algorithm and the motivation behind using it in a contact network.

The *degree* of a node is defined as the number of edges connected to it. In the context of a contact network, a high degree corresponds to an individual who is in contact

with a lot of people. Therefore, the likelihood that they would contract a disease is higher than an individual with a low degree.

A graph is considered a *clique* if every node is connected to the other. So, any typical graph has subgraphs that are a cliques, but the size of the largest clique depends on how well connected the graph is. Within epidemiological networks, viruses are known to spread rapidly within cliques, so if an individual is in a large clique they may have a high risk of contracting a disease.

Similar to clique, a *k-core* subgraph is defined as every node having degree  $k$  or more. The motivation behind this is similar to individuals with a high degree and clique; if someone is within a high  $k$ -core they not only have a high likelihood to contract or spread a disease, but also they are within a well connected core that may speed up the spread.

*Betweenness centrality* is a measure of the “centrality” of a node, meaning how many of the shortest paths go through it. For example, if buildings are nodes and roads are edges, the buildings downtown would have a high betweenness centrality, since there are a lot of paths which go through them to get to another part of town. The motivation behind using this metric is to catch those individuals who are “in between” certain social groups so that if they get infected, the disease has a hard time spreading to another part of the graph.

*Page rank* is a very popular algorithm developed primarily by Larry Page from Google to rank the likelihood of an individual of someone visiting a webpage. It does this by starting each webpage with the same number of “people” and evenly send each “person” to their neighbors until an equilibrium is reached. In the case of this algorithm, an equilibrium is reached when the values do not change between iterations. In the context of epidemiological networks, rather than of visualizing the algorithm as

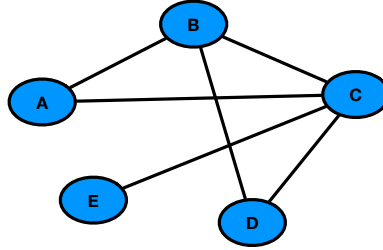


people visiting webpages we will be interpreting it as the diseases traveling to people. Therefore, the numbers can be interpreted as the likelihood of a disease reaching a person. Admittedly, this will not be as important in the early stages of a disease, but will hopefully become relevant as a disease becomes more well mixed. Research has suggested that nodes with high PageRank values are more likely to be super-spreaders of a disease compared to those with lower PageRank values [15], although this was in the context of cattle herds spreading a disease.

### 2.1.2 Contact Networks

Graphs can be modeled on much more than just land and bridges. For this thesis, nodes will represent people and edges represent an interaction, or contact, between two people that allows for transmission of COVID-19. Thus, we may define a contact network graph as  $G = (V, E)$ , where  $G$  = the contact network,  $V$  = the vertices, or the people in the contact network, and  $E$  = the edges, or the contacts between two people. The definition of a contact varies by who collected the contact network data, however it generally is defined as two people coming a certain distance from each other for greater than a set amount of time. Figure 2.1 shows an example contact network. In this example,  $V = \{A, B, C, D, E\}$  and  $E = \{(A, B), (B, C), (A, C), (B, D), (C, D), (E, C)\}$ . Since we are interpreting this as a contact network, we can imagine a scenario where  $A$ ,  $B$ , and  $C$  are roommates,  $B$ ,  $C$ , and  $D$  are in a club sport together, and  $E$  and  $C$  have a class together.

We must also introduce the idea of time, since different time slices of a day are different graphs, since two people may only interact at certain times of a day. Thus, graphs will now be denoted as  $G_\tau = (V_\tau, E_\tau)$ , where  $\tau$  is an hour time-slice of the contact network at hour  $\tau$ . Additionally, if a larger time-slice is needed, the graph will simply be denoted as  $G_{\tau_1-\tau_2}$ , where  $G_{\tau_1-\tau_2} = (V_{\tau_1} \cup V_{\tau_2}, E_{\tau_1} \cup E_{\tau_2})$ . In the scope



**Figure 2.1: An example contact network graph**

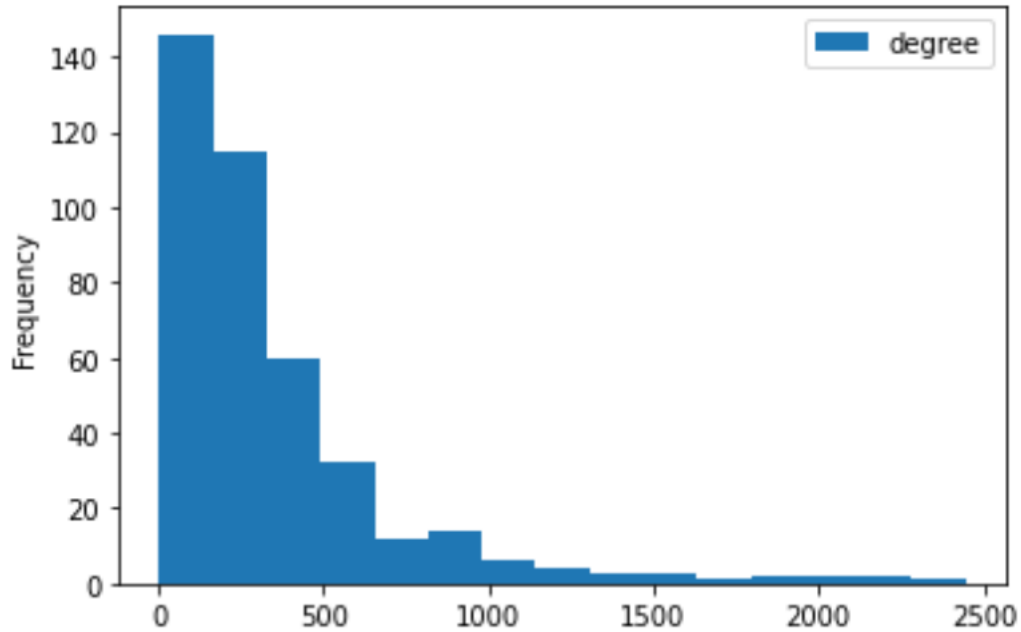
of this experiment, the largest time slice used was a day before to construct the graph for the testing strategies to use.

Going back to our example in Figure 2.1, we can construct a contact network such that  $G_{night}$  just has  $A$ ,  $B$ , and  $C$  connected, since they are roommates who regularly interact at night.

The contact networks used in this research are gathered from SocioPatterns, a multinational research group that gather contact data in several scenarios [4]. The scenarios used in this research include a high school, two office workplaces, a science conference, and an elementary school. The efficacy of the intervention strategies will be evaluated across each of these networks. The details of the SocioPatterns networks are discussed further in Chapter 4.

## 2.2 Epidemiological Models

The earliest account of modeling a disease mathematically was in 1760 by Daniel Bernoulli, where he formulated and solved a model of the smallpox disease [6]. In this model, he grouped the population into susceptible and immune, where those who already got sick are in immune and everyone else is susceptible. He did this to showcase the importance of inoculating the public, a primitive form of vaccination. From here, mathematical models have only improved, yet often assume a “full mixed”



**Figure 2.2: Histogram showcasing the power law nature of the number of interactions per person in the Conference contact network**

model where every person has equal chance of being in contact with another. While this is somewhat of an accurate model, it is very limited as people generally do not have an equal probability of interacting with every other person in a community. In fact, the degrees of real-world contact networks often follow a power law, meaning more people have a relatively small degree of contacts. This property is showcased in our own networks and displayed in Figure 2.2. This is why contact networks are so important in modeling the spread of a disease; they provide a more powerful framework by which to model the outcome of a disease [12].

### 2.2.1 General Terminology

There are several terms which are used in epidemiology which need to be defined in the scope of this research, which include  $R_0$ , *epidemic*, *endemic*, *herd immunity*, *false negatives and positives*, and *test sensitivity and specificity* [13, 14].

$R_0$  is sometimes referred to as the basic reproduction number, or how many people we can expect an infectious individual to infect. When  $R_0 = 1$ , each person only infects one person while they are sick and a disease is referred to as an endemic. If  $R_0 > 1$ , the disease is now an epidemic, and the spread of the disease becomes very rapid. There is no term for when  $R_0 < 1$ , but if this is the basic reproduction number of a disease then the number of people infected will be rapidly decreasing. The  $R_0$  value differs across diseases, and refers to the spread of a disease *initially* [7]. As a disease spreads, this number changes and may be simply referred to as the  $R$  value of a disease.

*Herd immunity* is the percentage of the population that must be immune to the disease in order for it to not persist. In terms of  $R_0$ , if a population is at the herd immunity threshold,  $R_0 < 1$  [13].

A *false negative* result means that an individual with a disease incorrectly receives a negative test result for a disease. Conversely, a *false positive* result means that an individual without a disease receives a positive result for the disease.

In the study of medical diagnosis, there are two statistical features of a medical test that are important to look at: the *sensitivity* and the *specificity* of the test. *Sensitivity* is defined as the proportion of people with the disease who correctly test positive using the test, and *specificity* is defined as the percentage of people without the disease who have a negative test result. Some more commonly known statistics for testing is the *false negative rate* and the *false positive rate*. The false negative rate equals  $1 - \text{sensitivity}$ , and the false positive rate equals  $1 - \text{specificity}$ . The false negative rate and false positive rate give us values that tell us exactly how often the tests are wrong for any given individual (sick or healthy), so these values are what are actually used in the simulation.

### 2.2.2 Compartmental Models

Models now have built on Bernoulli’s “susceptible” and “immune” model to a Susceptible, Infectious, and Recovered model, or SIR model. These types of models are generally referred to compartmental models where the entire population is partitioned into one of the compartments listed. Variations of these models include SEIR model, which includes individuals exposed to the disease, and SEQIR, where a percentage of the population is quarantined because they are showing symptoms [5]. Much like Bernoulli, there can be a mathematical formula from one compartment to another. However, this thesis focuses on the movement of individuals between compartments based on the contact networks that have been gathered. The compartmental model used in this paper can be seen in the experimental design section.

### 2.3 RT-PCR Based Tests

When determining if an individual has contracted a disease, a medical test must be conducted. A widely available test used for detecting SARS-CoV-2 is the reverse transcriptase polymerase chain reaction test, or RT-PCR [10]. This test is often administered as a swab in the nasal region, although other methods of extracting samples may include saliva, throat samples, or other bodily fluids [16]. For medical tests such as this one, it is important to consider the *sensitivity* and *specificity* as discussed in subsection 2.2.1. What values are used is discussed further in Chapter 6, however it is important to note that this test has a relatively high sensitivity, and a high specificity, meaning those with the disease are likely to test positive, and those without the disease are not.

## Chapter 3

### RELATED WORKS

A work by St-Onge et al. first tried to model a more realistic spread of a hypothetical disease using an improved SIR model [18]. Given the timing of the experiment, it may be assumed that this model was loosely based on COVID-19. The authors then showcased the importance of intervention strategies in mitigating the spread of disease in networks with higher-order structure (i.e. a large community). The evaluation of the intervention strategies is of particular interest in this research since this paper also explores intervention strategies to mitigate the spread of a disease.

Imai et al. was one of the first studies to show that the transmission of SARS-CoV-2 was self-sustaining, or that its  $R$  value is greater than 1 [7]. Additionally, they calculated the  $R_0$  value of COVID-19 to be 2.6. This study was conducted in Wuhan, China at the beginning of the COVID-19 pandemic, January 2020 and observed all the estimated amount of cases prior to the publication of the paper. This paper is of particular interest because it gives us the realistic values of the spread of COVID-19 in the absence of disease intervention strategies such as mask wearing which to compare the results of the simulations of this research to.

Kucharski et al. explored the effect of several intervention strategies on the effective  $R_0$  value of COVID-19 [9]. This work found that a combined testing and tracing strategy was the most effective at lowering the  $R$  value of a disease. Similar to Imai et al., the evaluation of these intervention strategies is important with this research as this study explores similar intervention strategies to this thesis.

Siu et al. attempt to mitigate the spread of a disease by creating vaccine intervention strategies using graph theory [17]. Additionally, rather than simulating the epidemic using a mathematical model, they also use a contact network gathered from Copenhagen [19]. This work also implements vaccination strategies using underlying reasoning which is similar to a testing/ quarantining strategy.

G'enois et al. explore if using co-location information can be down sampled to accurately model a real life contact network [4]. Here, co-location data is defined as two individuals being in the same general area such as a room. This paper used the SocioPatterns datasets and was able to show that there was no down-sampling technique which is able to accurately model real world interactions across any scenarios. This is helpful in the realm of this thesis, because there are many more contact networks that look more like a co-location interaction rather than a face to face interaction, so this paper shows that using these networks is not as accurate.

Estrada et al. have a very in-depth paper on how to mathematically model SARS-CoV-2 using a modified SIR model [2]. While I'm more focused on modeling using contact networks rather than mathematical models, this paper is still extremely helpful in creating my modified compartmental model, and for parameters for my simulation such as probability of infection.

Shah et al. explore the correlation between the super spreaders of a disease and the PageRank of those spreaders [15]. One of the results of this study found that the nodes with a high PageRank value contain a higher proportion of super spreaders than the nodes with lower PageRank values. It should be noted that this study was done in the context of the spread of disease among herds of livestock, however the results should still apply in a human contact network.

## Chapter 4

### NETWORK DESCRIPTIONS

The experiments done in this research use contact networks gathered by SocioPatterns. SocioPatterns is a collaboration between researchers and developers across the world to collect various contact network data for networks ranging from Baboons’ interactions to interactions among people within a hospital [4]. All of the data is collected using the same system where each participant wears an RFID tag and reader which records if two individuals come into contact. In the case of this network, a contact is defined as the readers of both individuals register the RFID tag of the other for a 20s time window. This collaboration states that two individuals must be within a 1.5 meters of each other to record a contact, or about 5 feet. However, intensity is not specified meaning we must therefore assume that all contacts are equal and is sufficient duration and distance to transmit COVID-19 from an infectious individual to a susceptible one. Table 4.1 summarizes each contact network used in this research and Table 4.2 summaries the properties of each network.

**Table 4.1: A description of each contact network used for this research [4].**

Network	Year	Participants	Duration
Workplace 1	2013	92	2 weeks
Workplace 2	2015	232	2 weeks
High School	2013	326	1 week
Conference	2009	403	2 days
Pre-School	2009	242	2 days

Looking at Table 4.2, we can see that these networks have very different structures. Most notably, the Workplace networks have a much smaller average degree and network density than the rest of the networks, and the Pre-School network has a significantly higher average degree and density than the rest. Intuitively, this makes



**Table 4.2: Properties of each contact network. Values are averaged over all the days of the study [4].**

Network	Average Degree	Network Density	Clique Number
Workplace 1	2.9	0.030	4.4
Workplace 2	6.4	0.028	7.6
High School	13.5	0.041	9.4
Conference	28.8	0.072	11.0
Pre-School	47.3	0.196	22.5

sense because an office workplace will likely have a hierarchy where in a typical day, each individual only comes into contact with the people in their group; i.e. people who work in Human Resources will typically only interact with people in the same department. In a school on the other hand, the nature of the students' schedules causes them to switch around classrooms frequently and have a very high mixture of contacts between their peers and teachers. Additionally, by design, in the Conference dataset we can expect a very well mixed network, which results in its higher average degree and density. The clique numbers are also interesting to note, as the schools and conference networks have a higher clique number than the office workplaces.

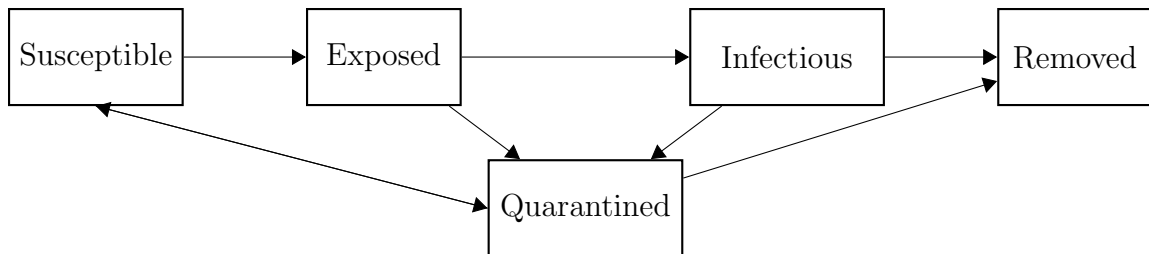
These aspects of the contact networks will be interesting in evaluation as, by design, some individuals will be frequently tested whereas others will be able to evade testing. Due to the wide variation in average degrees and network density, we can expect those with smaller average degrees to produce a smaller  $R_0$  value, and have a much lower peak infection numbers than the more well mixed networks. Additionally, any testing strategy that calculates clique number may be expected to perform better in the Pre-school network as its clique numbers are much higher.

## SYSTEM DESIGN

**5.1 Bucket Model**

When creating the simulation for a COVID-19 epidemic, the first consideration was the bucket model used. The bucket model for this research is shown in Figure 5.1, and will be discussed in this section. The model starts with the entire network in the SUSCEPTIBLE bucket, except one randomly chosen person is moved to the EXPOSED bucket. Which person is chosen can have a large impact on the initial stage of the infection simulation. For example, if the person has a small number of contacts, there are less chances to infect others which means that there may not even be an outbreak. As a result, each simulation uses a different seed value to ensure an adequate range of scenarios for this research. Then, the simulation runs by moving the individuals along the arrows defined by their interactions, testing strategies, and time.

From the SUSCEPTIBLE bucket, the nodes have two ways to move out: either they come into contact with someone in the INFECTIOUS bucket and are moved to the



**Figure 5.1: Proposed SEQIR Model**

EXPOSED bucket, or they receive a false positive test and are quarantined. An individual has probability  $0 < \alpha < 1$  of being moved to the EXPOSED bucket upon contact with an infectious person, which is based on research of SARS-CoV-2. This research used an  $\alpha$  value of 0.03, which was taken from the range of infectious rates used by Kucharski et al. [9]. It should be noted that Kucharski differentiated between household contacts and close secondary contact infectious rates, and this research used the middle value of the secondary contact infectious rate. This is in hopes to mimic real life contacts, where the shorter duration contacts have a low chance to transmit the disease, whereas the repeated contacts have a high likelihood to transmit. This infectious rate resulted in an average  $R_0$  value of 2.69 in the High School network without any interventions, which is similar to that found by Imai et al [7].

Then, from the EXPOSED bucket, individuals can either become infectious after a set amount of time and move to the INFECTIOUS bucket, or they can receive a positive test and be moved to the QUARANTINED bucket (although based on the testing statistics, an individual is very unlikely to test positive while in the EXPOSED bucket). Then, from the INFECTIOUS bucket, a person can either be moved to REMOVED, or they can test positive and be moved to the QUARANTINED bucket. The name from the REMOVED bucket comes from the idea that the individuals are effectively removed from the simulation; they either recover and will not become infectious again, or they are killed from the infection. From the QUARANTINED bucket, a person is moved out after two weeks they are put into it, and that person can either move to the SUSCEPTIBLE bucket if they falsely tested positive, or to the REMOVED bucket if the infection has run its course. Finally, for the purposes of this experiment, once a person is in the REMOVED bucket, they can never be moved out. It should be noted that there has been reports of multiple COVID-19 infections in a single patient [8], however this simulation assumes one infection is sufficient due to the rare nature of a second reinfection in the time span of the simulations.

## 5.2 Simulation

Below outlines the algorithm used for this simulation. The movement for each bucket is based on the bucket model described in section 5.1.

---

**Algorithm 1:** SIR SIMULATION ( $G_t = (V_t, E_t)$ )

---

**Input** : Temporal graph  $G_t = (V_t, E_t)$ , where  $G_t$  is the graph of all the contacts at time  $t$ . Similarly,  $V_t$  and  $E_t$  are the vertices and edges of contacts at time  $t$ . (Note,  $G = \bigcap_{t=0}^{\infty} G_t = (\bigcap_{t=0}^{\infty} V_t, \bigcap_{t=0}^{\infty} E_t) = (V, E)$ )

**Output:** A CSV file of the quantities of the buckets throughout the simulation.

- 1 Let  $n$  be a randomly chosen node in  $V$ ;
- 2 Let  $S = V \setminus \{n\}$ ,  $Exp = n$ ,  $I = Q = R = \emptyset$ ;
- 3  $t = 0$ ;
- 4 **while**  $Exp \neq \emptyset$  and  $I \neq \emptyset$  **do**
  - 5 **if** *it is a new day* **then**
    - 6 | Test the population and update  $Q$  bucket;
  - 7 **end**
  - 8 Update  $R$  bucket based on time passed for those in  $Q$  bucket and  $I$  bucket;
  - 9 Update  $I$  bucket based on time passed for those in  $Exp$  bucket;
  - 10 Update  $Exp$  bucket based on interactions with  $S$  and  $I$  bucket from  $E_t$ ;
  - 11 Update  $S$  bucket with quarantine false positives;
  - 12 Store  $SEQIR$  to database;
  - 13 |  $t = t + 1$
- 14 **end**
- 15 OUTPUT database to CSV;

---

Readers may notice the order of the bucket movements; assuming no tests, the removed bucket is updated first, then the infectious buckets, and so on. This is because in the case that a person has run the course of their infection, we want to make sure that they are moved to the proper removed bucket so that they do not infect any additional individuals. A similar line of reasoning goes for the rest of the buckets.

## Chapter 6

### INTERVENTION STRATEGIES

This research explores seven different testing strategies with a varying amount of available tests. This chapter outlines each strategy, the motivation behind creating this strategy, and potential benefits and pitfalls. In the simulation, the testing strategy is implemented once a day in the morning. Additionally, only the eligible population is tested, meaning only the population in the SUSCEPTIBLE, EXPOSED, and INFECTIOUS bucket.

An important part of these strategies is the method which the tests are conducted. This research attempts to simulate the testing that we have seen for COVID-19 as closely as possible. Because of this, we use the statistical information discussed in subsection 2.2.1 from the Reverse Transcriptase Polymerase Chain Reaction (RT-PCR) based test, which is both widely used and has research conducted regarding the false negative rate [10]. Unfortunately, getting exact values for the false positive and false negative rate is extremely difficult in practice, and these values vary based on the day of infection. So, this thesis attempts to approximate the false negative rates found by Kucirka et al. by starting at 100% at day 1 of infection, then each day having each value change according to the results found by the researchers. It should be noted that they found a considerable uncertainty in their numbers, however for the sake of this experiment we will assume the numbers found are correct. Then within the simulation, testing works by checking how long it has been since exposure, and based on the false negative or positive rate for that day, have the person be marked as “positive” or “negative” using random number generation. It should be noted that Kucirka et al. did not try to determine the false positive rate of the tests, however Dr.

Shmerling of Harvard Medical Health says that we may expect this value to be at or near zero, as any false positives are likely due to lab equipment error [16]. So, for this simulation the false positive rate will be a constant 2%. Finally, for these simulations we assume immediate feedback from the test results, which more closely resembles antigen tests rather than RT-PCR [16]. However, this assumption is made because the simulations were designed to showcase the effectiveness of the testing strategies rather than the tests themselves. Table 6.1 summarizes the false negative rates used in this simulation.

**Table 6.1: The False Negative rates of RT-PCR based testing, where Day is days since exposure, and FNR is the False Negative Rate, or percent chance someone will get a false negative result**

Day	1	2	3	4	5	6	7	8	9	10
FNR	100	100	97.7	71.0	38.7	24.8	20.1	19.1	20	22
Day	11	12	13	14	15	16	17	18	19	20
FNR	25	28.6	32.5	36.8	41.2	45.5	49.6	53.5	57.0	60.2

Another important note is that these strategies do not differentiate between asymptomatic and symptomatic infectious individuals. So, any infectious individual has the same probability of testing positive given that the day they were infected was the same. Additionally, if a person is symptomatic, they will still not be tested unless the following strategies dictate that it is necessary.

## 6.1 No Testing

In order to create a control which to compare the other testing strategies, simulations without any intervention were run. These simulations omitted any testing, and therefore no one was ever moved to the quarantined bucket. These results are expected to be unrealistic as no behavior changes are implemented to avoid getting sick, and will see infection until the simulation reaches the herd immunity threshold. Yet, we

will consider this the baseline to determine how much impact testing a population can have, assuming everything else equal.

## **6.2 Random Batch Testing**

This strategy is the simplest which one can come up with which tests a random subset of the population up to the number of tests available. While relatively rudimentary, this actually somewhat simulates the initial testing in the United States. Anecdotally, there were only a limited number of tests available in the early stages of the COVID-19 pandemic, and there was not a strong system in place for asymptomatic individuals to be tested. Therefore, we could view this testing strategy as an approximation of the early development of testing in the US. It can also give us a baseline to compare other intervention strategies to; if a strategy does worse than this one, then we can safely say that it is not an effective strategy. The pitfalls of this strategy are that it is random and therefore inherently unreliable. Because the testing is random, we are not guaranteed equal coverage of testing of the population. Depending on the number of tests available, this could make it probable that an individual who is infectious will never be quarantined at the early stages of the outbreak.

## **6.3 Individual Testing**

This strategy is motivated by the current testing strategy in place at California Polytechnic State University San Luis Obispo (Cal Poly SLO). This strategy applies to all students who want to access campus resources, which includes first-year students in the dorms. Cal Poly dictates that in order for a student to be compliant, they must be tested twice a week, or about every three days. For the sake of this simulation with a varying amount of tests, we will say that students who have not been tested



the longest will be given the tests first. To match the conditions at Cal Poly, there would need to be about  $N * 2/7$  tests per day, where  $N$  is the number of students. This testing strategy should be stronger than the last in that it guarantees that the entire population is tested, however lacks in that an individual could infect a lot of other individuals in that span of three days between tests.

#### **6.4 Testing using Degree**

This is the first strategy that utilizes the properties of the graph for the contact network. A timeslice of the previous day is recorded, then the people are tested in order of highest degree. The motivation behind this comes from the idea that individuals with the highest degree have a higher likelihood of being exposed to someone. While a person who comes into contact with an infectious individual will not test positive the next day, generally people stick to their habits and those with a high degree will retain a high degree. While this strategy gives a strong likelihood that those who tend to have a low number of interactions per day are never tested, it also guarantees that those with a high number of interactions like teachers are tested first and are immediately quarantined if necessary. This effectively lowers the density of the contact network which in theory lowers the number of individuals infected.

#### **6.5 Testing using Graph Properties**

Similar to section 6.4, we will implement four more strategies which are based in the properties of a graph. Again, the graph is constructed from a timeslice of the previous day, then people will be tested based on their clique number, their k-Core value, their betweenness centrality, and their PageRank. For sorting by clique, each person is assigned the number of the largest clique they are in, then those with

the largest clique numbers are tested first. The same principle applies for k-Core, betweenness centrality, and PageRank. These strategies will have interesting analysis to be done, as each will likely perform differently across each contact network due to their prioritization of their respective populations. For example, clique and k-core will likely do better in the dense contact networks, whereas the centrality measurements may have a constant performance across each contact network.

## Chapter 7

### RESULTS

Here, we consolidate the results from each simulation. For each contact network, all seven testing strategies were simulated with the number of tests conducted per day varying from zero to the size of the population. Furthermore, each test size experiment was run ten times, and the results were recorded and consolidated in this chapter. The results of the experiments are given in several formats: the output of an experiment, the maximum number infected, and the average R-value. The explanation for each of these is given in their respective sections, and the reason for using these metrics is discussed in Chapter 8.

#### 7.1 Single Simulation Results

Figure 7.1 through Figure 7.8 show the results of a single simulation for each testing strategy. This is relatively straightforward, as we have time on the X axis, and the number of people currently infected on the Y axis. Note, number of people currently infected only refers to number of people infectious, and does not refer to those who have been exposed, but are not yet transmitting the disease. Additionally, each of the results shown in this section were from the High School network, and the testing strategies used 90 tests per day, about a quarter of the population.

These figures show an interesting progression of the disease given the different testing strategies (or lack thereof). However, it should be noted that these graphs were taken from a single experiment on a single contact network, and are not necessarily indicative of how well the test performed as a whole. However, the experiments chosen

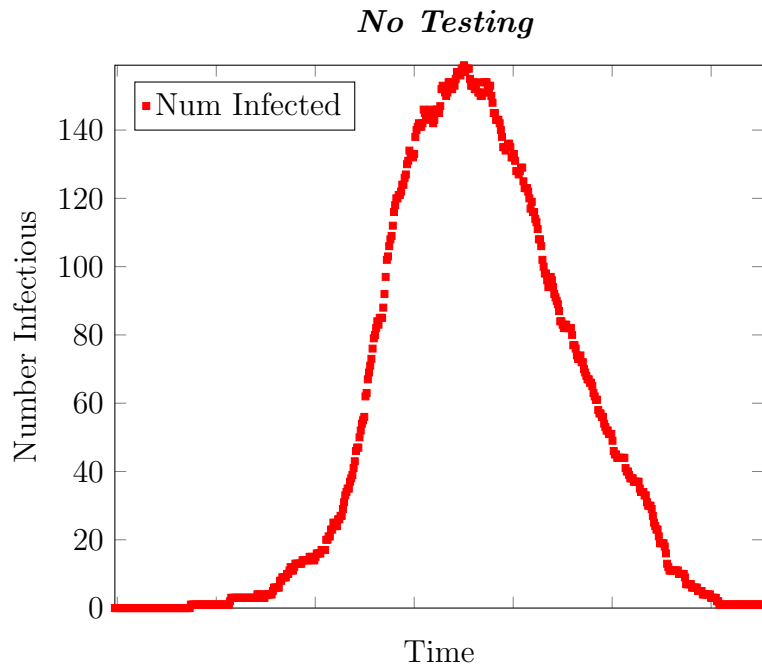


Figure 7.1: Results of a single simulation without any intervention

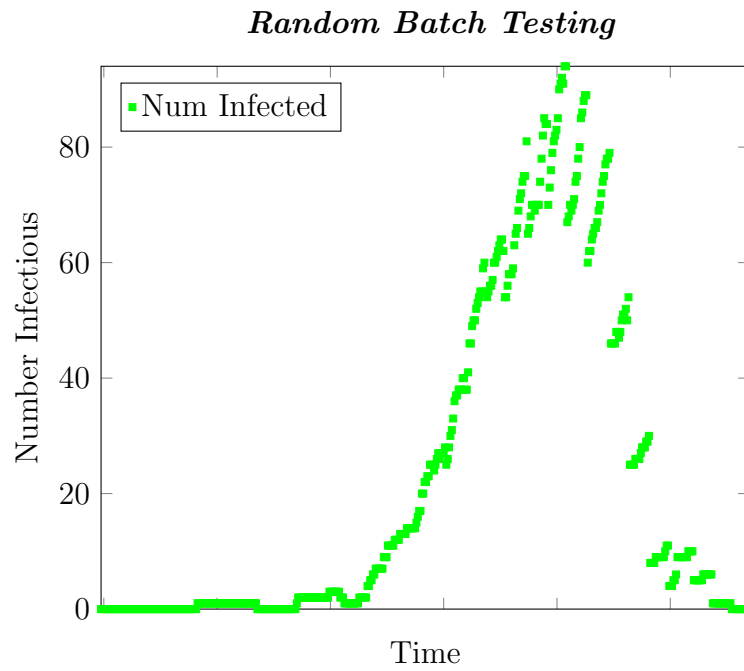


Figure 7.2: Results of a single simulation with testing 90 students per day with the random batch strategy

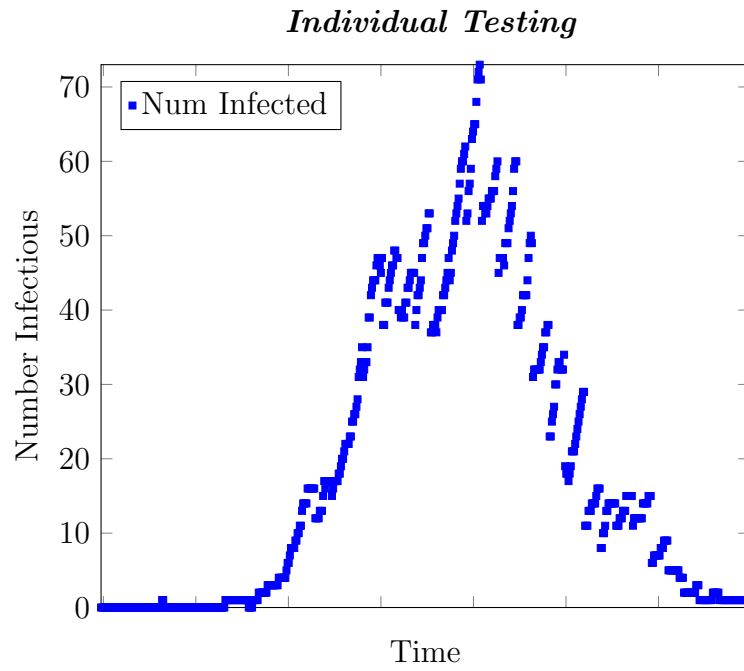


Figure 7.3: Results of a single simulation with testing 90 students per day with the individual testing strategy

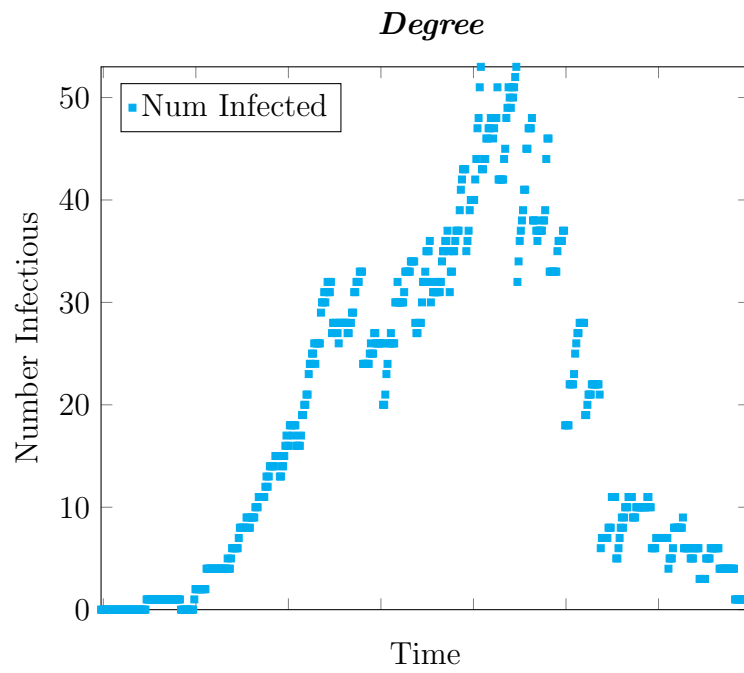


Figure 7.4: Results of a single simulation with testing 90 students per day prioritizing those with a high degree

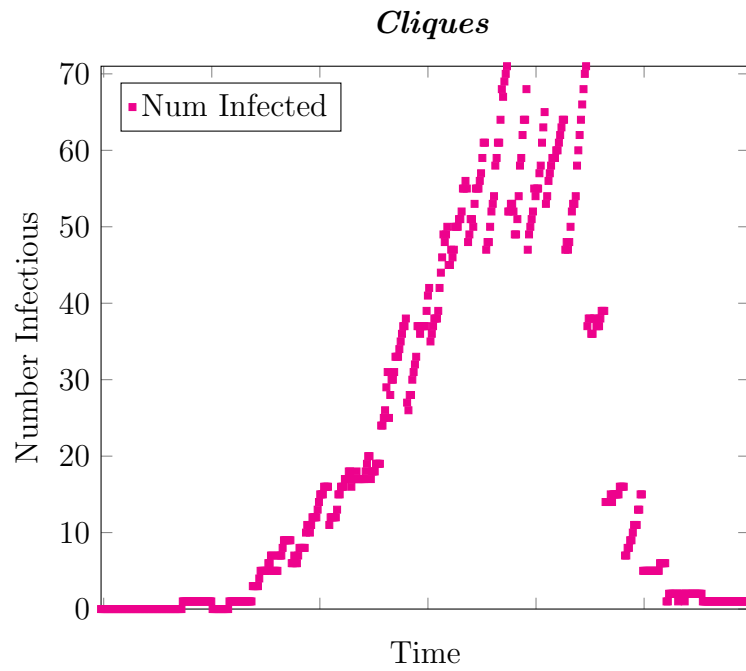


Figure 7.5: Results of a single simulation with testing 90 students per day prioritizing those with high clique numbers

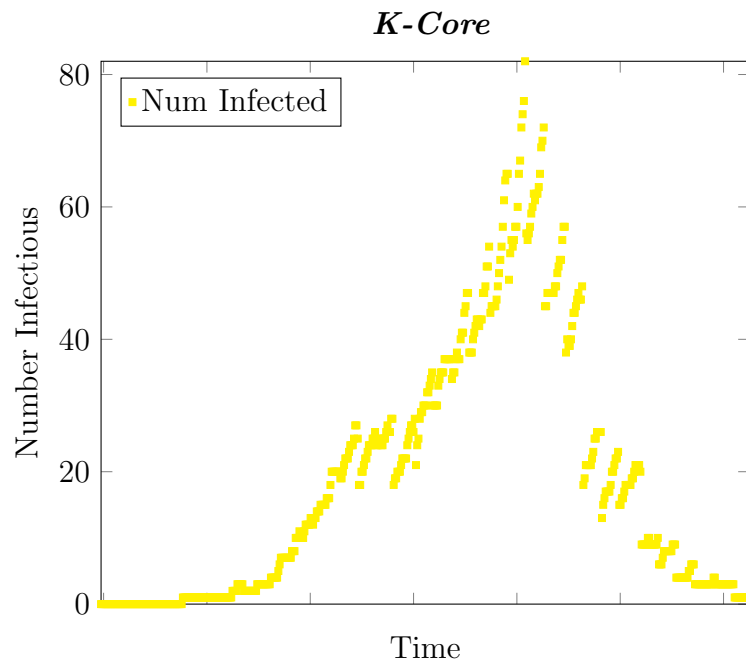


Figure 7.6: Results of a single simulation with testing 90 students per day prioritizing those with high k-core numbers

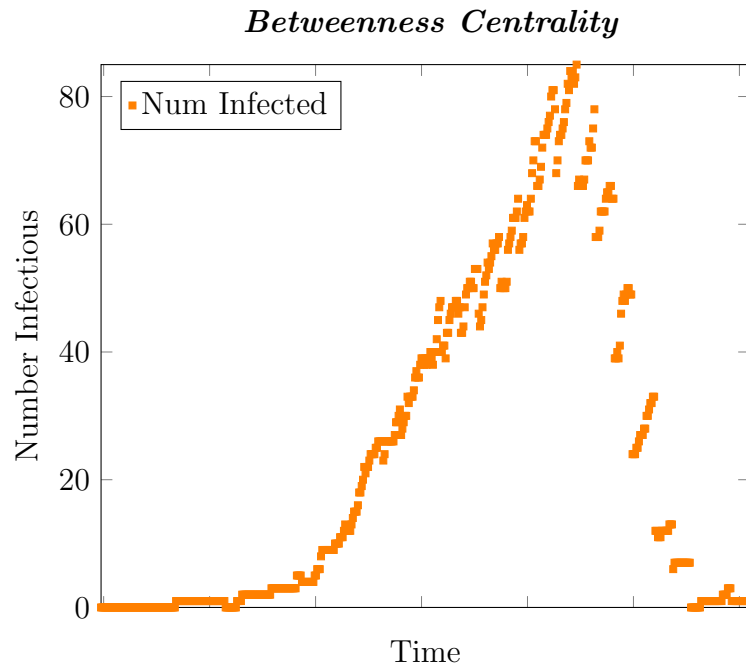


Figure 7.7: Results of a single simulation with testing 90 students per day with a high betweenness centrality

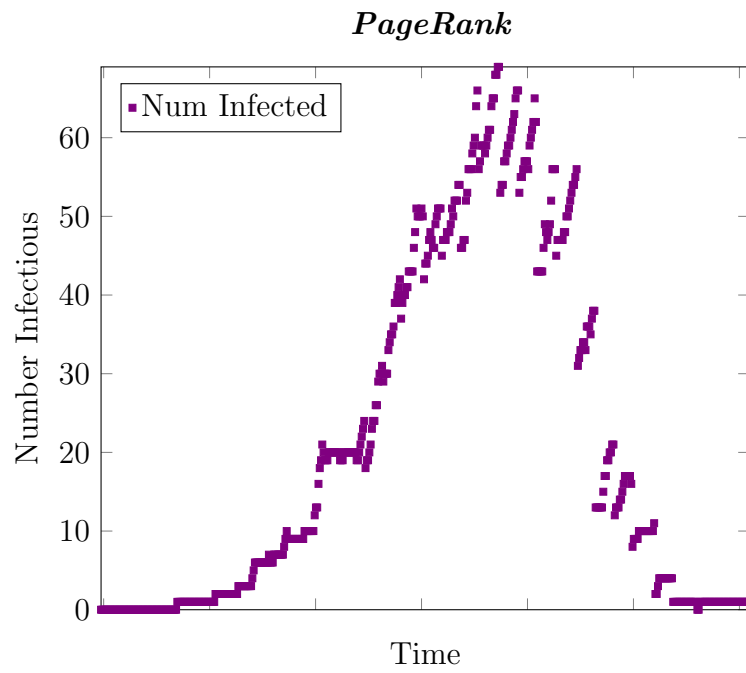


Figure 7.8: Results of a single simulation with testing 90 students per day with a high PageRank

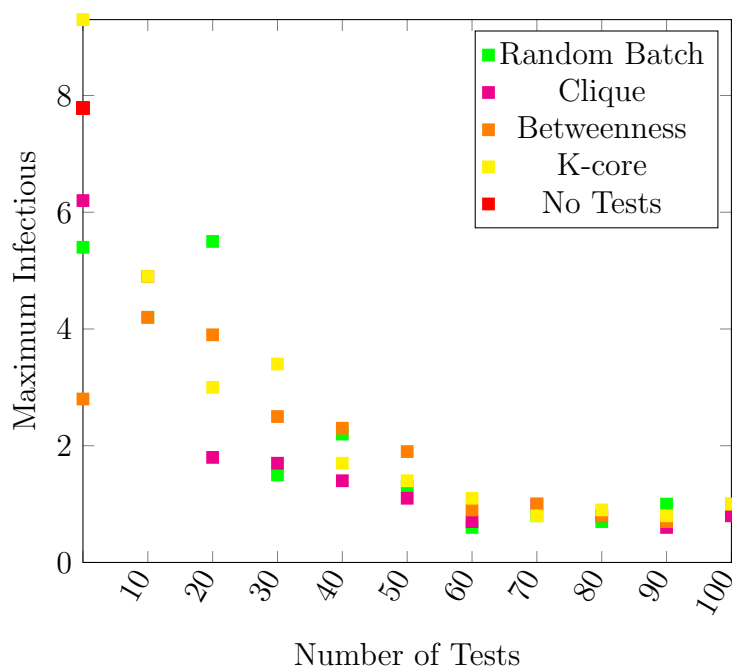
attempted to be as close to the average for the respective testing strategy as possible. Thus, assuming each of these are a typical result of the simulation, we are able to visually see how well each testing strategy does, based on the initial slopes of the outbreaks, the peak number of people infected, and finally how long it takes to have the population fully recover.

## 7.2 Maximum Number Infected

Figure 7.9 through Figure 7.13 takes the peak of the graphs in section 7.1 and averages the value across all ten simulations for a given contact network and testing strategy. This gives a single data point for a given testing strategy for the number of tests given. This is done for all simulations and are plotted with number of tests on the X axis and average peak number infectious on the Y. However, since there are so many testing strategies, these plots are split up into two plots with the three best testing strategies and random batch on the bottom, and the three worst and random batch on the top.

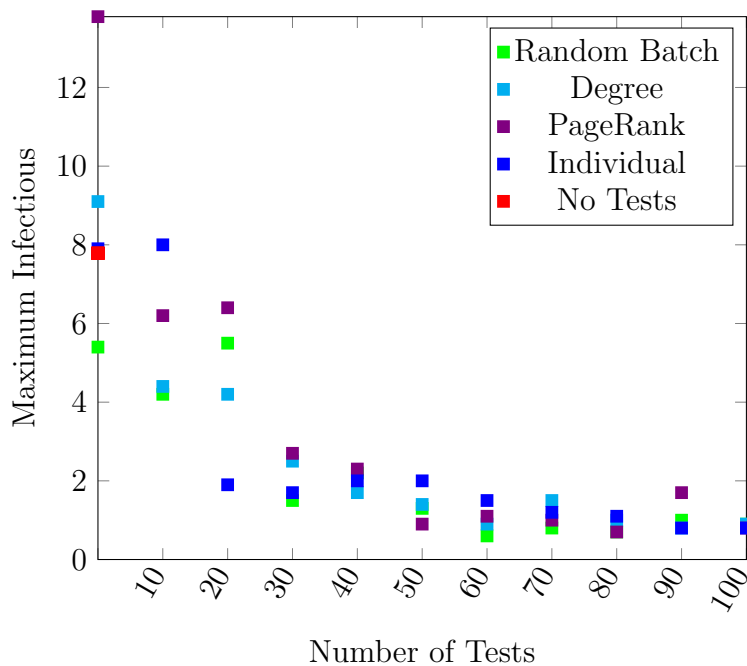


### Workplace 1 Simulations



(a)

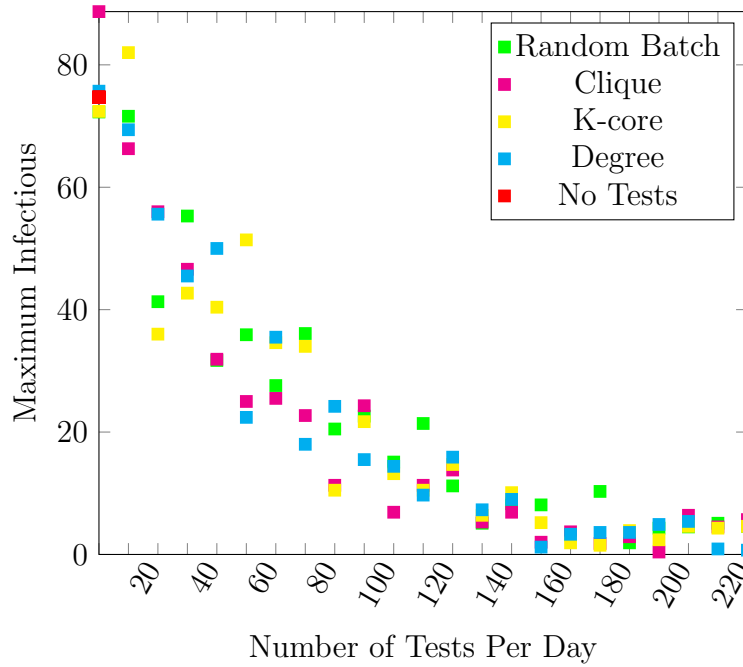
### Workplace 1 Simulations



(b)

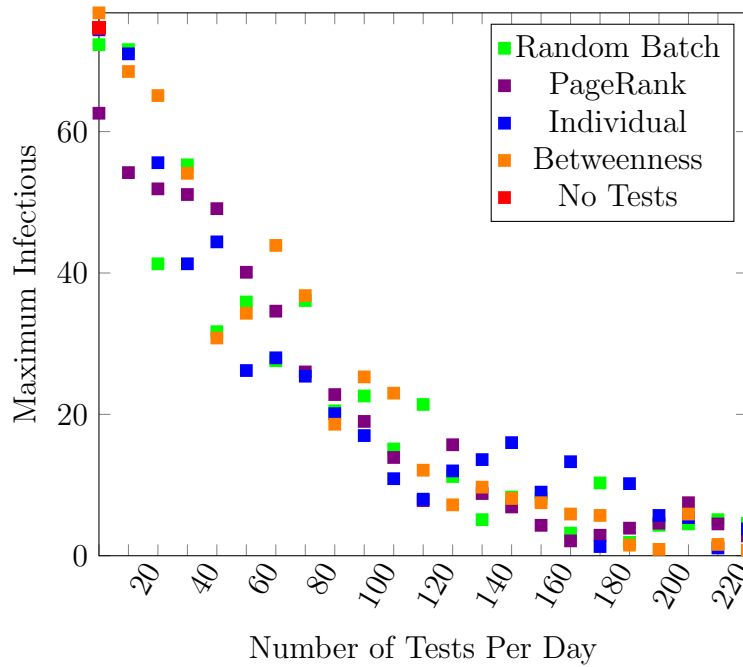
Figure 7.9: Averaged results of the maximum number infected for each strategy on the Workplace 1 contact network

*Workplace 2 Simulations*



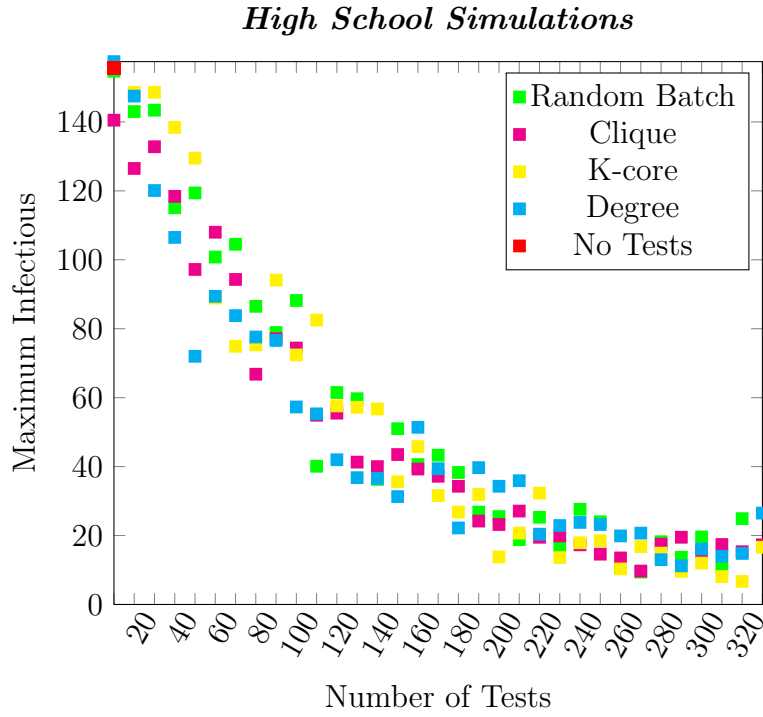
(a)

*Workplace 2 Simulations*

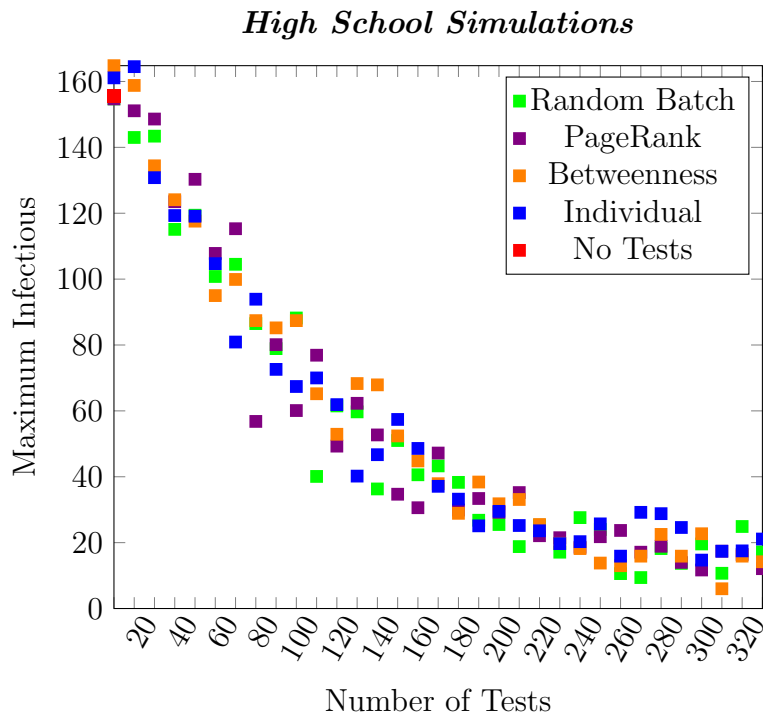


(b)

Figure 7.10: Averaged results of the maximum number infected for each strategy on the Workplace 2 contact network



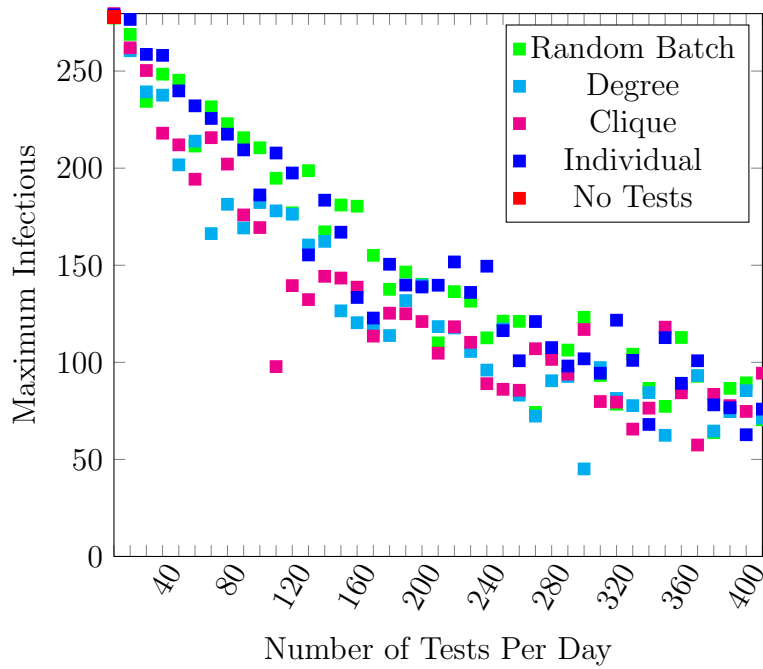
(a)



(b)

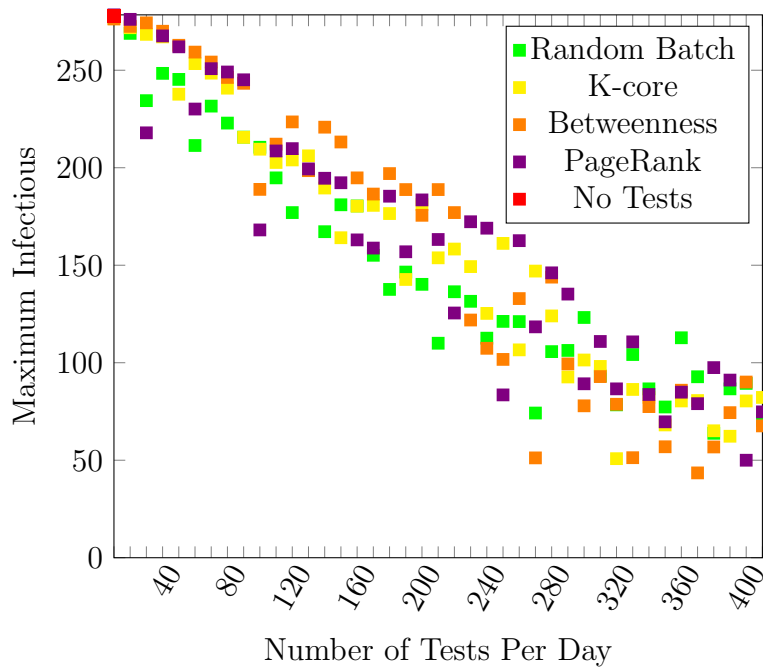
**Figure 7.11:** Averaged results of the maximum number infected for each strategy on the High School contact network

*Conference Simulations*



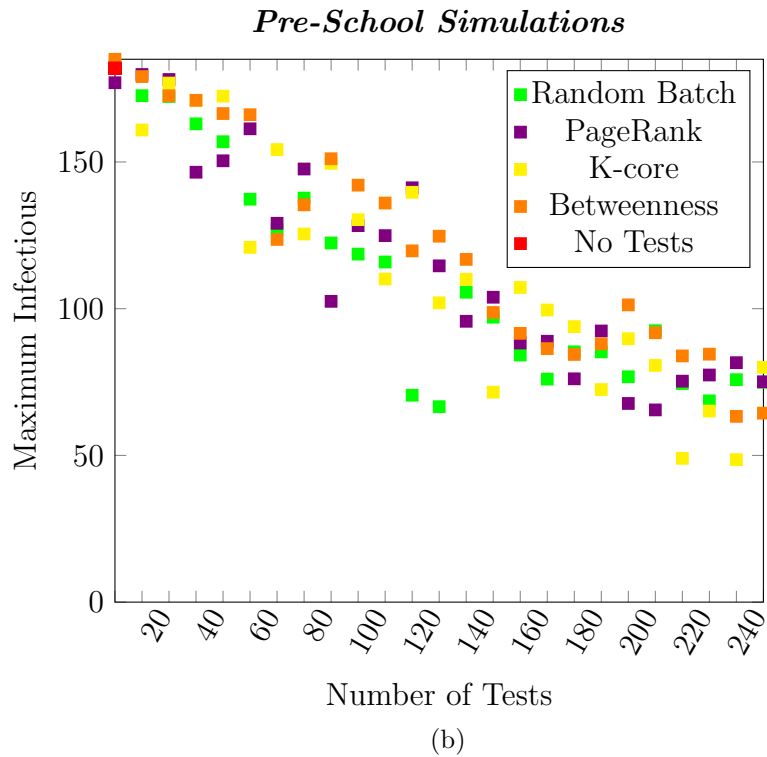
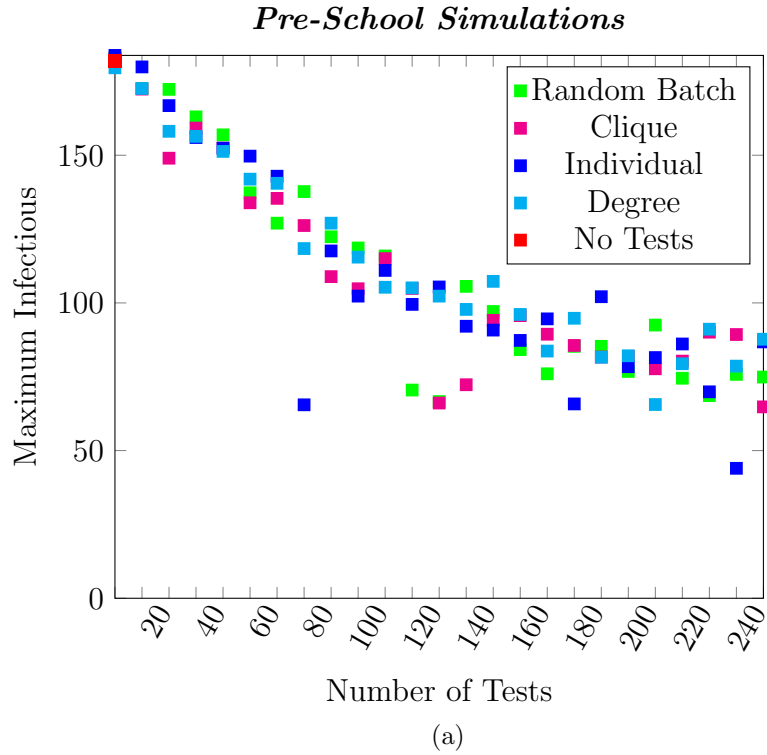
(a)

*Conference Simulations*



(b)

**Figure 7.12:** Averaged results of the maximum number infected for each strategy on the Conference contact network



**Figure 7.13:** Averaged results of the maximum number infected for each strategy on the Pre-School contact network

The method by which the “best” strategies was determined was a simple scoring system. For each point, in the graph, the point with the smallest value for maximum infectious was given a score of 6, the second smallest a score of 5, and so on. Table 7.1 summarizes these scores.

**Table 7.1: A summary of the score for each strategy across each contact network when considering the maximum number infected**

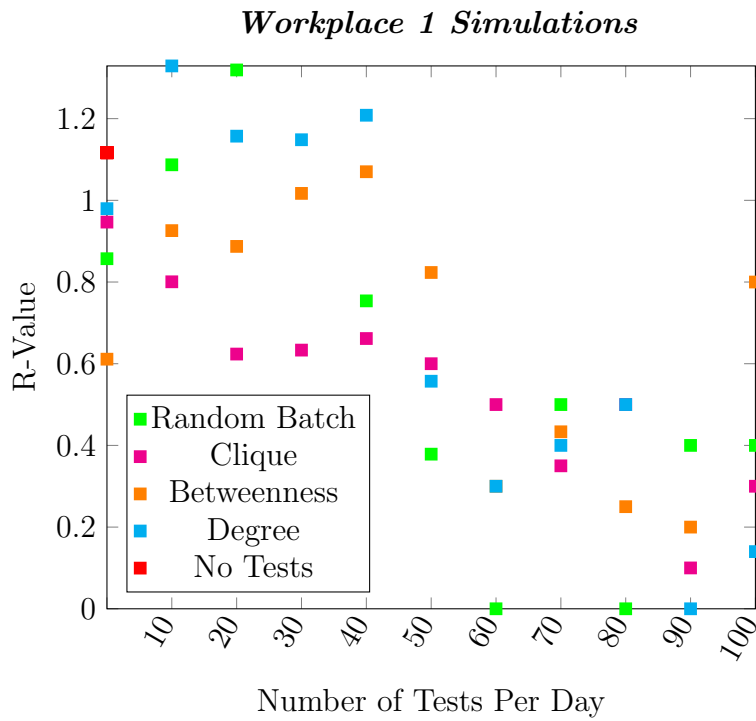
	Workplace 1	Workplace 2	High School	Conference	Pre-School
Random	57	87	131	154	112
Individual	38	91	110	153	108
Degree	42	96	146	225	102
Clique	65	104	158	223	115
K-Core	42	99	152	144	96
Betweenness	49	80	112	136	70
PageRank	38	94	120	114	100

### 7.3 R-value

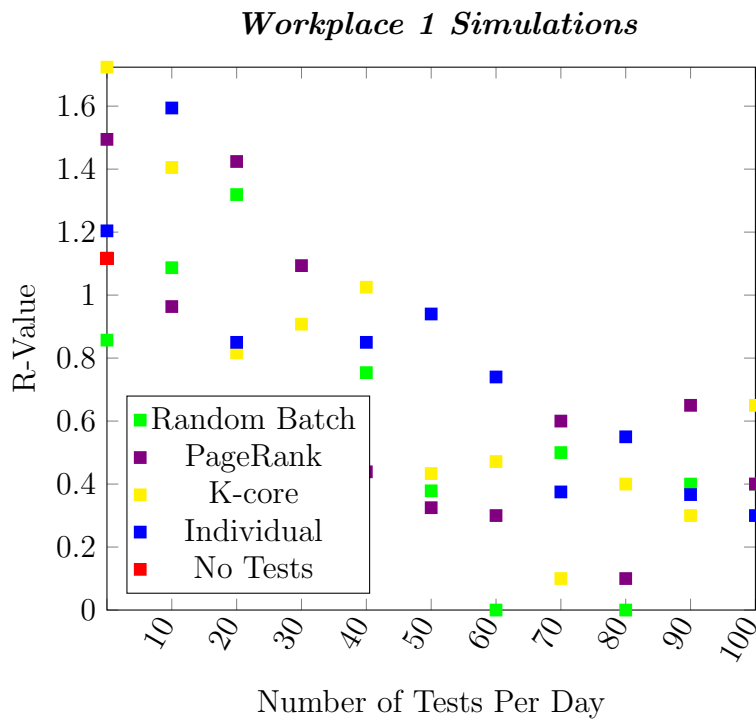
Figure 7.14 through Figure 7.18 takes the average of an R-value for a simulation and plots the average across all simulations, similar to section 7.2. The method which the R values are calculated was by tracking how many people a given individual infects in the duration of their sickness, then the average of all of these values, including those who did not infect anyone, is taken at the end of the simulation. These plots are also split up using the same scoring method of section 7.2. Table 7.2 summarizes these scores.

**Table 7.2: A summary of the score for each strategy across each contact network when considering the R value**

	Workplace 1	Workplace 2	High School	Conference	Pre-School
Random	52	93	126	137	110
Individual	36	85	115	159	105
Degree	43	100	122	205	86
Clique	56	96	130	210	115
K-Core	41	152	157	147	105
Betweenness	44	112	135	156	73
PageRank	42	120	139	134	106



(a)



(b)

**Figure 7.14:** Averaged results of the R-value for each strategy on the Workplace 1 contact network



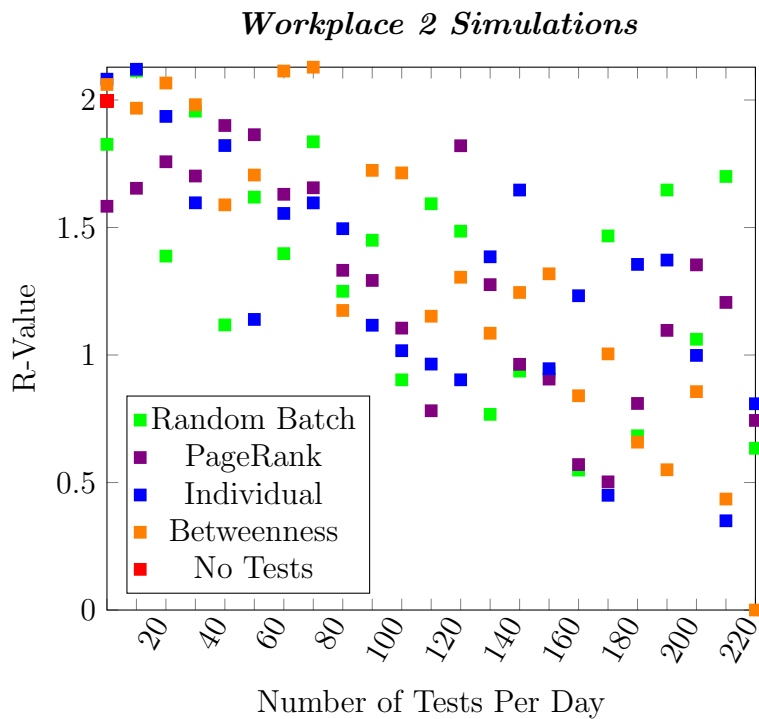
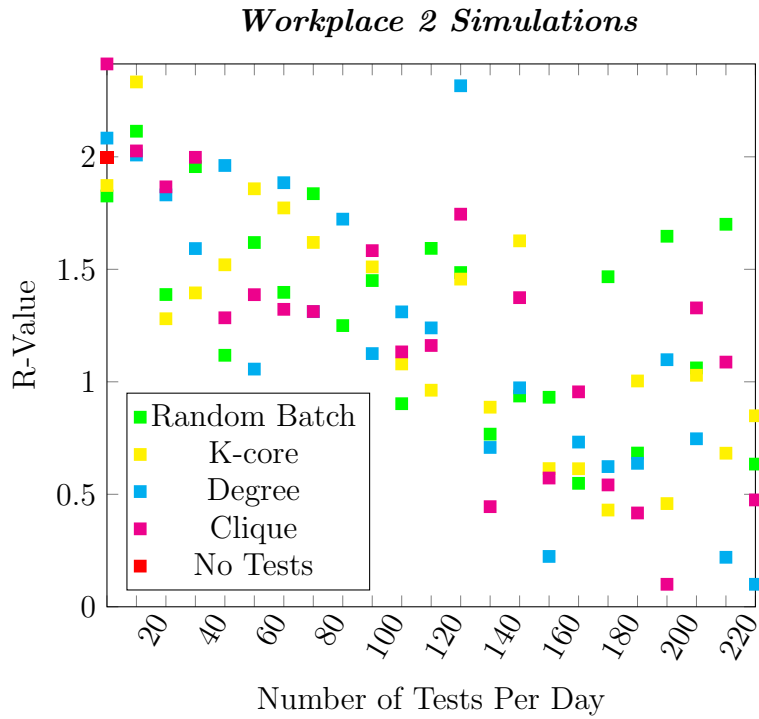
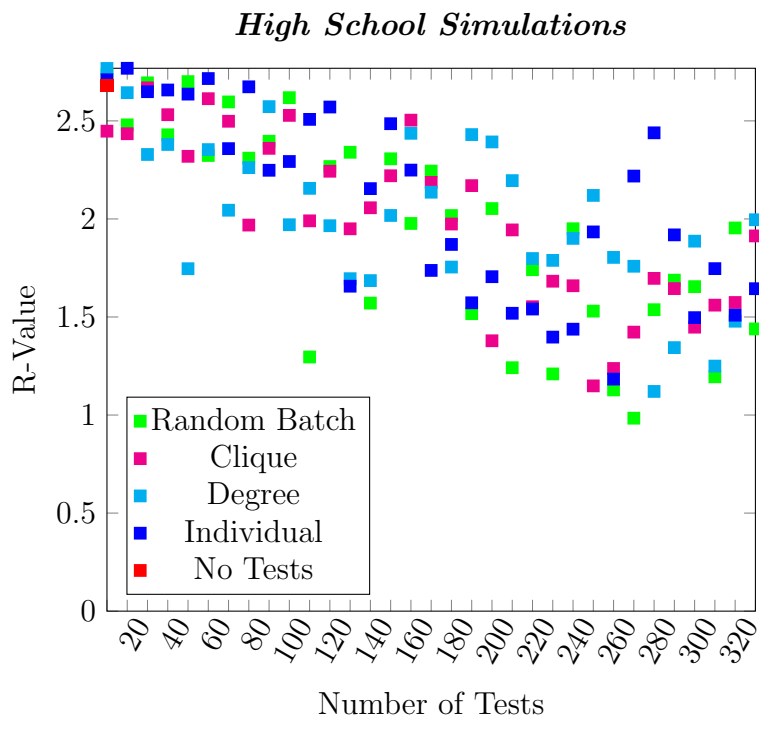
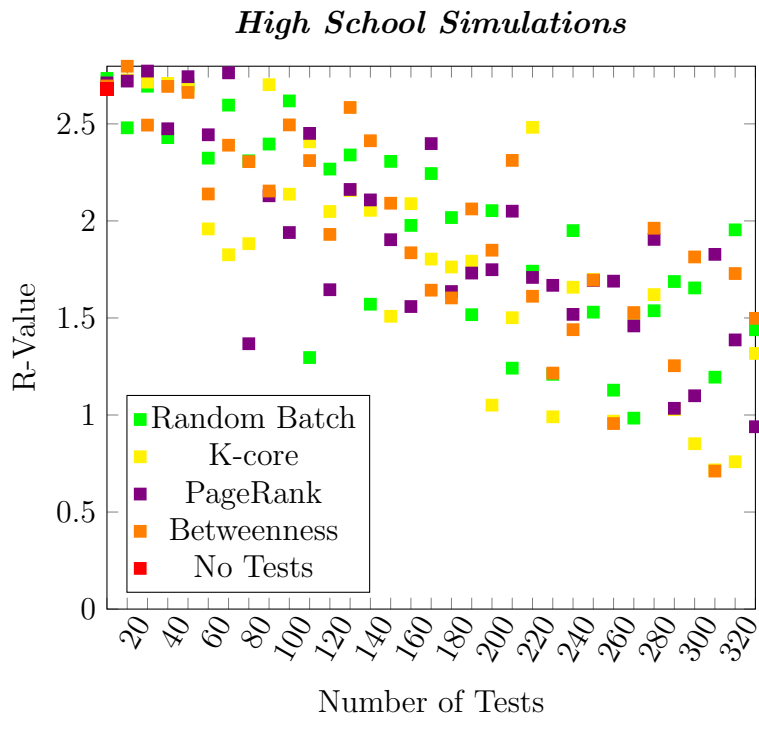
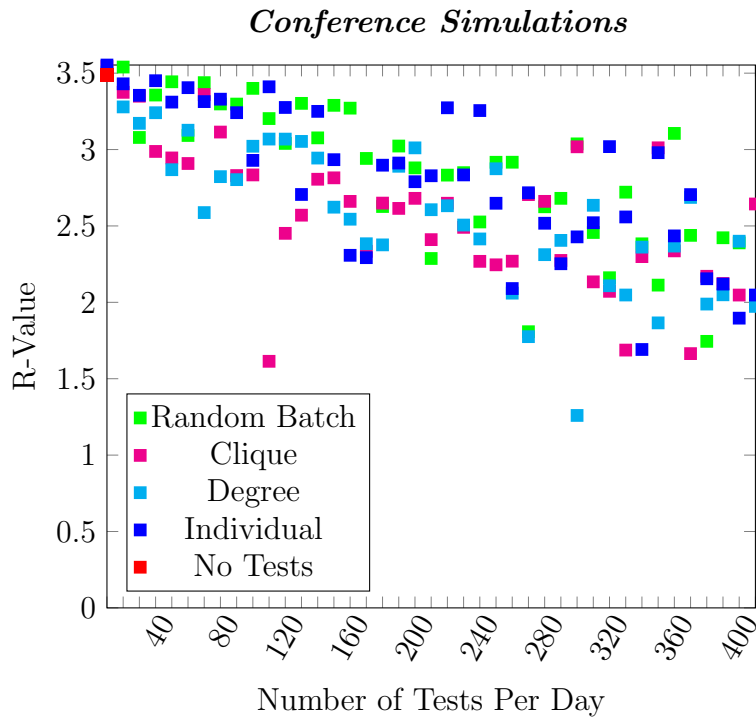


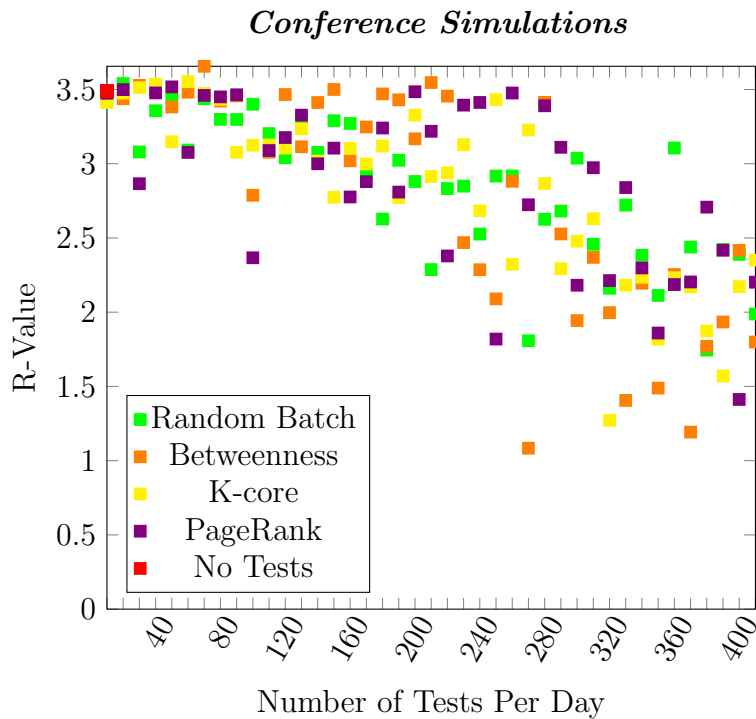
Figure 7.15: Averaged results of the R-value for each strategy on the Workplace 2 contact network



**Figure 7.16:** Averaged results of the R-value for each strategy on the High School contact network

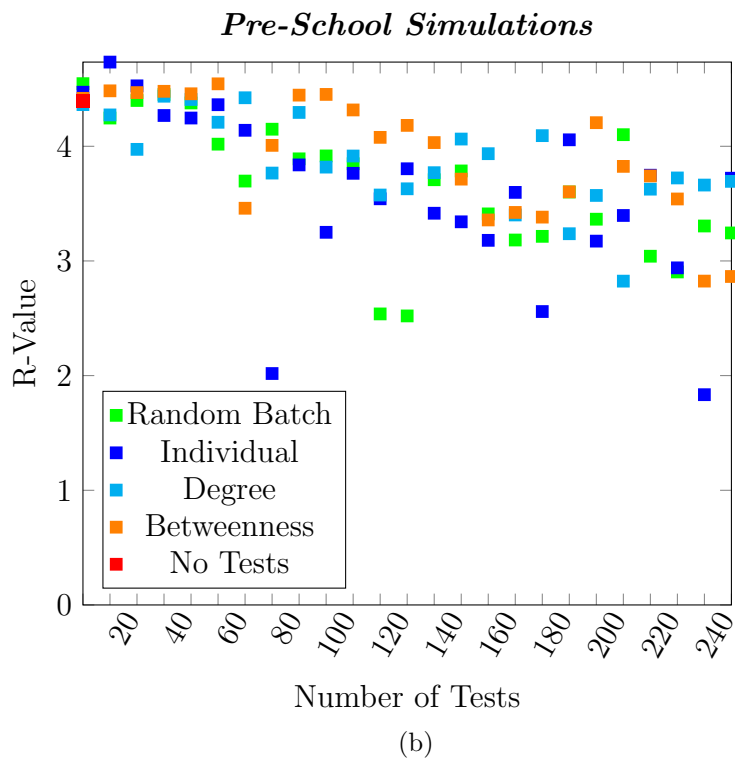
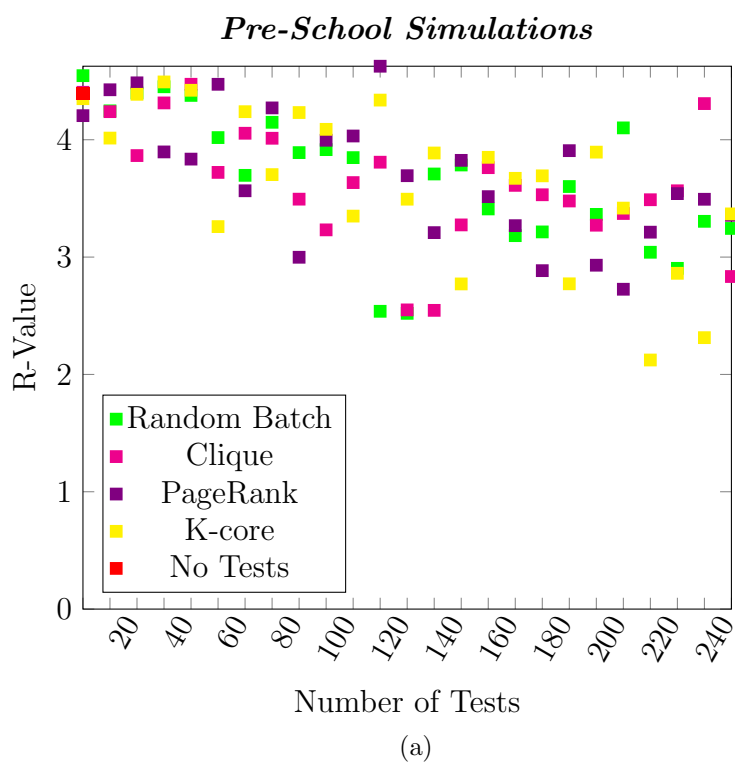


(a)



(b)

Figure 7.17: Averaged results of the R-value for each strategy on the Conference contact network



**Figure 7.18: Averaged results of the R-value for each strategy on the Pre-School contact network**

## Chapter 8

### EVALUATION

The first question when looking at the efficacy of an intervention strategy is how to quantify how well it performs. This research is evaluated using two metrics from the simulations: the maximum number infected and the R-value. The maximum number infected is a good measure to use in the simulations, as it is something we often heard in the duration of the COVID-19 pandemic phrased as “flattening the curve.” The curve referred to is the number of people currently infected, and one of the primary motivations behind the call to flatten this curve is to minimize the impact of those infected with COVID-19 on the capacity of ICUs, and to slow the rate of infection to others. The maximum number infected is therefore a strong indicator of how well an intervention strategy is doing as it gives a measure to how flat the infectious curve is (although is not perfect since some of the currently infected go into the QUARANTINED bucket). Additionally, we will be looking at the trend of the R-values across all simulations. In the case of these simulations, we can interpret the R-value as the number of people we can expect an infectious person to infect. Thus, if an intervention strategy is effective, we should see a decrease in the number of people we can expect an individual to infect.

#### **8.1 Contact Network Factors**

An important feature of the contact networks to note is in their difference in their population size, as well as the setting of the contact network which results in varying densities and average degrees. Looking at Figure 7.9 and Figure 7.14, we see that the

Workplace 1 contact network has a maximum of 100 people which makes analyzing a trend slightly harder due to the wide variation in noise. Additionally, from Chapter 4, we know that the Workplace 1 contact network has a much smaller average degree and density resulting in an end to the simulation much quicker. Another key factor to consider from Chapter 4 is the clique number, with all networks having a relatively small clique number, save for the Pre-School network which has a substantially higher clique number than the rest of the networks.

Additionally, it should be noted that some of these situations are not necessarily realistic to loop repeatedly, most notably the Conference dataset. However, we chose to include this network because it allows us to have an idea of what the outcome of a disease would be like in what is likely near the worst case, in terms of disease transmission. Additionally, the varying densities have a notable impact on the efficacy of how well the testing strategies mitigate the spread, which will be discussed later in this Chapter.

## 8.2 No Testing

The control for each simulation is important to consider, as it shows how much the strategies improve. However, an interesting analysis is to see the peaks of the number infected across each network, as the ratio of maximum number infected to number of people in the network has a wide variation. The values range from about 8% of the population on the Workplace 1 network, to about 75% of the population with the Pre-School network. It is tempting to claim this result comes solely from the density of the network, further inspection says this is not true. For instance, the Workplace 1 and Workplace 2 networks have roughly the same density, yet have infection ratios of 8% and 34%, respectively. Additionally, the Conference dataset has a little less than

twice the density of the High School, yet only saw an 8% increase in the infection ratio. We must therefore conclude that while density of a graph certainly has an impact on the results of this controlled experiment, it is not the most important factor; the most important factor being the average degree of the network.

### 8.3 Random Batch Testing

This strategy, while simplistic, allows us a baseline to compare the other strategies to. If this strategy outperforms another, we can say with confidence that the other strategy is not worth pursuing, as every other strategy is more computationally expensive, and most require information of a contact network which a community may not have.

When analyzing the results of this strategy on its own, we clearly see a decrease in both maximum infections and R-value across all networks with a larger amount of tests. While this fact is intuitive, we can also approximate the minimum number of tests per day required to reach maximum efficacy for the intervention strategies. For the office workplace networks, this value appears to be just over half the network before this value levels out. However, we should note that it appears that as the average degree and density of a network increases, so does the minimum number of tests required. This trend is followed until the number of tests required surpasses the population.

The results from this strategy is very interesting as it takes Kucharski et al.'s findings that random testing has a small impact on the spread of the disease and showcases how much of the population is required to be tested in order to reach maximum efficacy [9].

## 8.4 Individual Testing

This experiment yielded a surprising result as it shows effectively no difference between random batch testing. In fact, when using the scoring system defined in section 7.2, we see random batch testing slightly outperform this one. While the scoring system is notably flawed as many of the points are likely due to random chance, it does allow us to see that this strategy is definitively not more effective than Random Batch testing. The motivation behind this method was to guarantee the entire population tests regularly, yet it apparently does not make a difference as long as the number of tests per day is the same. In reflection, this makes logical sense, because even if we're guaranteeing that the entire population is being tested, each day can still be seen as a pseudo-random batch of individuals resulting in the similar trends that we see in Chapter 7. Additionally, this implementation of this strategy guarantees that a partition of the population is not tested every day which allows exposures to occur. However, a somewhat comforting note is that individual testing performed extremely consistently across each network density. Therefore, this strategy may be considered versatile and have an expected outcome for any network.

## 8.5 Testing Using Graph Properties

### 8.5.1 Degree

This is the first strategy we see consistently outperform the random batch strategy. In fact, based on Table 7.1, we see that this strategy is consistently one of the best strategies of this research. Furthermore, we can see that as the average degree and density increases, this strategy only does better. This fact is especially showcased in Figure 7.12. The success of this strategy likely is due to the repetitive nature of



these experiments, with the replays consistently giving people similar degree values. However, this fact mirrors real life as people tend to follow a consistent routine and are consistently exposed to a similar number of people day to day. For example, in an office workplace a manager would likely consistently have a high degree, and the same can be said for a teacher in a school. Since these individuals were targeted, it appears as though we were able to accomplish our goal and stop the super spreaders of a disease.

### **8.5.2 Clique**

Another exciting result, as we see testing by clique number not only matches the performance seen by degree, but surpasses it in the higher density networks. Using the results of Table 7.1, we see that this testing strategy appears to be the most effective across the table. This result slightly differs from testing by degree, as rather than targeting super spreader individuals, it targets super spreader groups and stops the rapid spread before it happens. The outcome of this testing strategy is very exciting, as we can definitively show that there is value in targeting the specific shape of a graph when trying to mitigate the spread of a disease.

### **8.5.3 K-Core**

While it appears that this strategy performed well, we can see that as the average degree and density of the graph increased, the performance of this strategy decreased. In fact, in the Conference and Pre-School networks, this strategy performed worse than random batch testing. This fact is somewhat surprising, as intuitively we would think that in high density networks, the k-cores would look similar to the cliques of a graph. However, upon reflection we may deduce that there may not likely be

high k-core values in contact networks such as the conference. For example, an individual in a given day may talk to many people in a conference, however there is no guarantee that those individuals also talked to the same people, and have a similar degree. Therefore, based on the results of these simulations, we may conclude that this strategy is ineffective in higher density networks.

#### **8.5.4 Betweenness Centrality**

A slightly different and less intuitive approach, this strategy attempted to slow the spread by targeting those individuals who were “in between” portions of a graph. Ideally, these would be the people who connect different portions of the contact network, i.e. those who carry the disease across social groups. However, we can see that this strategy consistently did not perform well. Reflecting on this strategy, this makes intuitive sense. Hypothetically, if there are 2 cliques in a graph and we are only testing the people who connects those cliques, the rest of the clique still can freely spread the disease. This strategy was outperformed by random batch across the table, so we may confidently say that this is an ineffective testing strategy.

#### **8.5.5 PageRank**

Another unique approach, the motivation of this strategy was to try to predict where the disease may be. However, this is another strategy that was typically outperformed by random batch testing. There is not necessarily an intuitive explanation to this, but we may conclude that PageRank simply is not an effective algorithm in determining the super spreaders of a disease.

## 8.6 Evaluation by R-Value

It should be noted that most of the evaluation was focused on the maximum number infected rather than the R-value. This was done upon reflection of how the R-value was calculated. While  $R_0$  has been shown to be a strong indicator of how infectious a disease is in the world, it does not necessarily help us with the setup of this experiment. This is because once the infection reaches herd immunity, we do not expect each infectious person to spread the disease to another person. This means that the R-value peaks when the number infected peaks, then lowers over time. This can affect the outcome of this value in several ways, depending on how many people are currently infectious. However, this is still an important metric to look at as we can clearly see that the intervention strategies lowers this value with more tests.

## Chapter 9

### CONCLUSION

This research explored numerous testing intervention strategies, and determined the viability of each strategy across various contact networks. The complexity of each strategy ranged from simply randomly testing a portion of the population every day to adapting an algorithm originally intending for ranking the relevance of web pages. Yet, this research often found that complexity does not necessarily mean effective. In fact, the more complex intervention strategies performed the worst, and the simpler testing strategies were difficult to beat. Although, this research was able to use the underlying properties of the contact network graph and say with confidence that testing by clique number is the most effective among the strategies explored across all contact networks.

Even with the success of the testing strategies, we must also reflect upon the shortcomings of the simulations. First was the assumption that the behavior of individuals would not change with an outbreak of an extremely infectious disease, save for quarantining if an individual tests positive. While this behavior to reduce the spread such as mask wearing and social distancing may be able to be adapted to these contact networks, it was intentionally omitted as this research's primary focus was to evaluate the performance of the testing strategies, not to evaluate the effectiveness of human behavior. Leaving out symptomatic and asymptomatic infectious individuals was done for the same reason; including symptomatic individuals would likely improve the testing and quarantining process as the strategies would be forced to always test symptomatic people.

This research certainly yielded some interesting results, however we must remember to do our part in mitigating the spread of COVID-19 by listening to public health officials and following proven methods such as wearing masks and social distancing until they say it is safe to return to normal.

## Chapter 10

### FUTURE WORK

Given the results of the experiments, we must remember the assumptions made with this experiment. Much of the potential future work may change these assumptions in order to make a more realistic simulation and explore a different aspect of disease transmission. The first assumption is that the behavior of the contact networks would remain the same in a pandemic. As we have seen, the behaviors of individuals does not remain the same in a pandemic situation, with people limiting contact with each other as much as possible. Additionally, we assume a constant infection probability, when in reality it differs depending on the individual infected, the distance of interactions, if one or both of the individuals are wearing a face covering, and so on. One of the most notable assumptions is that there is no distinguishment between the asymptomatic infectious individuals and symptomatic infectious individuals. In the simulations for this research, every individual who is infected is treated as asymptomatic in that they do not self-quarantine or get tested if they feel sick. Lastly, we assumed a closed network for this research, meaning there is no outside influence for infections. However, an improvement that may be made is to combine the networks and make households to make the final simulation more granular. All of these considerations allow interesting directions to be taken for future work.

Furthermore, more future work may include making more robust strategies. This may include combining the best graph-based strategies, perfect contact tracing, or even training models to predict the super spreaders in a contact network, although the latter may be difficult with limited contact network data. One of the more exciting possibilities is developing a pooling strategy. Part of the success of Duke

University's COVID-19 response is attributed to the fact that they implemented a pooling strategy which allowed them to do more tests per day [3]. Yet, these tests were pooled randomly, and an interesting direction may be to try to use the properties of a graph to decide which samples to pool together.

Finally, since the beginning of this research, a vaccine has been developed. Experiments that simulate vaccinating the population with the 2-dose vaccine like Moderna or Pfizer with varying efficacy would be an interesting direction to take this research as well. In short, this research could go in many directions and is limited to what aspect of the simulation is wanted to make more realistic or how elaborate the proposed intervention strategy would be.

## BIBLIOGRAPHY

- [1] S. C. Carlson. graph theory | Problems & Applications.
- [2] E. Estrada. COVID-19 and SARS-CoV-2. Modeling the present, looking at the future. *Physics Reports*, 869:1–51, 2020.
- [3] E. Ferreri. Students 'Nothing Short of Phenomenal' in Minimizing COVID Spread, Nov. 2020.
- [4] M. Génois and A. Barrat. Can co-location be used as a proxy for face-to-face contacts? *EPJ Data Science*, 7(1):11, Dec. 2018.
- [5] H. W. Hethcote. Three Basic Epidemiological Models. In S. A. Levin, T. G. Hallam, and L. J. Gross, editors, *Applied Mathematical Ecology*, Biomathematics, pages 119–144. Springer, Berlin, Heidelberg, 1989.
- [6] H. W. Hethcote. The Mathematics of Infectious Diseases. *SIAM Review*, 42(4):599–653, Jan. 2000.
- [7] N. Imai, A. Cori, I. Dorigatti, M. Baguelin, C. A. Donnelly, and S. Riley. Report 3: Transmissibility of 2019-nCoV. page 5, Jan. 2020.
- [8] A. Iwasaki. What reinfections mean for COVID-19. *The Lancet Infectious Diseases*, 21(1):3–5, Jan. 2021. Publisher: Elsevier.
- [9] A. J. Kucharski, P. Klepac, A. J. K. Conlan, S. M. Kissler, M. L. Tang, H. Fry, J. R. Gog, W. J. Edmunds, J. C. Emery, G. Medley, J. D. Munday, T. W. Russell, Q. J. Leclerc, C. Diamond, S. R. Procter, A. Gimma, F. Y. Sun, H. P. Gibbs, A. Rosello, K. van Zandvoort, S. Hué, S. R. Meakin, A. K. Deol, G. Knight, T. Jombart, A. M. Foss, N. I. Bosse, K. E. Atkins, B. J.



- Quilty, R. Lowe, K. Prem, S. Flasche, C. A. B. Pearson, R. M. G. J. Houben, E. S. Nightingale, A. Endo, D. C. Tully, Y. Liu, J. Villabona-Arenas, K. O'Reilly, S. Funk, R. M. Eggo, M. Jit, E. M. Rees, J. Hellewell, S. Clifford, C. I. Jarvis, S. Abbott, M. Auzenbergs, N. G. Davies, and D. Simons. Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study. *The Lancet Infectious Diseases*, 20(10):1151–1160, Oct. 2020.
- [10] L. M. Kucirka, S. A. Lauer, O. Laeyendecker, D. Boon, and J. Lessler. Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction–Based SARS-CoV-2 Tests by Time Since Exposure. *Annals of Internal Medicine*, May 2020.
- [11] R. H. S. MD. Which test is best for COVID-19?, Aug. 2020.
- [12] L. A. Meyers. Contact network epidemiology: Bond percolation applied to infectiousdisease prediction and control. *Bulletin of the American Mathematical Society*, 44(01):63–87, Oct. 2006.
- [13] R. M. Mullner. Epidemiology.
- [14] R. E. Rakel. Diagnosis - Mental examination, Nov. 2018.
- [15] N. Shah, H. Shah, M. Malensek, S. L. Pallickara, and S. Pallickara. Network analysis for identifying and characterizing disease outbreak influence from voluminous epidemiology data. In *2016 IEEE International Conference on Big Data (Big Data)*, pages 1222–1231, Dec. 2016.
- [16] R. H. Shmerling. Which test is best for COVID-19?, Aug. 2020.

- [17] C. Siu and T. Migler. Vaccination Strategies on a Robust Contact Network. In H. Cherifi, S. Gaito, J. F. Mendes, E. Moro, and L. M. Rocha, editors, *Complex Networks and Their Applications VIII*, pages 313–324, Cham, 2020. Springer International Publishing.
- [18] G. St-Onge, V. Thibeault, A. Allard, L. J. Dubé, and L. Hébert-Dufresne. Social confinement and mesoscopic localization of epidemics on networks. *Physical Review Letters*, 126(9):098301, Mar. 2021. arXiv: 2003.05924.
- [19] A. Stopczynski, V. Sekara, P. Sapiezynski, A. Cuttone, M. M. Madsen, J. E. Larsen, and S. Lehmann. Measuring Large-Scale Social Networks with High Resolution. *PLoS ONE*, 9(4):e95978, Apr. 2014.