

SIMULATING EPIDEMICS AND INTERVENTIONS ON HIGH RESOLUTION
SOCIAL 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
Christopher E. Siu
June 2019

© 2019
Christopher E. Siu
ALL RIGHTS RESERVED

COMMITTEE MEMBERSHIP

TITLE: Simulating Epidemics and Interventions on
High Resolution Social Networks

AUTHOR: Christopher E. Siu

DATE SUBMITTED: June 2019

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

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

COMMITTEE MEMBER: Nathaniel Martinez, M.D., Ph.D.
Assistant Professor of Biological Sciences

COMMITTEE MEMBER: Julie Workman
Lecturer, Dept. of Computer Science

ABSTRACT

Simulating Epidemics and Interventions on High Resolution Social Networks

Christopher E. Siu

Mathematical models of disease spreading are a key factor of ensuring that we are prepared to deal with the next epidemic. They allow us to predict how an infection will spread throughout a population, thereby allowing us to make intelligent choices when attempting to contain the disease. Whether due to a lack of empirical data, a lack of computational power, a lack of biological understanding, or some combination thereof, traditional models must make sweeping assumptions about the behavior of a population during an epidemic.

In this thesis, we implement granular epidemic simulations using a rich social network constructed from real-world interactions. We develop computational models for three diseases, and we use these simulations to demonstrate the effects of twelve potential intervention strategies, both before and during a hypothetical epidemic. We show how representing a population as a temporal graph and applying existing graph metrics can lead to more effective interventions.

ACKNOWLEDGMENTS

Thanks to:

- To my advisor, Professor Theresa Migler-VonDollen: you have given me so much support and encouragement over the last three years. You helped me discover my interest in theoretical computer science, and you constantly demonstrate the quality I strive for as an educator.
- To Professor John Clements: in many ways, your classes were my first introduction to navigating academia, both as a graduate student and as a novice researcher.
- To Professor Nathaniel Martinez: I am not an epidemiologist; I am not even a biologist. Your expertise and input provided invaluable guidance towards pragmatic, relevant applications for my work.
- To Professors Julie Workman and Paul Hatalsky: you taught me everything I know about the fundamentals of computer science. You helped me realize my interest and belonging in this field, and, because of you, the dream job I didn't yet know what I wanted practically fell into my lap when I was nineteen.
- To Jon, Caleb, and Chris: I have wasted countless hours in video games and on the Internet because of you three, and I don't regret any of it. You should have stayed in grad school with me — who needs money?
- To all my students: you have all played a role in motivating me to finish this thesis. Most of you motivated me by reminding me how much I enjoy teaching. Some of you motivated me by defending your theses before I could.

- To my parents, César and Sandra, and my sisters, Karina and Marissa: you were my sole teachers for eighteen years. You took me on all your vacations, you made sure I didn't stay in my room reading books all day, you figured out what I should study, and you gave me your unwavering love. Thank you.

TABLE OF CONTENTS

	Page
LIST OF TABLES	x
LIST OF FIGURES	xi
CHAPTER	
1 Introduction	1
2 Background	3
2.1 Epidemiological Models	3
2.2 Compartmental Models	4
2.2.1 Susceptible-Infected-Recovered	5
2.2.2 Susceptible-Infected-Susceptible	5
2.2.3 SIRS, SEIR, MSIR, and Other Models	6
2.3 Full Mixing Simulations	6
2.4 Epidemiological Metrics	7
2.4.1 Basic Reproduction Number	8
2.4.2 Epidemic and Herd Immunity Thresholds	8
2.5 Graphs	9
2.6 The Copenhagen Network Study	11
3 Related Works	15
3.1 Disease Spreading on Graphs	15
3.2 Disease Spreading with Mobility Patterns	16
3.3 Disease Spreading on Graphs with Mobility Patterns	18
3.4 Span-Cores in Temporal Networks	19
4 Implementation	21

4.1	Preprocessing	21
4.2	Graph Representation	22
4.3	Simulation Architecture	24
4.3.1	The Flu Model	24
4.3.2	The Measles Model	26
4.3.3	The Norovirus Model	28
4.4	Interventions	31
4.4.1	Uniform Vaccination	33
4.4.2	Vaccination by Degree Distribution	33
4.4.3	Vaccination by k -Cores	36
4.4.4	Vaccination by Density Decomposition	37
4.4.5	Quarantining	38
4.5	Summary of Simulations	39
5	Methodology	42
5.1	Baseline Models	42
5.2	Models with Interventions	43
6	Results	46
6.1	Baseline Scenarios	46
6.1.1	Influenza	46
6.1.2	Measles	49
6.1.3	Norovirus	50
6.2	Uniform Vaccination	53
6.3	Targeted Vaccination	56
6.3.1	Vaccination by Degree	56
6.3.2	Vaccination by k -Core	58

6.3.3	Vaccination by Ring	58
6.4	Quarantine Scenarios	61
7	Conclusion	64
7.1	Future Work	64
7.1.1	Ineffective Vaccinations	64
7.1.2	Additional Community Detection Algorithms	64
7.1.3	Combinations of Interventions	65
7.1.4	Simultaneous Infections	65
7.1.5	External Contacts and Deaths	66
7.2	Summary of Contributions	66
	References	67

LIST OF TABLES

Table		Page
1	Example Format of Bluetooth Data	40
2	Summary of Simulated Interventions	41
3	Summary of Simulated Scenarios	45

LIST OF FIGURES

Figure	Page
1 Compartments of Bernoulli's Model	4
2 The SEIR Model	5
3 Drawing of The Seven Bridges of Königsberg	9
4 Graph of the Seven Bridges of Königsberg	10
5 Cliques and Stars	11
6 The SEIR Model Applied to a Social Network	11
7 Selected Interactions in the Copenhagen Network Study	12
8 Interactions by Time in the Copenhagen Dataset	14
9 Numbers of Unique Interactions in the Copenhagen Dataset	14
10 Graph Snapshot of One Timestamp	23
11 Cluster Detection over Four Snapshots	29
12 Example Degree Distribution	34
13 Example Flattened Graph	35
14 Example k -Cores	36
15 Affected Individuals in a Baseline Influenza Scenario	47
16 Affected Individuals in Baseline Influenza Scenarios	48
17 Infected Individuals in Baseline Influenza Scenarios	48
18 Affected Individuals in Baseline Measles Scenarios	50
19 Exposed and Infected Individuals in Baseline Measles Scenarios	51
20 SEIR Compartments in a Baseline Measles Scenario	51
21 Affected Individuals in Baseline Norovirus Scenarios	52
22 Exposed and Infected Individuals in Baseline Norovirus Scenarios	52

23	SEIR Compartments in a Uniform Vaccination Scenario	54
24	Affected Individuals in Uniform Vaccination Scenarios	55
25	Summary of Uniform Vaccination Scenarios	55
26	Affected Individuals when Vaccinated by Degree	57
27	Affected Individuals when Vaccinated by k -Core	59
28	Affected Individuals when Vaccinated by Ring	60
29	Affected Individuals in Quarantine Scenarios	63
30	Affected Individuals in Vaccination-Quarantine Scenarios	63

Chapter 1

INTRODUCTION

“Epidemich...a sicknesse common vnto all people, or to the moste part of them.”

— Thomas Lodge, *A Treatise of the Plague*, 1603

From the Greek “*επιδημιος*” via the French “*épidémique*”, the word “epidemic” was first used in the English language in 1603, and it has come to describe a disease which has quickly infected an unusually high number of people within some well-defined population (OED Online, n.d.). Epidemics may further be termed “pandemics” to distinguish those diseases which have spread on a continental or global scale, and epidemics may be contrasted with “epizootics”, which affect animals.

Epidemiology, then, is the study of diseases that infect people, including the biological and social mechanisms involved in their outbreak, transmission, containment, and, hopefully, eventual eradication. Of that last goal, to date, only one disease that affects humans has been eradicated: a naturally occurring case of smallpox has not been reported since October of 1977, the conclusion of almost two hundred years of work immunizing the general population (Fenner, Henderson, Arita, Jezek, & Ladnyi, 1988). This required a dedicated program on the part of the World Health Organization, involving mass vaccinations, health surveillance, and targeted interventions on continental scales over more than a decade.

Should we ever be confronted with another smallpox-esque disease, computer models will be crucial to effective coordination and utilization of our limited resources, a logistical problem which, in many respects, we are not currently prepared for (Gates, 2015). Developing models to simulate and predict the spread of diseases could allow

us to contain outbreaks before they become epidemics, and, failing that, to determine the appropriate methods of intervention. Mathematical approaches for studying diseases have existed for just about as long — slightly longer, in fact — as the vaccination processes that can prevent them (Dietz & Heesterbeek, 2002). However, because of the complexity of the involved social and biological mechanisms, traditional mathematical models must make a handful of key unrealistic assumptions. They cannot account for the transmission of disease on an individual, per-person scale.

In this thesis, we implement granular epidemic simulations using a rich social network constructed from real-world interactions. We use these simulations to demonstrate the effects of various intervention strategies, both prior to and during a hypothetical outbreak. Specifically, we represent the population as a temporal graph and investigate the possibility that existing graph metrics can inform more effective interventions. In Chapter 2, we give a brief historical background to the separate disciplines of epidemic modeling and graph theory, and we introduce the dataset on which we build our simulations. In Chapter 3, we describe similar approaches which have become more common recently, as both computing power and relevant data have become more available. Then, in Chapter 4 and Chapter 5, we build and run computational models for simulating three different diseases and twelve different intervention strategies. The results of these simulations are discussed in Chapter 6. Finally, we summarize our findings and enumerate potential followup work in Chapter 7.

Chapter 2

BACKGROUND

2.1 Epidemiological Models

Use of mathematical models to characterize the spread of disease and explore potential treatments dates to around 1766, when Daniel Bernoulli developed a simplistic model to investigate the feasibility of inoculation against smallpox (Dietz & Heesterbeek, 2002), inoculation being the primitive precursor to vaccination. He reasoned that the population could be divided into two groups: those who had not yet contracted the disease, and were therefore susceptible, and those who had already survived the disease, and were therefore immune. In Bernoulli's model, surviving smallpox moves an individual from the susceptible group to the immune group; death removes an individual from their current group.

By combining a differential equation describing the rates at which these two groups grew and shrank with the crude empirical estimates of the day, Bernoulli estimated that the median age could be more than doubled, increased roughly from 11.5 to 25.5 years, if smallpox were eliminated as a cause of death. Life expectancy at birth would increase from 26 years and 7 months to 29 years and 9 months. It is, unfortunately, difficult to validate these estimates. Smallpox would not be eradicated in Europe until nearly two hundred years later (Fenner et al., 1988), by which time other factors — notably, childhood mortality rates — had drastically changed.

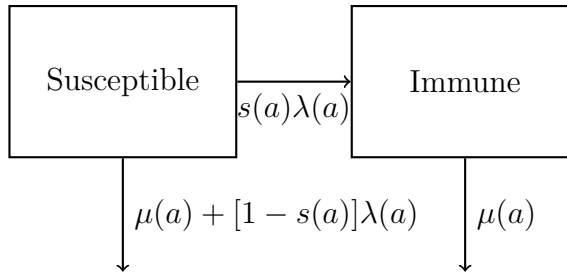


Figure 1: Compartments of Bernoulli’s Model. Here, a is age, $\lambda(a)$ is probability of infection, $s(a)$ is that of survival, and $\mu(a)$ is that of death by other causes. (Dietz & Heesterbeek, 2002)

2.2 Compartmental Models

In the modern nomenclature, Bernoulli’s work employs a simple *compartmental model*: a model that divides the population into “compartments”, depending on how the disease has already affected or might potentially affect them. The movement of an individual between compartments is governed by a simple, typically probabilistic, state machine. Namely, Bernoulli’s model has “Susceptible” and “Immune” compartments, reproduced in Figure 1.

Kermack and McKendrick in 1927 took a similar approach (Kermack & McKendrick, 1927), but they introduced a series of intermediate compartments for those individuals currently at various stages of infection at a given time. They labeled their individuals’ states as either “Unaffected”, “Ill”, or “Removed”. As with Bernoulli’s model, removal can mean recovery or death; neither Bernoulli’s nor Kermack and McKendrick’s models include the possibility of the susceptible population’s growing through birth.

2.2.1 Susceptible-Infected-Recovered

These historical models are most similar to the modern Susceptible-Infected-Recovered, or SIR, model, which has direct analogs to their compartments. In this model, every individual is either at risk of contracting the disease, currently infected and capable of transmitting the infection, or immune (whether by recovery or by death).

2.2.2 Susceptible-Infected-Susceptible

Depending on the particular disease that we are interested in modeling, the Susceptible-Immune compartments may be too simplistic, and the Susceptible-Infected-Recovered compartments, outright inaccurate. Surviving influenza, for example, does not make an individual immune to alternative strains and future mutations of the virus (Kramer et al., 2018). If we were interested in simulating influenza on time scales long enough for this to matter (and if we made the questionably realistic assumption that all strains could be simulated as the same disease), we might apply the Susceptible-Infected-Susceptible model (SIS), in which individuals are once again at risk after having recovered.

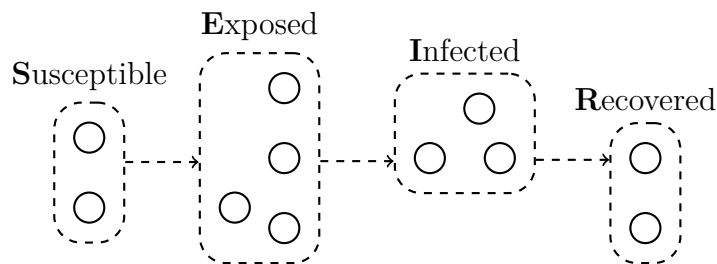


Figure 2: The SEIR Model. This example population is divided into Susceptible, Exposed, Infected, and Recovered compartments.

2.2.3 SIRS, SEIR, MSIR, and Other Models

There are, unsurprisingly, a plethora of variations on the basic compartmental models, depending the disease and desired granularity of the particular application. The previously mentioned influenza outbreaks, whose strains can change from year to year, also tend to occur in seasonal patterns (Krammer et al., 2018), so a model that allows for a temporary period of effective immunity may be more appropriate. Malaria has a non-trivial incubation time (Phillips et al., 2017), so a model should allow for the possibility that an individual has been infected but does not yet exhibit any symptoms. Infants inherit about nine months of immunity to measles from their mother (Rota et al., 2016), so a model may need to include this initial period. Models such as Susceptible-Infected-Recovered-Susceptible (SIRS), Susceptible-Exposed-Infected-Recovered (SEIR), and Maternal-Susceptible-Infected-Recovered (MSIR) attempt to account for these situations, respectively. An example of the SEIR model is depicted in Figure 2.

2.3 Full Mixing Simulations

Having developed the necessary compartments and associated differential equations, the afore-mentioned Kermack and McKendrick applied their SIR-esque mathematical model to bubonic plague, for which they had empirical data from a relatively recent outbreak in then-Bombay. Upon this application, they immediately identified the key assumption that must be made by all traditional epidemic models: “We are, in fact, assuming...that the uninfected population was uniformly susceptible...that all susceptible rats in the island had an equal chance of being infected...[and] that the condition approximated to one of contact infection.” (Kermack & McKendrick, 1927) That is to say, these *full mixing* models assume that every individual has an

equal chance of coming into contact with any other member of the population, and, thus, an equal chance of contracting the disease via that contact (itself an unrealistic assumption in the case of the bubonic plague, which spreads via fleas that live on rats (McEvedy, 1988)). Kermack and McKendrick therefore conclude that “None of these conditions are strictly fulfilled and consequently the numerical equation can only be a very rough approximation.”

Nevertheless, traditional epidemic modeling continues to use the same general approach. Given a sequence of compartments describing the possible states and transitions of infected individuals, they simulate the spread of disease by using differential equations to describe the movements of individuals between compartments. To some extent, these systems of equations can be extended to approximate the interactions of individuals (Murray, Stanley, & Brown, 1986). In effect, impersonal though it may seem, we may think of individuals in different states as a mix of different fluid chemicals in a reaction-diffusion system. In this application, the “chemical reactions” are the transmission of disease from contagious individuals to susceptible individuals, while “diffusion” describes the random movement of those individuals throughout the population.

2.4 Epidemiological Metrics

Numerous metrics exist for characterizing the virulence of a disease. In this thesis, we make use of two basic concepts.

2.4.1 Basic Reproduction Number

The *basic reproduction number* R_0 is the number of individuals we can expect a single contagious host to infect over the duration of their infectious period, assuming that every other individual is susceptible. In order for the disease to spread, R_0 must be strictly greater than 1. The specific number will vary from disease to disease, and can be estimated from empirical data from past outbreaks. A higher basic reproduction number implies a more virulent disease. Since we know how many unique contacts the average individual makes in one day in our dataset, and we know how many days such an individual remains infectious, the basic reproduction number is all we need to estimate the probability of transmission per contact.

2.4.2 Epidemic and Herd Immunity Thresholds

By definition, if an individual is immune to a disease, they will not contract the disease. Thus, generally speaking, if an individual is immune to a disease, they are incapable of spreading it. Therefore, as the fraction of the population that is immunized rises, the spreading of the disease is impeded, as there are fewer contacts by which it can be transmitted. If enough of the population is immune, an infectious individual may not have enough opportunities to spread the disease to the entire susceptible population, a phenomenon known as “herd immunity”.

The *epidemic threshold* β_c is the fraction of individuals that must be susceptible in order for a disease to persist. Below this threshold, the disease can be expected to die out before spreading beyond its local susceptible population. Alternatively, $(1 - \beta_c)$ is the *herd immunity threshold*, the fraction of individuals that must be immunized in order to prevent a widespread outbreak. In simplistic models, it is possible to predict these thresholds using the basic reproduction number, R_0 . These

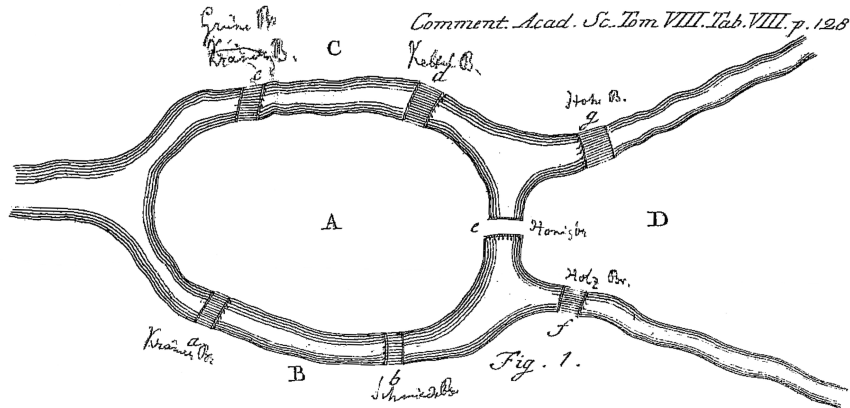


Figure 3: Drawing of The Seven Bridges of Königsberg. Euler’s illustration of the Seven Bridges labels the landmasses as A , B , C , and D . (Euler, 1741)

predictions, however, assume a full mixing model, and are therefore not applicable to our simulations.

2.5 Graphs

Perhaps by coincidence, at almost the same time Daniel Bernoulli was modeling diseases with compartments, Leonhard Euler was considering the bridges that connected an island in the Pregel River with then-Königsberg, Prussia. To solve the problem that has become known as the Seven Bridges of Königsberg, Euler developed what he called the “geometry of position” (Euler, 1741).

A *graph* $G = (V, E)$ comprises a set of *vertices*, V , and a set of *edges*, E . By convention, the number of vertices is denoted $|V| = n$, and that of edges, $|E| = m$. Two vertices u and v may be connected by an edge $e = (u, v)$. One important and fundamental characteristic of a vertex is the number of edges to which it is connected, the *degree* of that vertex. In some situations, an edge may connect a vertex back to itself, and there is no requirement that any two vertices must be connected. If there is a sequence of edges that leads from a vertex u to a vertex v , we say that there is

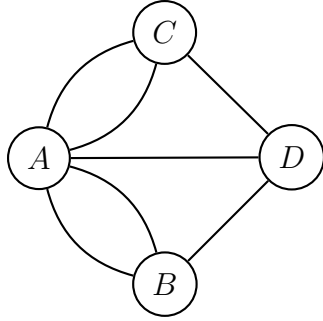


Figure 4: Graph of the Seven Bridges of Königsberg. The landmasses and bridges in Figure 3 are represented by this graph.

a *path* between u and v and that they are in the same *connected component*. Graphs allow the modeling of any arbitrary collection of entities and their relationships. In the case of the Seven Bridges of Königsberg, the vertices represent landmasses and the edges represent the bridges that connect them, as drawn in Figure 3 and Figure 4.

In representing the Seven Bridges as a graph, Euler abstracted away geometrical and geographical details of the problem that had no impact on its resolution, leaving behind only the underlying structure of edges and vertices. There are, in graph theory, certain special structures; in this thesis, we will encounter two: a *clique* is a subgraph in which every vertex is connected by an edge to every other vertex, and a *star* is a subgraph in which one vertex is connected to every other vertex. Examples of both can be found in Figure 5.

One potential application of graphs is to represent social interactions: vertices will represent people, and edges will represent some form of interaction, such as making a phone call, shaking hands, or eating a meal together. Such a graph would be called a *social network*. If we have such a network for our population, we can simulate the spread of a disease without resorting to full mixing: by spreading the disease exclusively across edges, we can simulate the individual transmission events from infectious individuals to susceptible individuals. An example of how we might augment the compartmental model of Figure 2 with a social network is shown in Figure 6.

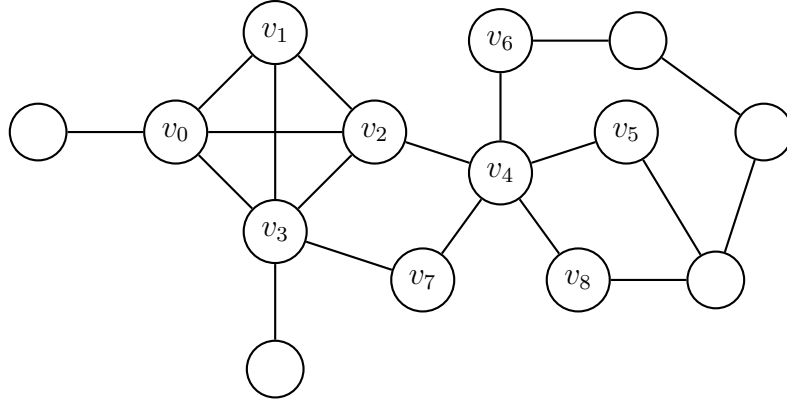


Figure 5: Cliques and Stars. Here (among others), $\{v_0, v_1, v_2, v_3\}$ form a clique and $\{v_2, v_4, v_5, v_6, v_7, v_8\}$ form a star.

2.6 The Copenhagen Network Study

The Copenhagen Network Study was conducted in two distinct iterations between 2012 and 2013, with the aim of creating a “high resolution” social network that was not owned by a private company or a government and could therefore be provided to researchers (Stopczynski et al., 2014). Rather than simple binary social interactions, the Copenhagen dataset includes the locations, face-to-face contacts, phone calls, text messages, Facebook interactions, genders, personality traits, and self-assessed mental health of about a thousand undergraduate students at the Technical University of Denmark (located about twenty minutes’ drive north of central Copenhagen).

Not all of this data is relevant to disease spreading. The details of our preprocess-

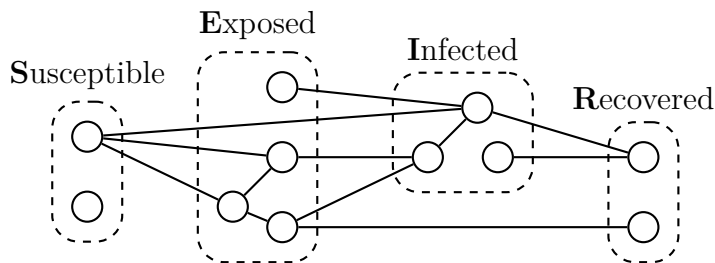


Figure 6: The SEIR Model Applied to a Social Network. The example population and compartments from Figure 2 could be augmented with a social network.

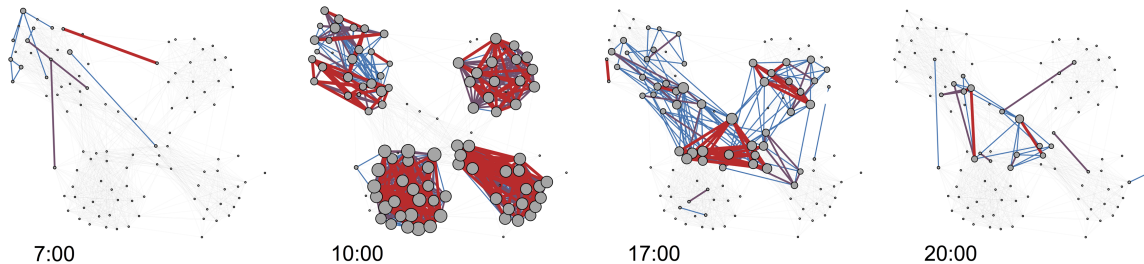


Figure 7: Selected Interactions in the Copenhagen Network Study. Blue edges indicate less frequent interactions, red, more frequent, and vertex size scales with degree (Stopczynski et al., 2014).

ing and subsequent graph representation of the dataset are described in Chapter 4. Specifically, we focus only on the face-to-face contact data. We were provided the interactions of about seven hundred individuals, over one month, reported every five minutes. A selection of those interactions is illustrated by Figure 7. The detail of this contact data gives us the opportunity to simulate transmission of disease with far more granularity, precision, and realism than would be possible in a full mixing model.

We highlight the number of interactions over time in the Copenhagen dataset, plotted in Figure 8. We find that not only do individuals make plenty of contact with each other (peaking at almost 1800 interactions within the span of five minutes on the third Monday), but their interactions also occur with a distinctive, immediately recognizable pattern. Every day, the most interactions occur around midday, and there are significantly more on what we infer to be weekdays compared to weekends.

Additionally, there is a wide variety to the total number of unique contacts each individual makes over the entire month, as shown in Figure 9. For instance, there are ten individuals who never come into face-to-face contact with more than one other, and there is one who interacts with more than five hundred. The average individual makes 229.86 unique contacts (sample standard deviation: 110.32, median: 236.50).

These characteristics of the Copenhagen dataset further reinforce the idea that full mixing simulations do not suffice: when modeling an infection, it is not a realistic assumption that every individual makes contact with every other. Perhaps we could justify approximating Figure 9 with a normal distribution, but we certainly cannot assume that these contacts are made at all times, neither can we assume that the individuals involved move randomly throughout the population.

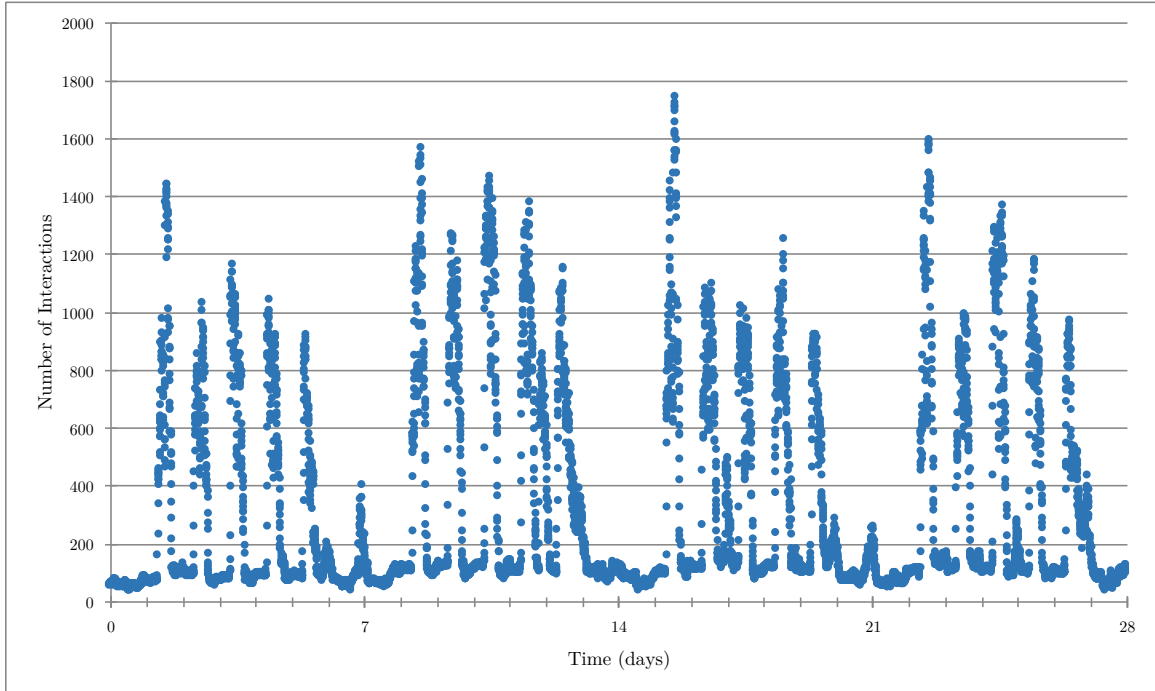


Figure 8: Interactions by Time in the Copenhagen Dataset. Preprocessed as described in Chapter 4.

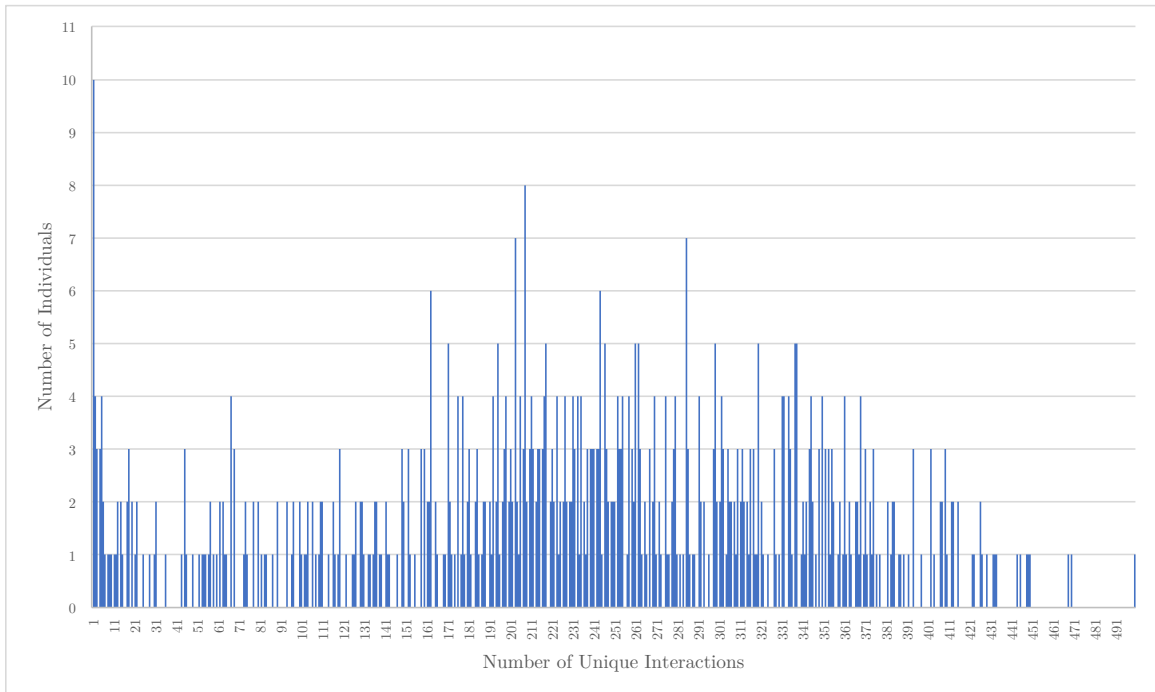


Figure 9: Numbers of Unique Interactions in the Copenhagen Dataset. Preprocessed as described in Chapter 4.

Chapter 3

RELATED WORKS

As this thesis is concerned with epidemic spreading in social networks that incorporate a richer level of detail, its related works fall broadly into two categories: those that explore epidemics on graphs and those that develop more detailed epidemic models.

3.1 Disease Spreading on Graphs

Although they employed neither a sophisticated epidemiological model nor a particularly detailed network, Kephart and White were, in 1991, among the first to propose modeling the spread of a virus using a graph (Kephart & White, 1991). Working to adapt existing mathematical models of infections, they argue that the traditional methods are not quite precise enough when dealing with computer viruses. Unlike their biological counterparts, which, in many cases, can be transmitted via the surrounding environment, computer viruses must generally spread across an active connection between two individual members of the population. It is unreasonable to assume that a computer will routinely interact with every other computer on the network. Compared to the assumptions made by a full mixing model when simulating a biological disease, it is even more unrealistic to assume that an infected computer is equally likely to infect any other computer on the network.

Modeling networked computers as vertices in a directed graph, Kephart and White apply an SIS model, assuming that every computer either is infected and capable of spreading the infection or is susceptible to being infected. Assigning to each edge a probability of spreading the virus, they run simulations on randomly generated

graphs. These simulations allow them to confirm that concepts in traditional epidemiology can also be applicable to network security: for instance, they show that a perfect defense against malware is not strictly necessary, provided individual infections are detected and corrected quickly enough.

Though their work is groundbreaking, Kephart and White acknowledge that it is overly simplistic and that, at the time, they lacked a sufficient basis of real-world data. They concede that the SIS model represents an extreme situation in which users take no precautions after their computers have been infected for the first time. Their simulations do not consider how the structure of the network might change over time: they do not account for the possibility that a user might disconnect their computer from the network upon discovering the infection.

3.2 Disease Spreading with Mobility Patterns

The incorporation of individual movements, such as we might find in high resolution social networks, is not exclusive to models based on graphs. Granell and Mucha argue that the traditional reaction-diffusion system for modeling the movements of individuals is inadequate when considering diseases that affect humans (Granell & Mucha, 2018). Humans are habitual; they do not diffuse uniformly or randomly.

Granell and Mucha partition the population based on individuals' locations at discrete time steps, differentiating between residences and common spaces and between daytime and nighttime behavior. Their model attempts to account for patterns of movement. This allows them to identify the optimal amount of mobility in the presence of a burgeoning infection: where uninfected individuals move around enough to minimize exposure, yet infected individuals stay stationary enough to avoid assisting the spread of the infection. Their probabilistic model is not applied to graphs; it

is verified with a variety of Monte Carlo simulations of situations not based on any particular real-world data.

Similarly, Gemmetto, Barrat, and Cattuto used temporal information to refine a conventional epidemic simulation (Gemmetto, Barrat, & Cattuto, 2014). They, however, derived their data from a real-world high resolution contact network obtained from sensors worn by children at a primary school in Lyon, France. Although their work is not explicitly posed as a graph-based problem, their model did separately and individually compute the results of every contact between children as it simulated the spread of an influenza-like disease using SEIR.

Since theirs is a real-world dataset, Gemmetto *et al.* investigate typical real-world attempts to combat the spread of infections in this situation. Their model allows them to simulate the effects of varying levels of school closure, from canceling just one class to canceling all classes in a grade to closing the entire school temporarily. They are able to show that canceling just one class (the class containing the infected child) can drastically reduce the spread of an infection, and may actually be preferable to closing the entire school — not only do the children in other classes remain productive in school, but they are shielded from any infections spreading through the general community.

One acknowledged limitation of the dataset used by Gemmetto *et al.* is its specificity and scope; it includes fewer than three hundred people over just two days. Short-term patterns must be extrapolated and long-term results, especially concerning interactions with the enclosing community, must be approximated. The generality of their conclusions has yet to be thoroughly explored.

3.3 Disease Spreading on Graphs with Mobility Patterns

Combining both mobility data and graph-based simulation, the work of Frías-Martínez, Williamson, and Frías-Martínez more closely resembles that of this thesis: they constructed and infected a temporal social network of an entire city in Mexico (Frías-Martínez, Williamson, & Frías-Martínez, 2011). Their location data was collected from cell towers, so it is not nearly as precise as the wearable sensor data employed by Gemmetto *et al.* or the data collected in the Copenhagen Network Study, but Frías-Martínez *et al.* specifically chose a time and place that had been affected by the 2009 H1N1 swine flu pandemic. Their agent-based social network model simulates infections using compartmental SEIR while considering interactions and mobility patterns on an individual basis.

By including these real-world mobility patterns in their network, Frías-Martínez *et al.* could tweak the simulated movements of their population, allowing them to retrospectively analyze the effectiveness of the Mexican government’s response to the outbreak. They were able to tentatively show that closing schools within the first week, then prohibiting all non-essential movements within the following week delayed the exponential spread of infections by forty hours and reduced the overall number of cases by 10%.

Though they had a different goal, Stopczynski, Pentland, and Lehmann also infected a high resolution social network — the Copenhagen network (Stopczynski, Pentland, & Lehmann, 2018) (Stopczynski and Lehmann being two of the researchers behind the Copenhagen Network Study). They considered two networks formed from the same dataset: one including only interactions between individuals that came within $\lesssim 1$ meter of each other, and one that also included any interactions within $\lesssim 10$ meters. Since the long-range network is naturally a superset of the short-range network,

they then randomly sampled long-range interactions to ensure that both networks contained the same number of edges.

By applying an infection with a simple SIR model to these two networks, Stopczynski *et al.* are able to emphasize their structural differences (perhaps unsurprisingly, they find that their epidemic is more difficult to contain if it is allowed to spread through long-range interactions). This model is neither based on nor intended to simulate any real-world disease; it was developed only to highlight the differences between the short- and long-range networks.

3.4 Span-Cores in Temporal Networks

Finally, while it is neither exclusively applicable to the task of simulating epidemics nor to the concept of high resolution social networks, the theoretical work of Galimberti, Barrat, Bonchi, Cattuto, and Gullo is of tangential interest to this thesis. They introduce the concept of a “span-core”, essentially, a k -core for graphs that change over the course of discrete time steps (Galimberti, Barrat, Bonchi, Cattuto, & Gullo, 2018). The Copenhagen dataset is such a network, and we employ k -cores in Section 4.4.3.

Given a graph G , a k -core is a maximal subgraph of G within which every vertex has degree at least k . The collection of all k -cores, the core decomposition of G , can provide valuable insight into finding and approximating the densest regions of a graph. In a conventional, static graph, a graph that does not change over time, any k -core is itself a subgraph of some $(k - 1)$ -core for $k \geq 1$. This fact can be exploited in the computation of the core decomposition, which can be done in linear time (Matula & Beck, 1983). However, if the graph can change over time, this assumption no longer holds.

Galimberti *et al.* introduce the span-core, terminology they use to indicate a (k, Δ) -core, where Δ is an interval of time steps. As we might intuitively expect, a (k, Δ) -core is a maximal subgraph within which every vertex is, over the lifetime of the interval Δ , connected to at least k other vertices. As a simple example of a potential application, they show how span cores can be used to characterize patterns of interaction between school children. They use span cores to differentiate between students at a primary school in Lyon (the same as studied by Gemmetto *et al.*) and those at a high school in Marseilles.

IMPLEMENTATION

4.1 Preprocessing

Four week’s worth of the Bluetooth-based face-to-face data collected by the Copenhagen Network Study (Stopczynski et al., 2014) was provided to us in CSV form, an example of which is shown in Table 1. Each Bluetooth scan was given as a quadruple of a timestamp, a “User A ID” (the scanner), a “User B ID” (the scannee), and a Received Signal Strength Indicator (RSSI). Scans are reported at discrete intervals, and timestamps are integers representing the number of seconds since the start the study. Since scans were reported every five minutes, timestamps begin at 0 and increment by 300. A User B ID of “−1” indicates that User A reported scanning no Bluetooth devices at that time, while a User B ID of “−2” indicates that User A scanned a device that was not part of the study.

While every user ID in the raw Bluetooth data is a natural number, they are not contiguous. We first rename and rearrange user IDs to effectively “defragment” the assigned user IDs, according to the order in which they first appear in the raw data. This allows user IDs to correspond directly to indices in the vertex sequences employed by the iGraph library (The iGraph Core Team, 2015). We also prune every user who never makes contact with any other user over all four weeks of the study. Presumably, these users are left over from portions of the Copenhagen Network Study to which we do not have access — we find it generally unrealistic for a student to be isolated from their peers for an entire month. These isolated vertices have no effect on the outcomes of our simulations, as they can neither contract nor spread any disease.

4.2 Graph Representation

We model the scans of each timestamp as individual undirected, weighted graphs using the iGraph library for Python (The iGraph Core Team, 2015). These individual graphs can collectively be treated as snapshots of a network that changes over time. In each graph $G_i = (V_i, E_i)$, vertices represent users; there is an edge between vertex u and vertex v of weight w if user u scanned user v or vice versa with RSSI w at timestamp t_i . For each vertex v , we further record of the number of external devices scanned by user v .

We always assume that if user u has come face-to-face with user v , then user v must have come face-to-face with user u . In some rare situations, the raw data contains both a scan of user u by user v and a scan by user v by user u at time t_i . In these cases, we add only one edge between users u and v , weighting that edge with the higher (closer to zero, indicating a stronger signal) RSSI of the two scans.

Finally, we observe that not every user is present in a scan in every timestamp. This could simply be because their phone had no power at that time. For the purposes of our simulation, we add vertices to the network once their corresponding users appear in a scan, but we never remove them. Users that have appeared before yet fail to appear in any scan in a later timestamp will be persisted as isolated vertices — since we are interested in simulating the spread of a disease, we need to keep track of users' states with respect to the infection at all times, regardless of whether or not they successfully report their scanning status. One snapshot that results from parsing the data in Table 1 is shown in Figure 10.

While we chose to represent the RSSI values of individual scans as the weights of edges in our network, in practice, none of our simulations will use this information.

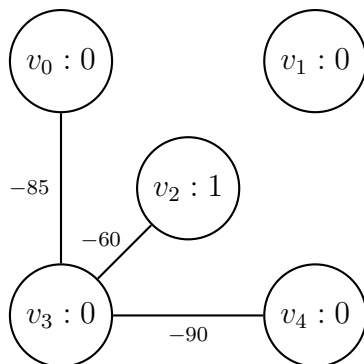


Figure 10: Graph Snapshot of One Timestamp. The snapshot produced by data at time $t_1 = 300$ in Table 1. Note that v_1 is presumed to persist from time $t_0 = 0$, although it is missing from the raw Bluetooth data. Note also that v_2 is unique in that it has scanned an external device at this time.

Recovering physical characteristics of interactions (distance between users, for example) would require that we have specific details of the scanning devices. Since we do not have these details, we assume that face-to-face interaction is binary. The same is true of the number of external devices scanned: while we had initially hoped that these scans could be interpreted as contacts with individuals who did not participate in the study, there is not necessarily any correlation between external devices scanned and number of face-to-face interactions. “External devices” includes wireless mice, for example, not just other phones (Sapieżyński, n.d.). So, although this information is represented in our graph as attributes of each vertex, none of our simulations make use of it.

Thus, the only meaningful feature of individual interactions is whether or not they occurred at all; the only meaningful characteristic of the network is the structure of each snapshot. It is important to note that the vertices and their edges need not exclusively form little cliques within the network: if user a scans user b , and user b scans user c , that does not imply that user a must have scanned user c . This could mean that, within the five minute interval represented by each timestamp, user b came into contact with users a and c separately. It could also mean that the three

users formed a line when they came into contact with each other, such that user b was close enough to scan both users a and c , but those users were not close enough to scan each other.

4.3 Simulation Architecture

We developed a modular system that allows the definition of compartmental models in Python. Each model is defined within a Python module and implements, at minimum, three functions: one to initialize new vertices when they are first added to the network, one to log information about the current state of the model, and one that simulates the spread of disease for a single timestep.

This architecture allows us easily develop and swap out variations of models for different diseases and different amounts of simulation complexity. For any time t_i , the individual model modules are given full access to the iGraph vertices of both the G_i and G_{i+1} snapshots, so they may adjust the attributes contained within as necessary — the times at which individuals enter each compartment, for instance, are stored as attributes of their corresponding vertices.

4.3.1 The Flu Model

We first validated our simulation architecture by developing an intentionally simplistic compartmental model for influenza. Influenza, commonly known as the “flu”, is an infection caused by viruses belonging to one of four families: Influenzavirus A, Influenzavirus B, Influenzavirus C, and Influenzavirus D (Krammer et al., 2018). Of these, Influenzavirus A is perhaps the most widely known; in particular, its subtype H1N1 was responsible for the 1918 Spanish flu and 2009 swine flu pandemics. The

1918 outbreak is estimated to have killed 50 million people worldwide (Johnson & Mueller, 2002), easily more than twice the total casualties of World War I, which ended that same year (Gilbert, 1994). The flu is typically transmitted by airborne particles released by a nearby infected individual’s coughing or sneezing.

While flu vaccines exist, the rapid mutation of the flu families of viruses prevents humans from developing any long-term immunity; roughly every year, they must reacquire (whether by vaccination or by recovery from infection) immunity to new strains of the virus (Krammer et al., 2018). Since our dataset only covers a month of interactions, our simulations do not run long enough for this lack of long-term immunity to have any effect. Therefore, we chose to simulate the flu with a simple SIR model: unless otherwise specified, individuals begin the simulation susceptible to infection. If infected, they immediately become infected and infectious for between 5 and 8 days, after which they are considered to be completely recovered.

In the case of the 1918 Spanish flu, the basic reproduction number has been estimated as $R_0 \in [2, 3]$ (Mills, Robins, & Lipsitch, 2004). On the first Monday of the Copenhagen Network study, individuals made contact with an average of 28.18 other participants. We thus estimate the transmission probability per contact as $\beta = \frac{2.5}{28.18 \times 6} = 0.015$. This is a crude estimate at best. Notably, an individual in the Copenhagen dataset does not necessarily make equal or singular amounts of contact with those 28.18 others. We reiterate that this simulation is only intended to validate our software architecture and establish that epidemic outbreaks are possible in the network. The interventions which we will apply in Section 4.4 are not based on such rough estimates.

After each discrete timestep — recall that our contact data was reported in five minute intervals — we consider the previous state of each individual separately. If an individual was previously in the susceptible state, they remain susceptible in the

next by default, though they may be infected later in the simulation of this timestep. Whenever an individual is infected, they are moved to the infected state in the next snapshot and assigned a random infectious period between 5 and 8 days. If an individual was previously in the infected state, and their infectious time has expired in the intervening five minutes, then they are moved to the recovered state. If, however, an infected individual is scheduled to remain infectious, we explore all of its neighbors in the previous snapshot, infecting any susceptible neighbors with probability 0.015. Finally, recovered individuals remain in the recovered state; as previously noted, our simulation would not run long enough for them to realistically move back into the susceptible state.

All of these model parameters are uniformly applied to every individual in the population, an assumption that is also true of our forthcoming measles and norovirus models. For example, we do not attempt to simulate other pre-existing conditions that could exacerbate a flu infection. We do not have enough demographic information to simulate otherwise. That being said, real-world populations similar to that of the Copenhagen Network Study should, in fact, be composed primarily of healthy individuals around the same age.

4.3.2 The Measles Model

Having validated the architecture of our simulation through the development of an influenza model, the main focus of this thesis concerns measles and interventions for measles-like diseases. A highly contagious disease which likely evolved from the same viral ancestor as the recently eradicated rinderpest, measles was once responsible for more than 2 million deaths annually (Rota et al., 2016). Nine out of ten susceptible individuals who come into contact with an infectious individual will contract the disease (CDC (Centers for Disease Control and Prevention), n.d.-e), which can continue

to spread effectively even when less than 10% of the population is at risk — in other words, its herd immunity threshold is at least 90% (Rota et al., 2016).

Once it infects its host, the measles virus incubates for 10–12 days (CDC (Centers for Disease Control and Prevention), n.d.-e). Prodrome, an initial period of milder symptoms, then lasts 2–4 days, followed by a rash that lasts 5–6 days, and the host will be contagious for up to 4 days before and after the rash appears. All told, an infectious individual will spend up to 14 days infecting up to 90% of those they meet. Thus, it is not surprising that measles has been targeted by the World Health Organization for elimination, though their 2015 global target for reduction in cases was not met (Rota et al., 2016).

We model measles using an SEIR compartmental model. Individuals begin in the susceptible state. Once infected, according to the timeline enumerated above, they are in the exposed state for between 10 and 12 days, then move to the infected state for 7 to 14 days. Individuals are only capable of transmitting the disease while in the infected state, which they do so at a rate per contact of $\beta = 0.90$ (CDC (Centers for Disease Control and Prevention), n.d.-e). As with influenza, those infected with measles eventually move to the recovered state, after which they are no longer at risk of being infected. While measles carries a non-zero chance of death or other complications (Rota et al., 2016), we chose not to simulate any of these alternative outcomes.

One factor that makes measles so contagious is its ability to linger in airborne droplets for hours (Rota et al., 2016). It is possible for a susceptible individual to contract the disease without ever coming into contact with an infected individual, simply by their occupying the same room within a few hours of each other. Because of the distinct lack of location data in the Copenhagen network dataset, we chose not to simulate this method of infection. As we will show in Section 6.1.2, the disease is more than

capable of spreading rapidly by face-to-face transmission alone. We also note that this possibility would not be simulatable with traditional full-mixing models, either.

4.3.3 The Norovirus Model

To demonstrate the flexibility of our architecture beyond simple compartments and face-to-face contact events, we implemented a model to simulate the spread of norovirus, the most common foodborne disease in the United States (CDC (Centers for Disease Control and Prevention), n.d.-d). Because norovirus is easily spread by contaminated food, by vomiting, and by toilet flushing, we believe that some amount of location-based information is essential to a realistic simulation model. In lieu of any such empirical data provided to us from the Copenhagen Network Study, we implemented a compartmental model that also conservatively identifies clusters of individuals in close contact with one another.

Ideally, we would be able to take the entire month’s worth of snapshots, compute the connected components of each, and attempt to track patterns of how vertices enter and exit components. This would allow us to estimate exactly when a cluster formed, and it could allow us to smooth over any potential hiccups in the reported data — if an individual remained in a cluster consistently but lost phone power for ten minutes during that time, for instance.

Unfortunately, such a computation would be prohibitively expensive, and the extraction of locations and mobility patterns from anonymized contact data is beyond the scope of this thesis. Rather, our norovirus model only considers snapshots within predetermined intervals. Vertices are considered to be within a cluster if and only if they are within the same connected component for every snapshot in the interval. Once the last snapshot is processed and clusters have been determined, we can simu-

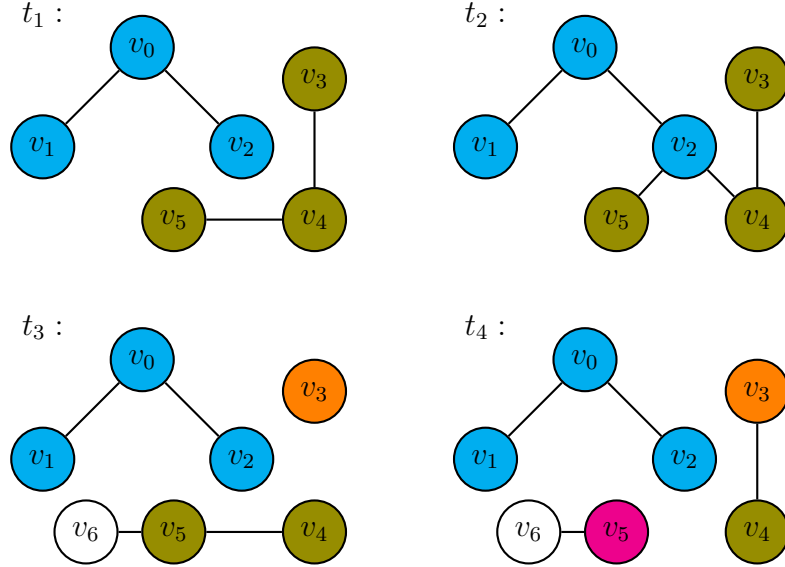


Figure 11: Cluster Detection over Four Snapshots. Note that v_6 is never considered, as it cannot be in any non-trivial cluster.

late the spread of norovirus among the individuals of a cluster with some alternative probability of transmission. Assuming our clustering interval is relatively short, on the order of hours, any individuals newly infected by this method would not have time to themselves spread the infection, so we do not need to retroactively explore their interactions with individuals outside the cluster. Importantly, this allows our model to identify clusters in a streaming fashion. At any given timestep, it need only consider the next snapshot and the existing clusters that it has computed up through the previous snapshot: if a vertex v belonged to cluster C at time t_i , and v belongs to component A at time t_{i+1} , then v belongs to cluster $C \cap A$ at time t_{i+1} .

For instance, consider the contact data over four snapshots visualized in Figure 11. At time t_1 , no previous snapshots have been considered, so the computed clusters are equal to the snapshot's connected components: v_0, v_1 , and v_2 form one cluster, while v_3, v_4 , and v_5 form another. At time t_2 , the entire snapshot is connected, thus, when each vertex's existing cluster is intersected with its current component, none of the clusters are changed. At time t_3 , v_3 appears in a different component from v_4 and

v_5 , so their cluster is split in two. Note that the choice to color v_3 's cluster orange while v_4 and v_5 's remains olive green is entirely arbitrary, and there is no significance within our model to which vertices “remain” in an existing cluster and which are “split off”. Finally, at time t_4 , the lack of an edge between v_5 and v_4 splits the former off into an fourth cluster. Note that although an edge between v_3 and v_4 appears at time t_4 , those vertices are not placed in the same cluster; they have been irrevocably separated into different clusters at time t_3 . Also note that v_6 is not placed into any cluster at all, although it always appears in the same connected component as v_5 . Since it does not appear in the initial snapshot, it cannot possibly appear in the same connected component as any other vertex in all snapshots. It cannot appear in any meaningful cluster, so, for the purposes of our model, it can be ignored.

Similar to influenza, immunity to norovirus is not permanent, but lasts long enough that individuals cannot be infected more than once within the time span of our simulations. We therefore simulate norovirus using an SEIR compartmental model, just as we did measles. Recent work based on reported outbreaks in Japan has estimated the basic reproduction number of norovirus to be $R_0 \in [1.11, 4.15]$ (Matsuyama, Miura, & Nishiura, 2017). Adopting the same crude estimation as we did for influenza, this gives our model a transmission probability per contact of $\beta = \frac{2.63}{28.18 \times 2} = 0.047$. While it may be possible for individuals to continue spreading the disease at some lower rate once recovered (CDC (Centers for Disease Control and Prevention), n.d.-d), we did not model this possibility. Individuals remain in the exposed state for 0.5 to 2 days, and in the infected state for 1 to 3 days.

That basic number only applies to airborne transmission. As previously noted, norovirus is spread especially well in shared restrooms and by contaminated food. To account for these possibilities, our model first identifies clusters of individuals between 12:00 midnight and 4:00am, between 11:00am and 12:00 noon, between 12:00

noon and 1:00pm, and between 1:00pm and 2:00pm. We assume that these individuals either share a dormitory or a meal with each other. Within each of these groups, every infected individual has the opportunity to spread the infection to everyone else in the cluster. This differs from the normal simulated mode of transmission in two ways: firstly, this potentially gives infected individuals far more influence over those nearby. Consider, for example, the clusters given by the four snapshots in Figure 11. Suppose that, in this situation, individual v_1 was infected. Since v_1 appears in the same cluster as v_0 and v_2 , it would have a chance to infect both of those individuals. Notice, however, that v_1 never makes direct face-to-face contact with v_2 . In the standard simulation, the infection would not be able to spread to v_2 within the time span of these four snapshots. In the standard simulation, v_1 could potentially infect v_0 , but v_0 would not become contagious quickly enough to then infect v_2 .

Additionally, we can apply an alternative probability of transmission to the contact events assumed to have taken place within a cluster. In a food service scenario, a single infectious employee who continued to work while symptomatic but took reasonable precautions (wearing gloves and frequently washing or disinfecting hands) was estimated to infect 167.4 customers over 2000 servings, where 80% of the population was susceptible (Duret et al., 2017). Assuming that no customer was served twice, this gives an increased transmission probability per cluster contact of

$$\beta_{cluster} = \frac{167.4}{0.8 \times 2000} = 0.1046.$$

4.4 Interventions

Having experimented with immunizing volunteers against smallpox by injecting them with material from cowpox lesions, Edward Jenner published the procedure he termed “vaccination” in 1798 (Riedel, 2005). The World Health Organization began an in-

tensified effort to contain the disease in 1966, the last naturally occurring case was reported in Somalia in 1977, and smallpox was declared eradicated in 1980 (Fenner et al., 1988). The primary plan was a mass vaccination campaign with a goal of immunizing 80% of the global population. Additionally, the WHO established an extensive surveillance program to preempt outbreaks: whenever a case was reported, the infected person would be isolated and their household and close contacts would be vaccinated, a strategy termed “ring vaccination”. This sort of targeted intervention could not be simulated in a full mixing model, which does not distinguish between the individual members of each compartment and their interactions.

Compared to traditional full mixing, our graph-based models give us far more control over the granularity of the simulation. They allow us to explore the transmission of a disease on an individual, per-contact basis. They also allow us to implement more sophisticated simulations based on the structure of the network, as we did with norovirus. And, as a result of their more fine-grained simulation, graph-based models also enable additional opportunities to explore interventions: strategies for preventing or containing an epidemic.

Using our measles model, we tested five major intervention strategies: four for prevention, based on vaccination, and one for containment, through quarantine. Further variations on these strategies ultimately result in twelve simulated interventions. We chose to test interventions using the measles model for three reasons: firstly, measles is highly contagious, so interventions should have a more pronounced effect in a shorter amount of time — if an individual escapes infection, that is more likely to be a result of the intervention than a result of random transmission chance. Secondly, while the measles virus is highly contagious, the measles vaccine is equally highly effective. Properly administered in two doses, the vaccine is 97% effective (CDC (Centers for Disease Control and Prevention), n.d.-c), so we avoid more complex scenarios where

we must account for ineffective vaccinations. Finally, at the time of this writing, measles outbreaks in the United States have recently risen drastically (CDC (Centers for Disease Control and Prevention), n.d.-b), so interventions that target measles are of greater immediate relevance. A summary of interventions is given in Table 2.

4.4.1 Uniform Vaccination

Our baseline intervention is vaccination with uniform probability. Whenever an individual first appears in Copenhagen Network Study — whenever a new vertex is added to the network — that individual has a chance of gaining complete immunity to the disease. Effectively, individuals start in the recovered state instead of in the susceptible state with some probability p . We simulated varying amounts of vaccination, from $p = 0.05$ to $p = 0.95$ in increments of 0.05.

4.4.2 Vaccination by Degree Distribution

Considering our lack of demographic information, vaccination with uniform probability is likely the most realistic preventative intervention we can employ. However, we imagine a scenario where there are a limited number of vaccines. Perhaps the vaccine is costly to manufacture, difficult to administer, or experimental in nature. Perhaps we are not dealing with measles proper, but with a new disease that spreads similarly to measles. In these situations, rather than randomly selecting individuals to vaccinate, we would like to vaccinate those that have the greatest effect on the spread of the infection.

Once an individual is vaccinated, they can neither contract nor spread the disease. In our graph-based models, the disease is spread along edges, thus, vaccinating an individual is analogous to removing all of the edges incident to their corresponding

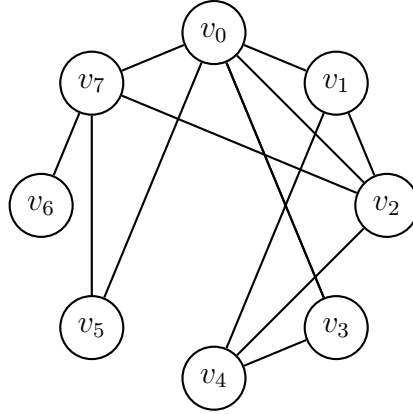


Figure 12: Example Degree Distribution. This graph has degree distribution $(0, 1, 2, 2, 2, 2, 1)$

vertex. Alternatively, vaccinating an individual effectively removes their vertex from the network: the individual is simply no longer capable of affecting the spread of the infection. If we are to contain an infection by removing a limited number of vertices, and the infection spreads along edges, then our ultimate goal is to remove vertices in order to disconnect the network into multiple connected components. If we can split the network into multiple connected components, any infection that begins in one component will not be able to spread to the others. Given this goal, if we have a limited number of vaccinations, one possible approach is to vaccinate those individuals who make the most contact with other individuals.

The degree distribution of a graph is simply the numbers of vertices having each of the possible degrees. For example, the graph in Figure 12 has degree distribution $(0, 1, 2, 2, 2, 2, 1)$, because there are zero vertices with degree 0, one vertex with degree 1, two vertices with degree 2, two vertices with degree 3, two vertices with degree 4, and one vertex with degree 5. Alternatively, we could express the degree distribution as a sequence of bins of vertices: $(\emptyset, \{v_6\}, \{v_3, v_5\}, \{v_1, v_4\}, \{v_2, v_7\}, \{v_0\})$.

Supposing that higher degree vertices would contribute to a faster or wider spread of the infection were they to be infected, and supposing that those vertices are re-

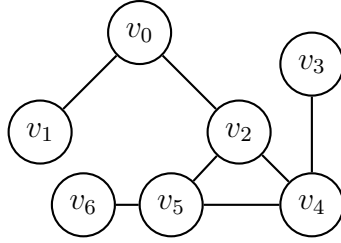


Figure 13: Example Flattened Graph. Flattening the four snapshots in Figure 11 produces this graph.

responsible for connecting more components together, we simulated the vaccination of limited numbers of individuals, prioritizing those of highest degree. Effectively, this is simply sorting the vertices in ascending order by degree and vaccinating the last λ corresponding individuals. For some values of λ , this will be close if not identical to vaccinating the individuals in the k highest bins of the degree distribution.

The network representing the Copenhagen Network Study, however, is not a single graph; it is composed of numerous graph snapshots. Before we can compute the degree distribution, we must decide how many snapshots to “flatten” into a single graph. Given a sequence of snapshots, the flattened graph contains an edge between vertices u and v if and only if there is an edge between u and v in at least one of the snapshots. The result of flattening the four snapshots in Figure 11 is shown in Figure 13.

One option is to simply flatten the entire network, combining the entire month’s worth of snapshots into a single graph. We can think of this as providing those in charge of administering vaccinations with advance knowledge of every individual’s movements for the next month. While this would give us the most information about the degrees of vertices, it is likely unrealistic. Therefore, we tested two additional variations on this intervention strategy: first, using the network flattened over the first week of the study, second, over just the first Monday.

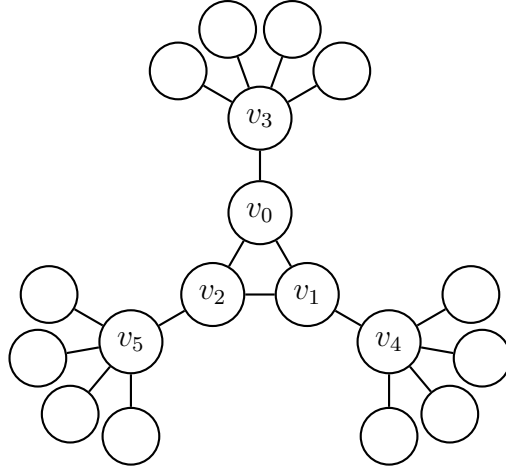


Figure 14: Example k -Cores. This graph’s highest k -core does not include its highest degree vertices.

In all three cases, we simulated varying numbers of vaccinations in increments of 10. There are, for instance, 659 unique individuals seen in the first week of the Copenhagen Network Study. We therefore simulated administering λ vaccinations, where $\lambda \in \{10, 20, 30, \dots, 650\}$, and the vaccinations were given to the λ vertices of highest degree based on the flattened graph of the first week.

4.4.3 Vaccination by k -Cores

With the same reasoning that motivated vaccination by degree, we next investigated vaccination informed by k -cores. A k -core is a maximal subgraph such that every vertex has degree at least k . A graph’s cores thus form a hierarchy of vertices, as any vertex that appears in a core must also appear in any cores for lower values of k . Note that a vertex with degree k is not necessarily contained in — and, moreover, does not imply the existence of — the graph’s k -core. Additionally, the highest degree vertices are not always found in the highest core. In Figure 14, for example, the highest degree vertices are v_3 , v_4 , and v_5 , all of which have degree 5. There is, however, no 5-core. The highest core is the 2-core, which contains vertices v_0 , v_1 , and v_2 .

Vertices belonging to higher k -cores roughly correspond to vertices in denser regions of the network, thus, we consider the possibility that these individuals have a greater effect on the spread of an infection. As we did with the degree distribution, we first considered three methods of flattening the network: flattening the entire month’s snapshots, flattening the first week, and flattening the first Monday. For each flattened graph, we computed every k -core. Given each set of k -cores, we then simulated limited numbers of vaccinations, and vaccinations were given to the individuals that appeared in the highest cores.

4.4.4 Vaccination by Density Decomposition

Similar to k -cores, the density decomposition partitions the vertices of a graph based on the density of their subgraphs, producing “rings” of vertices (Borradaile, Migler, & Wilfong, 2018). Compared to the highest core, the highest ring better approximates the densest subgraph. It can be shown that if the highest ring in the density decomposition is R_k , then the density of the densest subgraph is in the range $(k - 1, k]$, where density is the ratio of edges to vertices. Furthermore, every vertex in the densest subgraph must be in the highest ring.

The downside to the density decomposition is its computational complexity. Finding the density decomposition involves first finding an “egalitarian orientation”. Finding that orientation has a running time of $O(|E|^2)$, which is the lower bound on the running time of the decomposition. The k -cores of a graph, in contrast, can easily be found all at once with a simple, greedy, pruning algorithm in linear, $O(|V| + |E|)$ time (Matula & Beck, 1983).

As we did with k -cores, we suppose that the vertices in the highest rings, being in the densest regions of the network, represent individuals who are best positioned

to propagate the infection. The density decomposition was separately computed for the network flattened over the first Monday, over the first week, and over the entire month. For each decomposition, we simulated vaccinating a limited number of individuals drawn from the highest rings.

4.4.5 Quarantining

The interventions we have considered thus far all revolve around vaccination. They all rely on the concept of inducing immunity to a disease before coming into contact with an infectious individual. If an individual has already been infected, an alternative method of preventing an epidemic is to quarantine that individual: to isolate them so that they cannot spread the infection to others. Effectively, this involves augmenting our SEIR measles model to produce an SEIQR model, where individuals enter a quarantined state between the infected and recovered states. Such a compartmental model has previously been used to explore the impact of isolation on seasonal patterns of infections (Feng & Thieme, 1995).

In the case of measles, during the prodrome period, individuals are already contagious, but their externally perceivable symptoms are largely cold-like: a fever, coughing, a runny nose, and pink eye (CDC (Centers for Disease Control and Prevention), n.d.-e). Only after a rash appears is it obvious that measles is the cause. Therefore, we keep individuals in the infected state for 2–4 days before moving them into the quarantined state. During that time, they are still capable of spreading the infection. We assume a perfect quarantine with which those affected are completely compliant: there is no chance that individuals will ignore the quarantine and remain in the infected state, and, once quarantined, there is no chance that an individual could inadvertently transmit the infection. We also make no attempt to proactively quarantine others who might have recently made contact with a newly quarantined individual.

4.5 Summary of Simulations

We developed baseline models to simulate three different diseases on the Copenhagen network: an influenza model, to verify that an epidemic simulation was possible, a measles model, to serve as a basis for exploring interventions, and a norovirus model, to demonstrate that our models are capable of handling more complex transmission events than simple face-to-face contacts. Using our measles model, we then developed twelve variations of intervention strategies: ten preventative and two responsive.

timestamp	user_a	user_b	rss_i
0	0	1	-90
0	2	-1	0
300	0	3	-85
300	2	3	-80
300	2	-2	-70
300	3	2	-60
300	4	3	-90

Table 1: Example Format of Bluetooth Data.

Intervention	Probability of Intervention	Number of Interventions	Requisite Foreknowledge
Uniform Vaccination	$p \in \{0.05, 0.10, \dots, 0.95\}$	$\approx p \times 692$	<i>none</i>
Vaccination by Degree	1	$\lambda \in \{10, 20, 30, \dots, 690\}$	Degree Distribution, one month Degree Distribution, one week Degree Distribution, one Monday
Vaccination by Cores	1	$\lambda \in \{10, 20, 30, \dots, 690\}$	k -Cores, one month k -Cores, one week k -Cores, one Monday
Vaccination by Rings	1	$\lambda \in \{10, 20, 30, \dots, 690\}$	Density Decomposition, one month Density Decomposition, one week Density Decomposition, one Monday
Quarantine	1	<i>equal to number of infections</i>	<i>none</i>

Table 2: Summary of Simulated Interventions

Chapter 5

METHODOLOGY

In all simulations, one individual was randomly chosen from the first snapshot to be the “seed” (not to be confused with the seed of the random number generator) of the infection. Though it may not be realistic, for the purposes of our simulations, participants of the Copenhagen network study are assumed to make negligible contact with the outside world, and we do not simulate any chance of infection via such external contacts. In order to allow ample time for infections to run their course, the dataset, preprocessed as described in Section 4.1, was looped twice, creating eight weeks of individual contact events. A summary of scenarios simulated is shown in Table 3.

5.1 Baseline Models

Each model — SIR influenza, SEIR measles, and SEIR norovirus — was first simulated without any interventions. For each baseline model, 30 seeds (in both senses of the word: we tested different seeds of infection; to keep them consistent across trials, we set the seed of the random number generator) were tried, and for each seed, 5 simulations were run. Thus, in total, 150 simulations were run for each model, 450 overall.

At each timestep of each simulation, we logged the number of individuals in each compartment as well as the total number of affected individuals. Throughout this thesis, we refer to an *affected individual* as one that has contracted the disease at some point in the past: one who is or has been in the infected state, in the case of

the influenza model, or one who is or has been in the exposed or infected states, in the cases of the measles and norovirus models.

5.2 Models with Interventions

As noted in Section 4.4, all of our interventions were tested based on the measles model. For uniform vaccination, we simulated varying values of the vaccination probability p from $p = 0.05$ to $p = 0.95$, in increments of 0.05. For each value of p , we tested 10 different seeds for the initializing random number generator, and for each seed, we ran 5 simulations — while the random number generator is seeded when determining which individuals are initially vaccinated or infected, it is not when determining transmission of infection. In total, 950 simulations were run for the uniform vaccination scenario.

For the degree-, core-, and ring-based vaccination scenarios, we vaccinated λ individuals whose vertices were among those with the highest degree, core, or ring. The values of λ were varied in increments of 10 up to the total number of vertices in the corresponding flattened graphs. In the event of a tie between vertices, vaccinations were distributed randomly. For example, suppose that there were 20 vertices in the highest ring, ring R_k , and 30 vertices in R_{k-1} . If $n = 30$, we would vaccinate the 20 vertices in ring R_k along with 10 vertices randomly selected from R_{k-1} . If we then incremented to $n = 40$, we would vaccinate the same 20 vertices in ring R_k , and we would make a new random selection of 20 vertices from R_{k-1} .

As noted in Section 4.2, not every individual appears in every snapshot. There are numerous individuals who do not make their first appearance until after the first snapshot. In all of the vaccination scenarios, newly introduced individuals may be vaccinated according to the parameters of the intervention, regardless of when they

first appeared in the dataset. That is to say, in the uniform vaccination simulations, every individual is vaccinated with some probability p once they make their first appearance, even if that is not in the first snapshot. Likewise, in the degree-, core-, and ring-based simulations, selected individuals were vaccinated once they made their first appearance, even if that was not in the first snapshot. Additionally, because some individuals first appear in later snapshots, it is possible that the λ vertices of highest degree, core, or ring encompass all of the vertices in the first snapshot. In this situation — which is also theoretically possible with uniform vaccination, though highly improbable — it is not possible to seed the infection during the first timestep, so we continue attempting to infect individuals as they make their first appearances until we find a valid seed.

Specifically, there are 612 individuals who appear during the first Monday, 659 during the first week, and 692 over the entire month. For each value of λ , we ran 5 simulations. Thus, the 61 varying values of λ for the first Monday, 65 for the first week, and 69 for the first month collectively mandate 2925 simulations for the degree-, core-, and ring-based scenarios.

Finally, in the quarantine scenario, we simulated two approximations of real-world situations. Firstly, we tested quarantine with no other interventions. Secondly, we tested quarantine with a uniform vaccination probability of $p = 0.90$. Of the probabilities we had already previously tested, this best matches the percentage of children who received the measles vaccination in the United States as of 2017 (CDC (Centers for Disease Control and Prevention), n.d.-a). In both cases, as with the baseline simulations, 30 different infection seeds were tried, with 5 simulations per seed. We therefore ran 300 simulations of quarantine scenarios.

Scenario	Varying Parameters	Total Simulations
Baseline Influenza	30 seeds	150
Baseline Measles	30 seeds	150
Baseline Norovirus	30 seeds	150
Uniform Vaccination	19 values of p 10 seeds	950
Vaccination by Degree	61 values of λ (Monday)	305
	65 values of λ (week)	325
	69 values of λ (month)	345
Vaccination by k -Core	61 values of λ (Monday)	305
	65 values of λ (week)	325
	69 values of λ (month)	345
Vaccination by Ring	61 values of λ (Monday)	305
	65 values of λ (week)	325
	69 values of λ (month)	345
Quarantine	2 values of p 30 seeds	300
<i>total</i>		4625

Table 3: Summary of Simulated Scenarios

6.1 Baseline Scenarios

We begin by infecting the Copenhagen network using our three basic models. This serves to confirm that our models are functional, to establish that an epidemic outbreak is possible, and to provide a baseline against which to compare the efficacy of our eventual attempted interventions.

6.1.1 Influenza

We first consider a single baseline influenza scenario, illustrated by Figure 15. Here, we have randomly selected a single seed for the infection, then run five simulations based on that starting situation. In other words, while the randomly chosen seed was held constant in all five simulations, the individual transmission events were not.

The complete results of simulating influenza without any interventions are plotted in Figure 16 and Figure 17. We find that, despite the very low probability of transmission per contact ($\beta = 0.015$), the disease experiences the rapid growth indicative of an epidemic, managing to affect almost every individual in the population: the curve does not quite reach 692. In the majority of our simulations, the outbreak occurs within the first week. There were, however, two situations in which the randomly chosen seed was not favorably positioned (with respect to the virus): in these cases, the outbreak was delayed for about a week, as the number of individuals in the dataset increases over the first week.

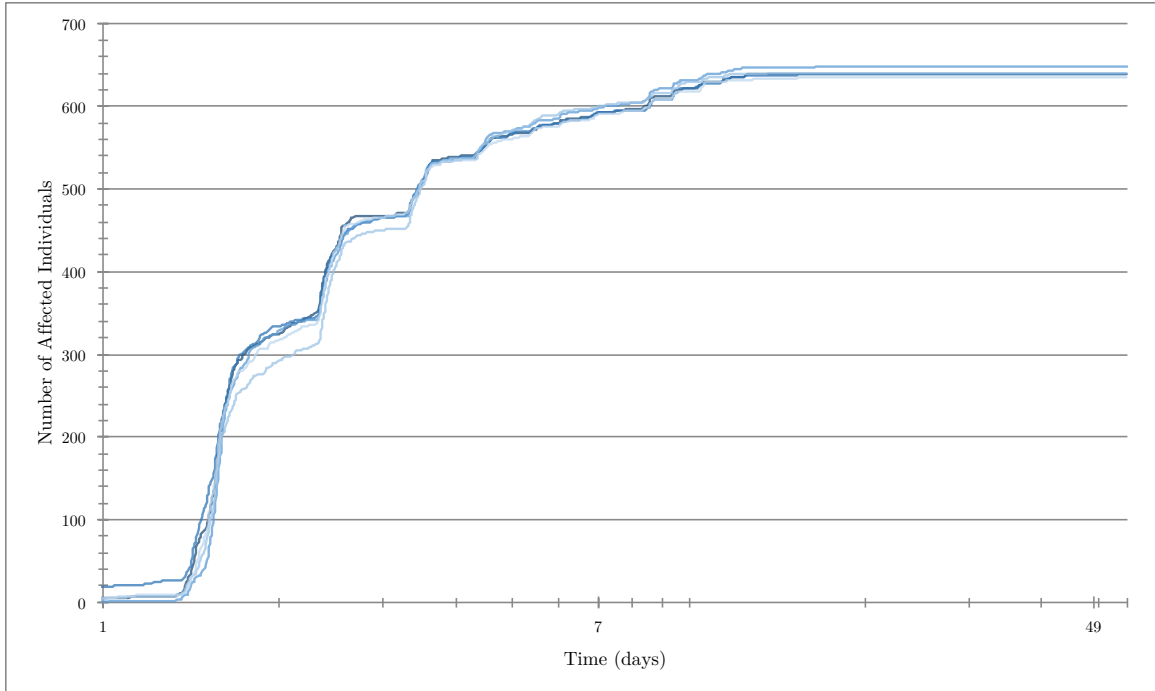


Figure 15: Affected Individuals in a Baseline Influenza Scenario. Five influenza simulations used the same initial seed of the infection. Note the logarithmic scale for time to better illustrate the individual curves.

These effects are reflected in the plot of infected individuals, Figure 17. In most simulations, the number of currently infected individuals reaches its peak by the beginning of the second week, but there are two where the curve is shifted to the right by one week. We also note the difference in the general shape of the curves to the left and to the right of the peak: because we simulate transmission of disease through individual contact events that occur with some daily pattern, the increasing portion of the curve is noticeably stepped. However, once an individual is infected, we roll a die with regards to when they will recover, and this randomness smooths out the decreasing portion of the curve.

Finally, we note that these plots are likely not (and were not intended to be) entirely representative of a real-world influenza epidemic. In particular, these outbreaks are accelerated by the simplicity of our SIR influenza model. In this model, as soon as individual is infected, they move directly from the susceptible state to the infected

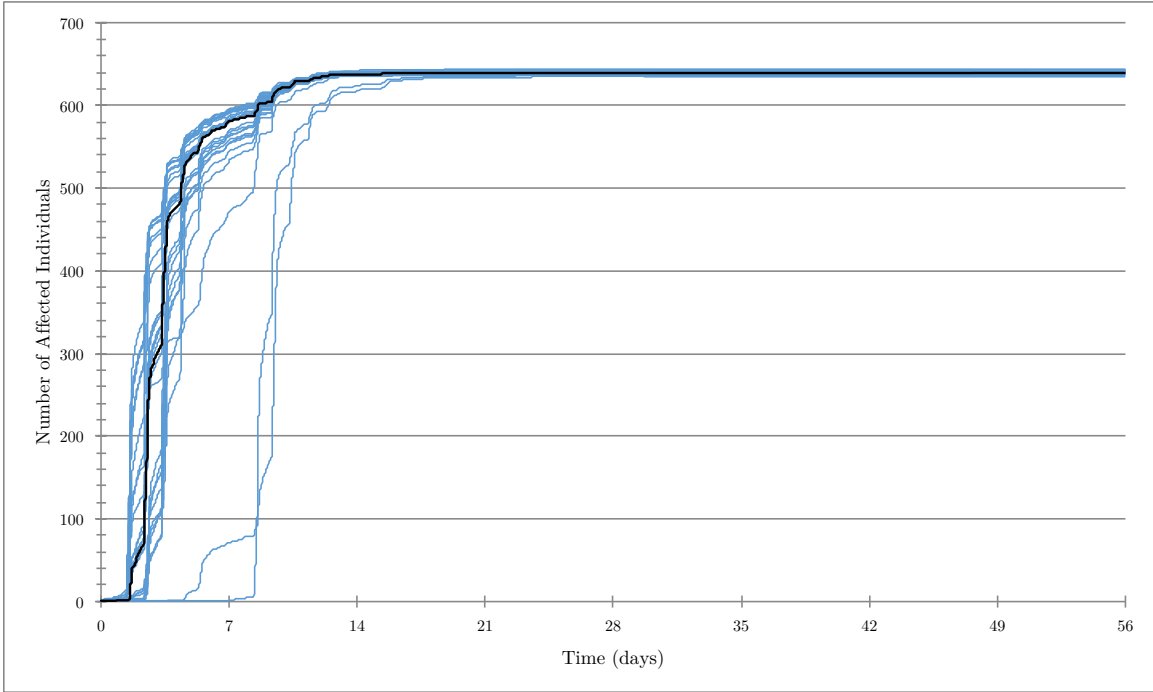


Figure 16: Affected Individuals in Baseline Influenza Scenarios. Each curve is an average of five simulations with the same infection seed and no interventions; the pointwise median of thirty curves is shown in black.

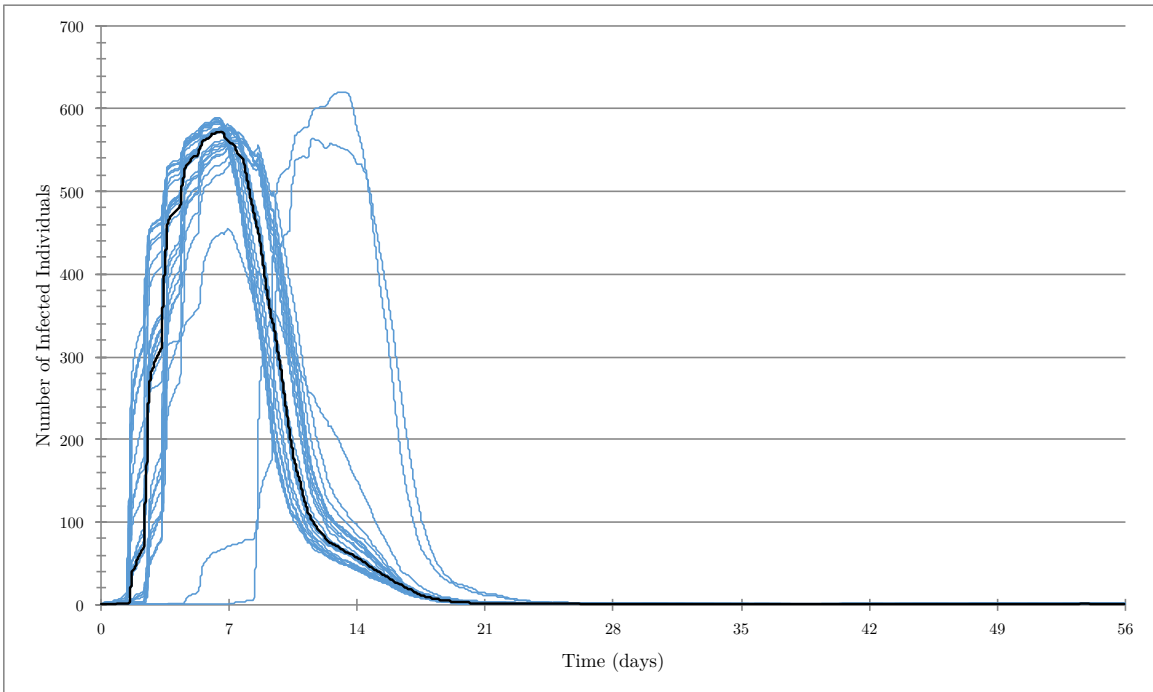


Figure 17: Infected Individuals in Baseline Influenza Scenarios. Each curve is an average of five simulations with the same infection seed and no interventions; the pointwise median of thirty curves is shown in black.

state; they are immediately capable of passing on the infection. Thus, the plotted curves begin to show unencumbered, exponential growth within the first couple of days.

6.1.2 Measles

In contrast to the immediate dispersal of influenza, the measles virus spreads with a more complex pattern, as shown in Figure 18. Because the measles model uses the additional exposed state, and exposed individuals cannot transmit the disease, the outbreak initially appears to slow after infecting about a hundred individuals — certainly, it spreads more slowly than our influenza simulations. In reality, however, these are just the individuals that the original seed is capable of directly infecting. The outbreak is initially limited by the number of individuals with whom that seed makes direct contact. The incubation period of measles is between 10 and 12 days, so by day 12, the initial batch of exposed individuals start transitioning into the infected state, and only then can the epidemic truly spread.

This behavior is more evident in Figure 19, where we plot the number of exposed individuals in blue alongside that of infected individuals in red. Up until about day 10, the initial seed is the only individual capable of spreading the disease, and the distinctive stepped shape of the exposed curves reflects that seed's daily movement patterns. Notably, the infection spreads more slowly on days 7 and 8, which are weekends. Much more rapid growth of the exposed curves begins around day 12, at the same time we begin to see growth of the infected curves.

As we saw with influenza, there are a couple of cases where the initial seed is in a sparser region of the network. This does nothing to prevent an outbreak; it only delays it by a few days. We also observe that measles infects practically the entire population,

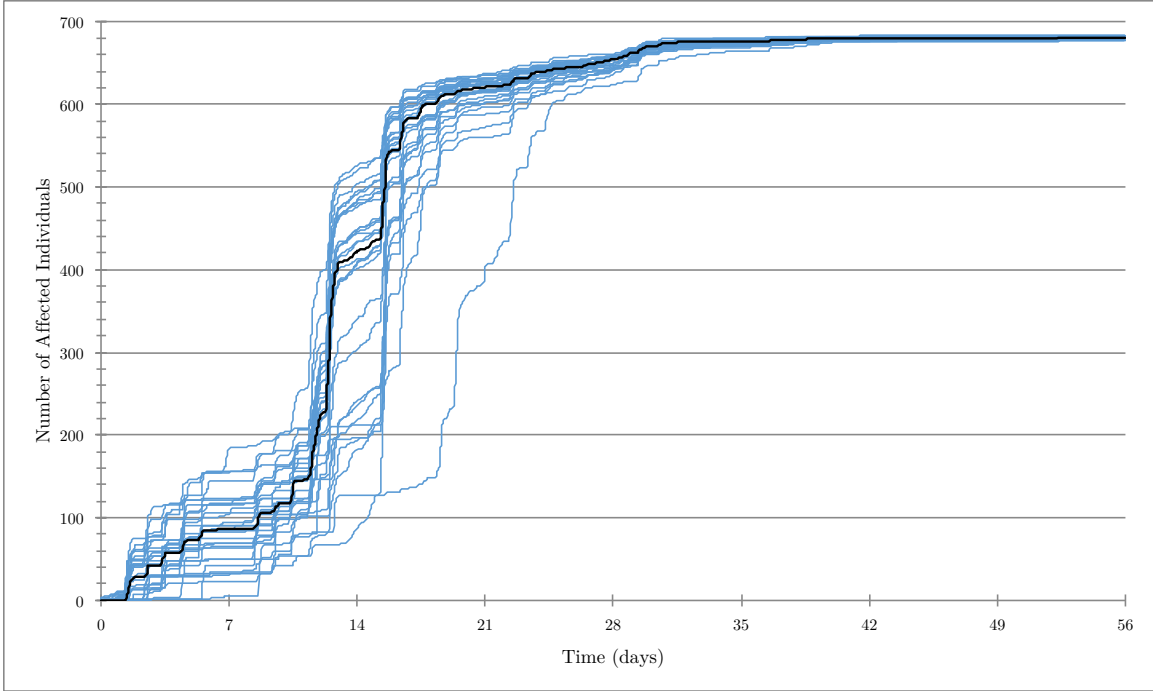


Figure 18: Affected Individuals in Baseline Measles Scenarios. Each curve is an average of five simulations with the same infection seed and no interventions; the pointwise median of thirty curves is shown in black.

as previously shown in Figure 18, but the peaks of the exposed and infected curves in Figure 19 do not get nearly so high. This is because the affected individuals are split between the exposed and infected states; they are never all in the same compartment at the same time. For the sake of completeness and later comparison, the sizes of all compartments for five individual simulations are presented in Figure 20.

6.1.3 Norovirus

Rounding out our baseline scenarios, the results of our norovirus simulations are summarized by Figure 21 and Figure 22. The norovirus model represents something of a middle ground between the influenza and measles models: it is generally much less virulent than measles, but its ability to infect clusters, as described in Section 4.3.3, allows it to spread faster than the flu. As with measles, individuals in the norovirus

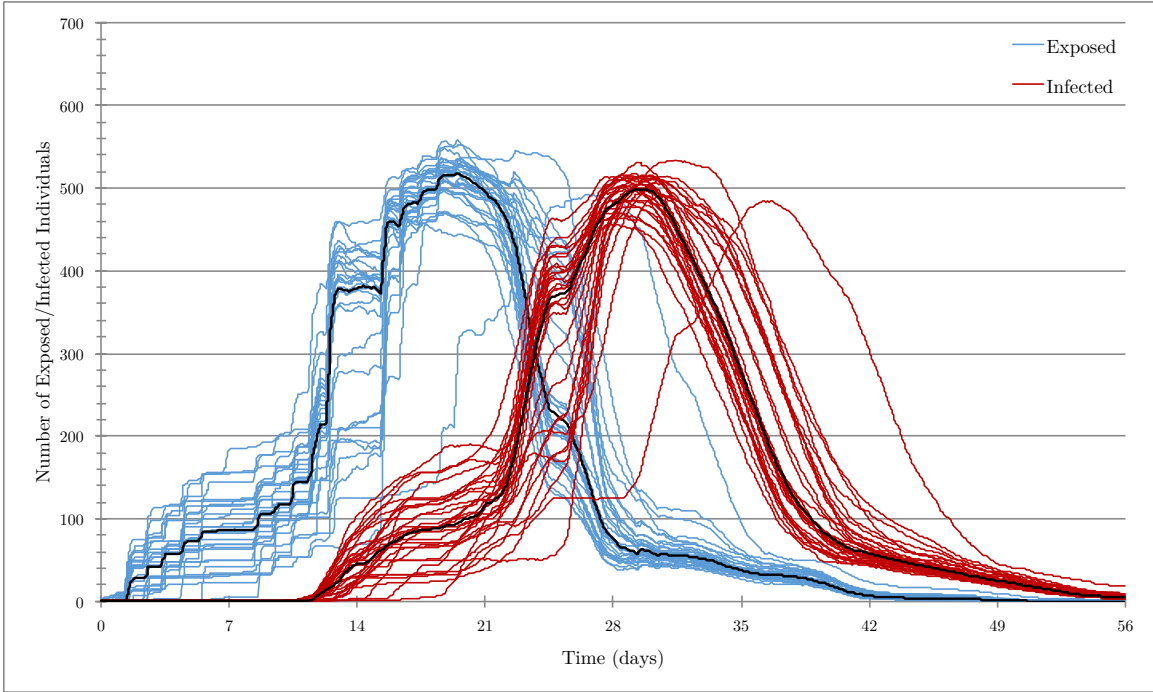


Figure 19: Exposed and Infected Individuals in Baseline Measles Scenarios. Each curve is an average of five simulations with the same seed and no interventions; the median of thirty curves is shown in black.

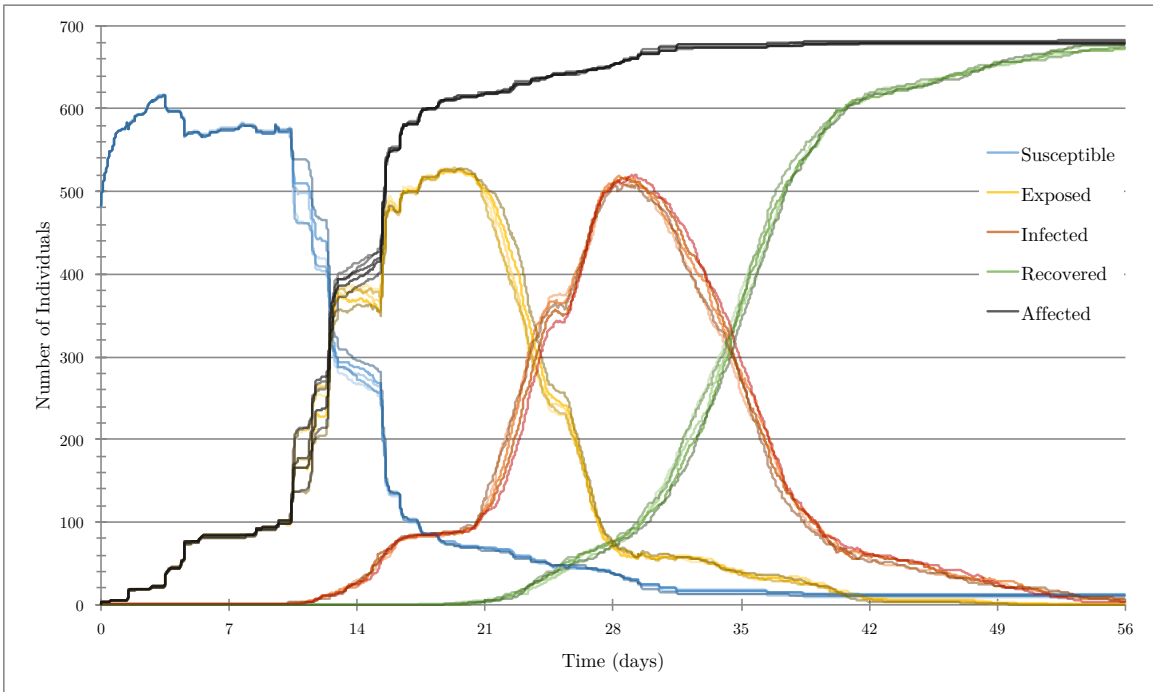


Figure 20: SEIR Compartments in a Baseline Measles Scenario. Five measles simulations used the same infection seed. Note the initial increase in susceptible individuals as they enter the Copenhagen network.

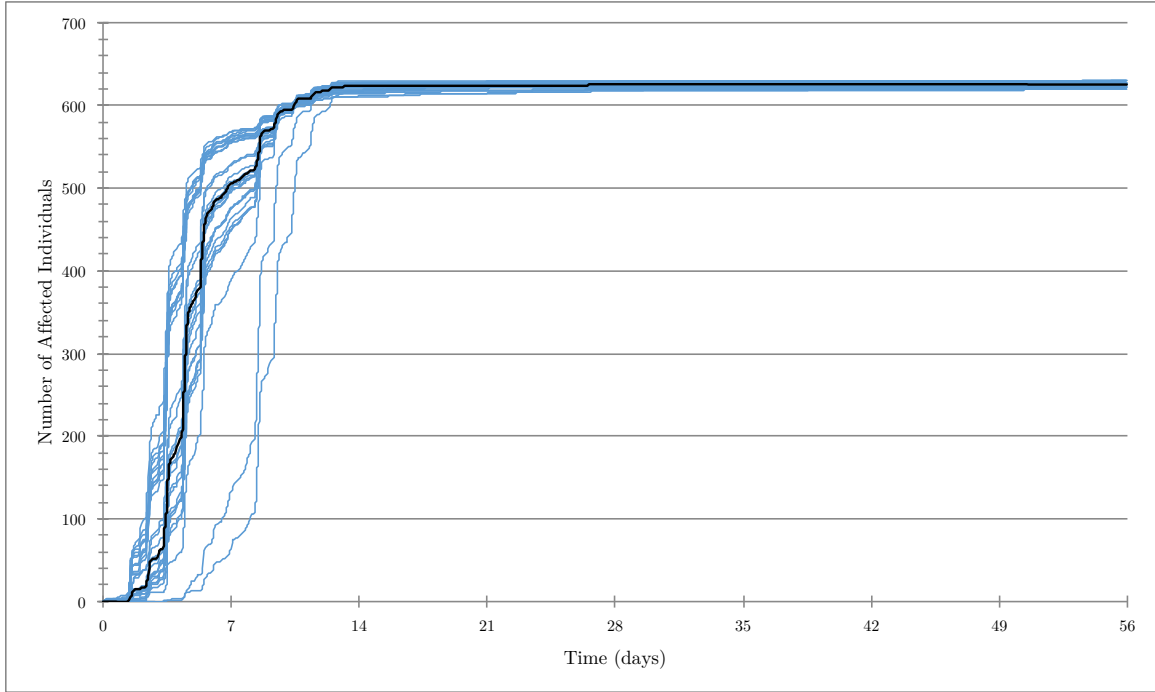


Figure 21: Affected Individuals in Baseline Norovirus Scenarios. Each curve is an average of five simulations with the same infection seed; the pointwise median of thirty curves is shown in black.

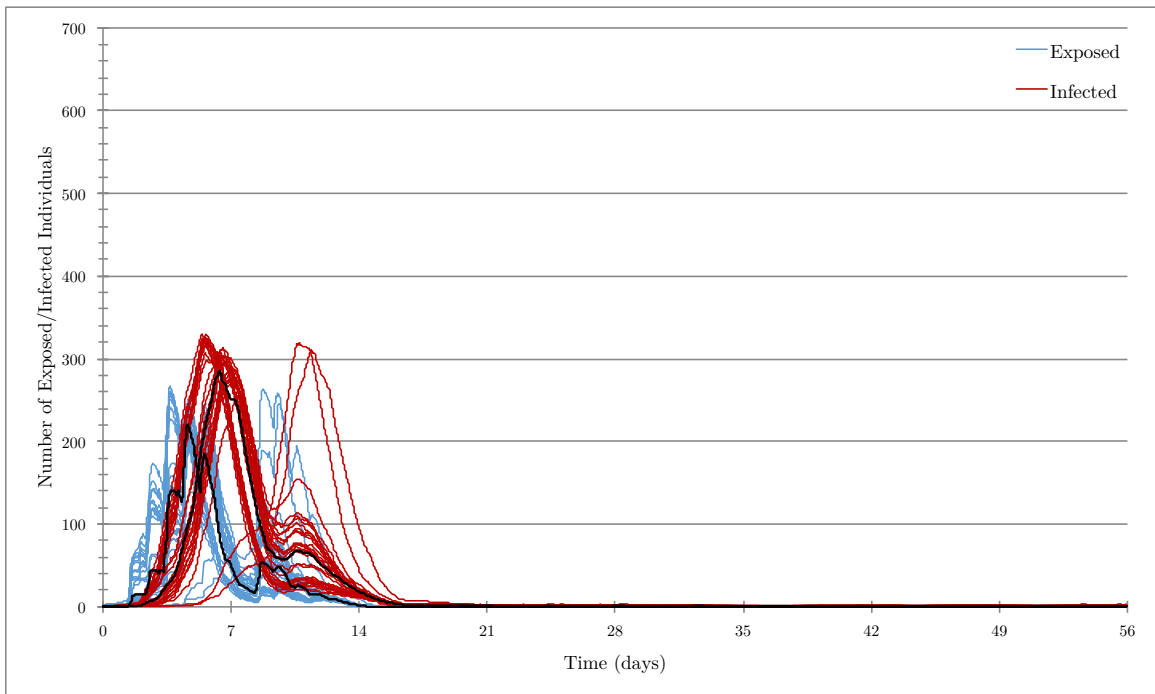


Figure 22: Exposed and Infected Individuals in Baseline Norovirus Scenarios. Each curve is an average of five simulations with the same seed; the median of thirty curves is shown in black.

model must pass through an exposed state before they can spread the disease, but that period is short enough for exponential growth to begin within a couple days, in similar fashion to the overly-simplistic flu simulation. Individuals also spend relatively little time in the infected state; the end result is that the exposed and susceptible curves of norovirus simulations rise and fall in rapid, overlapping spikes. Individuals quickly become exposed, transition to the infected state, infect others, and recover. As shown in Figure 22, the numbers of individuals in each compartment never reach the heights they do in flu or measles simulations, but as shown in Figure 21, the outbreak still affects the majority of the population.

6.2 Uniform Vaccination

In the uniform vaccination scenarios, individuals are placed in the recovered state instead of the susceptible state with some probability p upon their first appearance in the dataset. One selected example is shown in Figure 23. In this simulation, we set $p = 0.90$. Of the vaccination probabilities we tested, this most closely matches the vaccinated percentage of the population in the United States as of 2017 (CDC (Centers for Disease Control and Prevention), n.d.-a). The total number of individuals who have been affected by the disease is plotted in black. Though dramatically subdued compared to the baseline curve in Figure 20, this curve exhibits a period of faster-than-linear growth, and it flattens out as it approaches the maximum number of susceptible individuals, indicating that almost every susceptible individual has already been infected.

This general result is true of the other uniform vaccination interventions, summarized in Figure 24, which plots the number of affected individuals for each simulated vaccination probability. As enumerated in Table 3, we simulated 19 different probabilities.

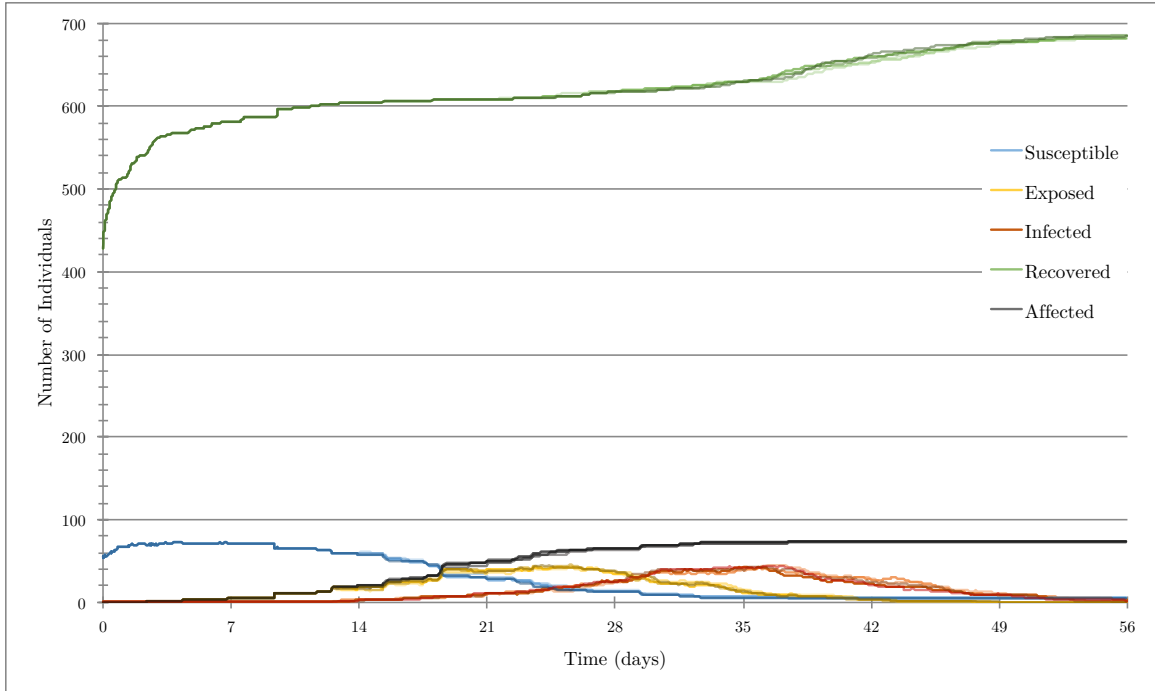


Figure 23: SEIR Compartments in a Uniform Vaccination Scenario. Five uniform vaccination simulations used the same initial seed, where $p = 0.90$. $\max_t\{|S_t|\} = 73$; $\min_t\{|S_t|\} = 5$; $\max_t\{|A_t|\} = 74$.

For each probability, we tested 10 different infection seeds, and for each seed, we ran 5 simulations. Thus, each curve in Figure 24 represents the median of 10 averages of 5 simulations each. As illustrated by the shapes of the curves, increasing the probability of vaccination neither alters nor delays the general pattern of the outbreak, it merely places a cap on the number of individuals that can eventually be infected.

This effect is reinforced by Figure 25. Here, we summarize the total number of affected individuals for each vaccination probability. Each box encapsulates the 50 simulations run for a single vaccination probability, and the dashed line indicates the expected number of individuals who would still be susceptible (i.e., $(1 - p) \times 692$). In uniform vaccination scenarios, essentially, everyone who could have been infected was eventually infected. This agrees with existing estimates that measles outbreaks can continue to occur even when $<10\%$ of the population is susceptible (Rota et al., 2016). We further infer that, due to the densely connected nature of the Copen-

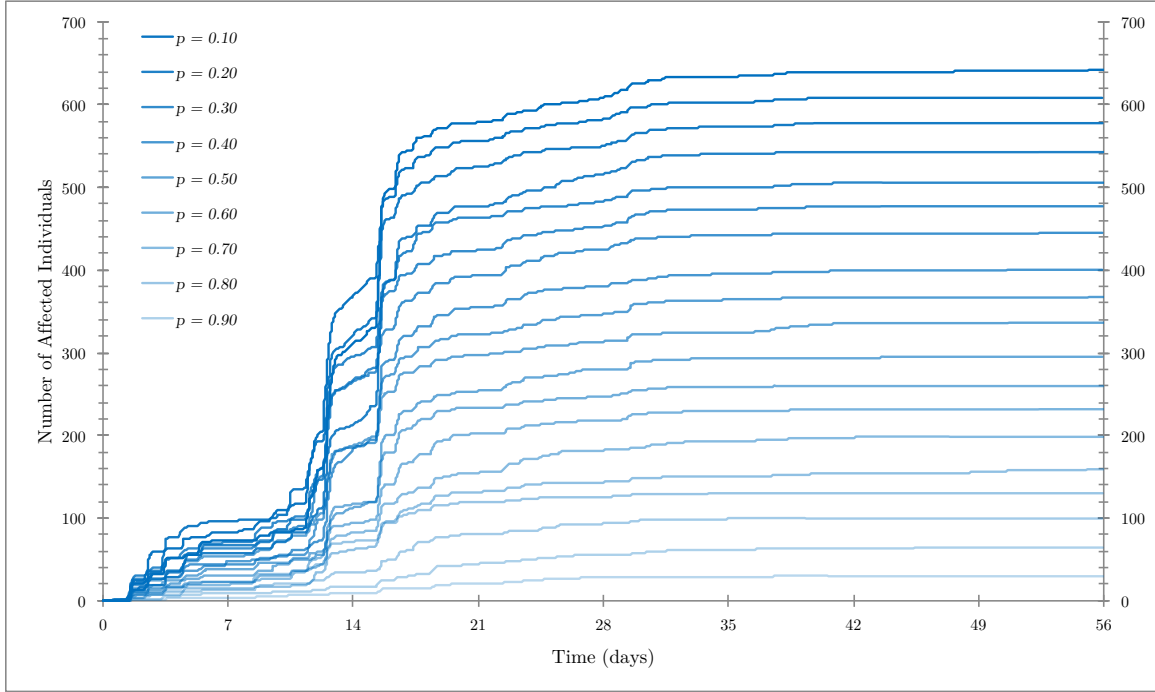


Figure 24: Affected Individuals in Uniform Vaccination Scenarios. Each line is the median of ten averages of five simulations each (ten infection seeds, five simulations per seed).

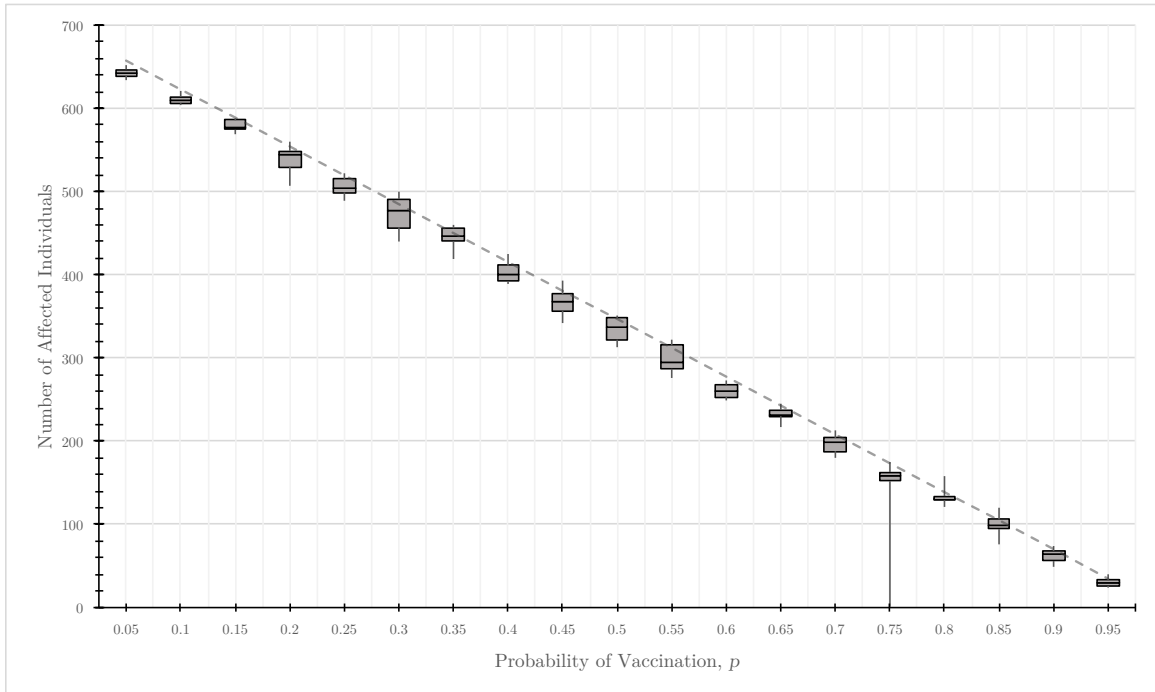


Figure 25: Summary of Uniform Vaccination Scenarios. Each box encapsulates fifty simulations. The dashed line indicates the expected number of susceptible (that is, not vaccinated) individuals.

hagen network, the herd immunity threshold (typically estimated between 89% and 94% (Rota et al., 2016)) appears to be higher than 95% in our simulations.

The only noticeable outlier occurs at $p = 0.75$: in this simulation, by random chance, we happened to vaccinate every individual immediately surrounding the one who would eventually be chosen as the seed of the infection, so the infection was never able to spread. This inadvertent targeted ring vaccination is, of course, the exception, not the trend.

6.3 Targeted Vaccination

We now imagine a hypothetical scenario in which we have a limited number of vaccines to distribute in preparation for an epidemic. In such a situation, we would like to try to target individuals for immunization in a more intelligent manner than uniform random vaccination.

6.3.1 Vaccination by Degree

We first simulated prioritizing individuals by the degrees of their vertices: in other words, those who made more unique contacts would receive the vaccine first. The results of this intervention are given by Figure 26. As detailed in Section 4.4.2, we tested three ways of computing the degree of each vertex in this network that changes over time. The simulation in the top-left of Figure 26 computes the degree based on the first Monday; in the top-right, based on the first week; in the center, based on the entire month.

We observe that vaccination by degree using just the first Monday’s data does not appear to make a difference compared to the uniform vaccination of Figure 24. This

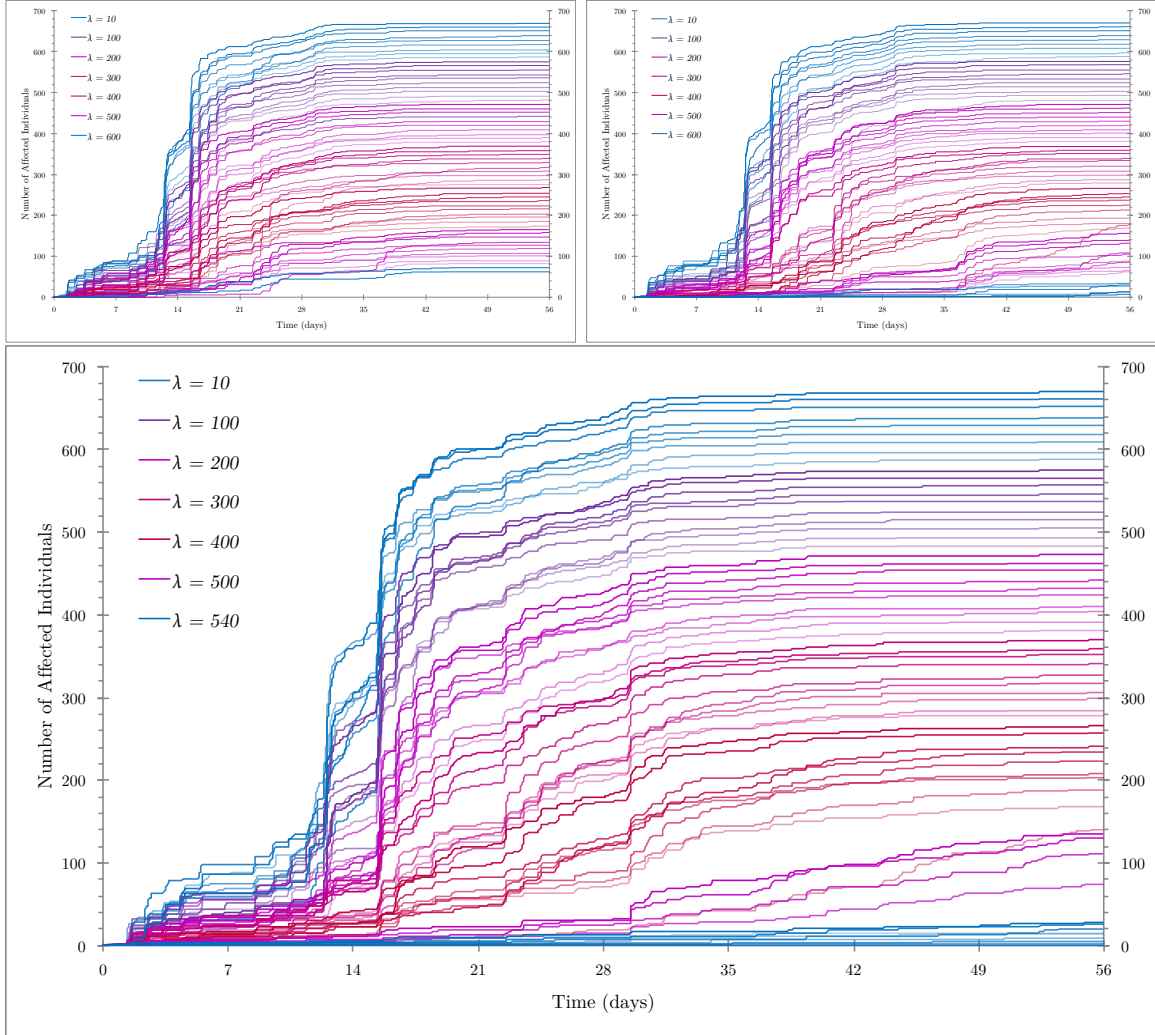


Figure 26: Affected Individuals when Vaccinated by Degree. Each curve is the median of five differently seeded simulations with λ vaccinations. Clockwise from top left: degrees computed from one Monday, from one week, and from one month.

makes intuitive sense: drawing from just the first Monday, there is simply not enough data about the contact patterns of vertices to make an intelligent decision regarding whom to vaccinate. Once we consider the data from the first week, we start to see some improvement at higher numbers of vaccinations. Starting at about $\lambda = 500$ vaccinations by highest degree, the outbreak has been delayed until the beginning of the fourth week. Our third simulation is, admittedly, not so pragmatic, since it requires knowing all the contacts each individual will make over a month. But if we happen to have that information, and we vaccinate according to highest degree, we

find that around $\lambda = 540$, the disease no longer infects the majority of susceptible individuals (540 vaccinated, thus, 152 susceptible, and a median of only 28 infected). In other words, the herd immunity threshold has been lowered to around 80% — a remarkable improvement over uniform vaccination.

6.3.2 Vaccination by k -Core

Our second targeted vaccination strategy involves the use of k -cores. As with vaccination by degree, we simulated administering a limited number of vaccinations, prioritizing individuals whose vertices were in higher cores, and we computed those cores based on the first Monday, the first week, and the entire month. These results are shown in Figure 27.

Disappointingly, we find that vaccination by k -core does not appear to perform any better than uniform random vaccination. In retrospect, however, this makes sense: higher k -cores identify subgraphs that are clique-like, whereas higher degrees identify the centers of subgraphs that are star-like. In a social network, we find it believable that there are more stars than cliques. Moreover, k -cores identify denser regions of a graph; they approximately identify clusters of vertices. In a targeted vaccination scenario, we do not necessarily want to vaccinate the clusters of individuals themselves — we want to vaccinate the individuals who connect those clusters to each other.

6.3.3 Vaccination by Ring

Finally, we turn our attention to the density decomposition; we prioritize the vaccination of individuals whose vertices appear in higher rings, in the same manner as we did with degrees and k -cores. These simulations are plotted in Figure 28, and they show that vaccination by ring has similar effects to vaccination by degree. Given only

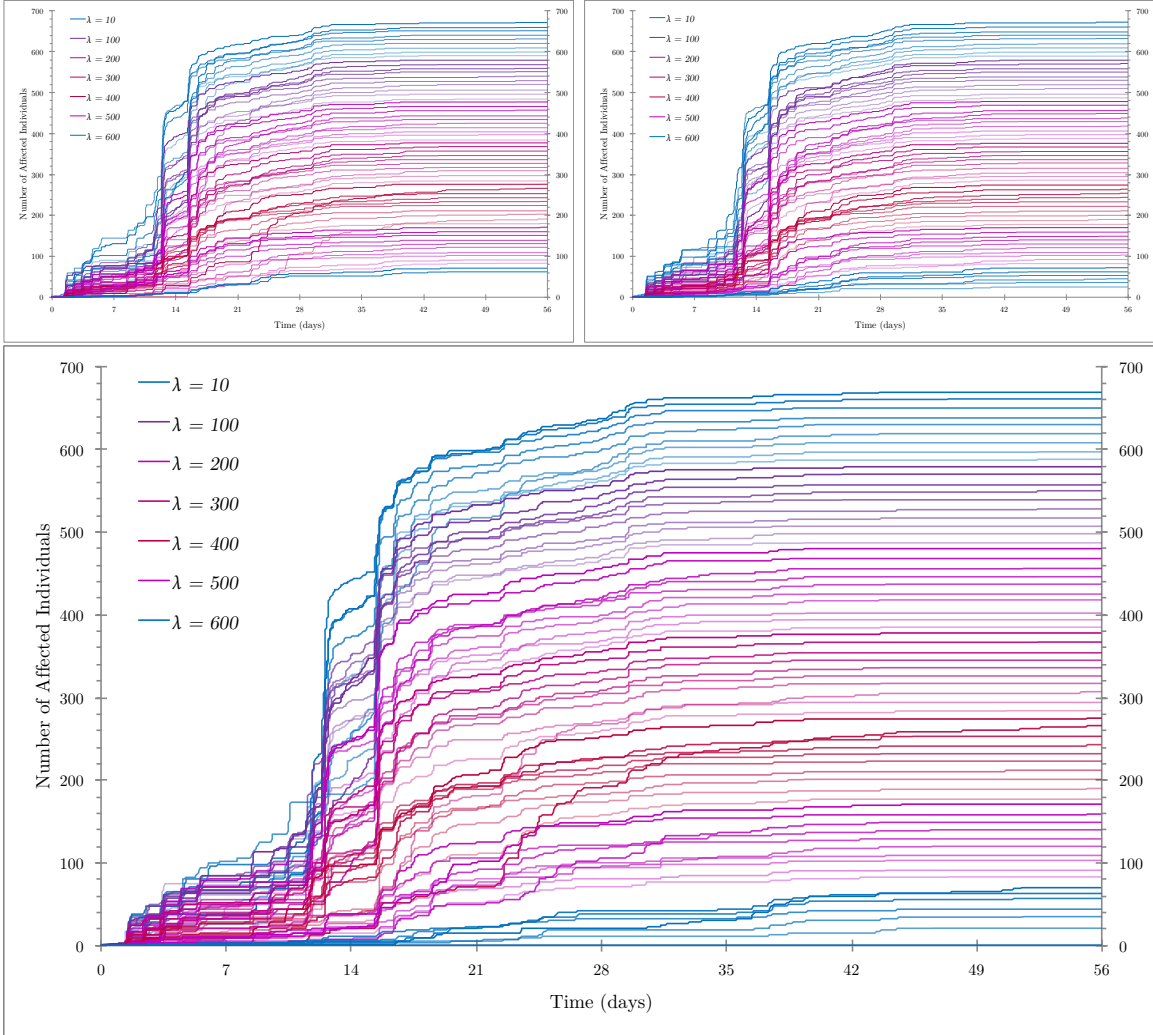


Figure 27: Affected Individuals when Vaccinated by k -Core. Each curve is the median of five differently seeded simulations with λ vaccinations. Clockwise from top left: k -cores computed from one Monday, from one week, and from one month.

the first Monday’s contact data, we see no improvement over uniform vaccination. Given the first week’s data, we can’t prevent the outbreak, but we can certainly delay it. And given the entire month’s data, the disease fails to become an epidemic around $\lambda = 550$ (142 susceptible, median of 26 infected), once again suggesting that the herd immunity threshold has been lowered to about 80%.

However, while vaccination by ring produces similar results to vaccination by degree, it takes more computational resources to orchestrate. The degree of every vertex in

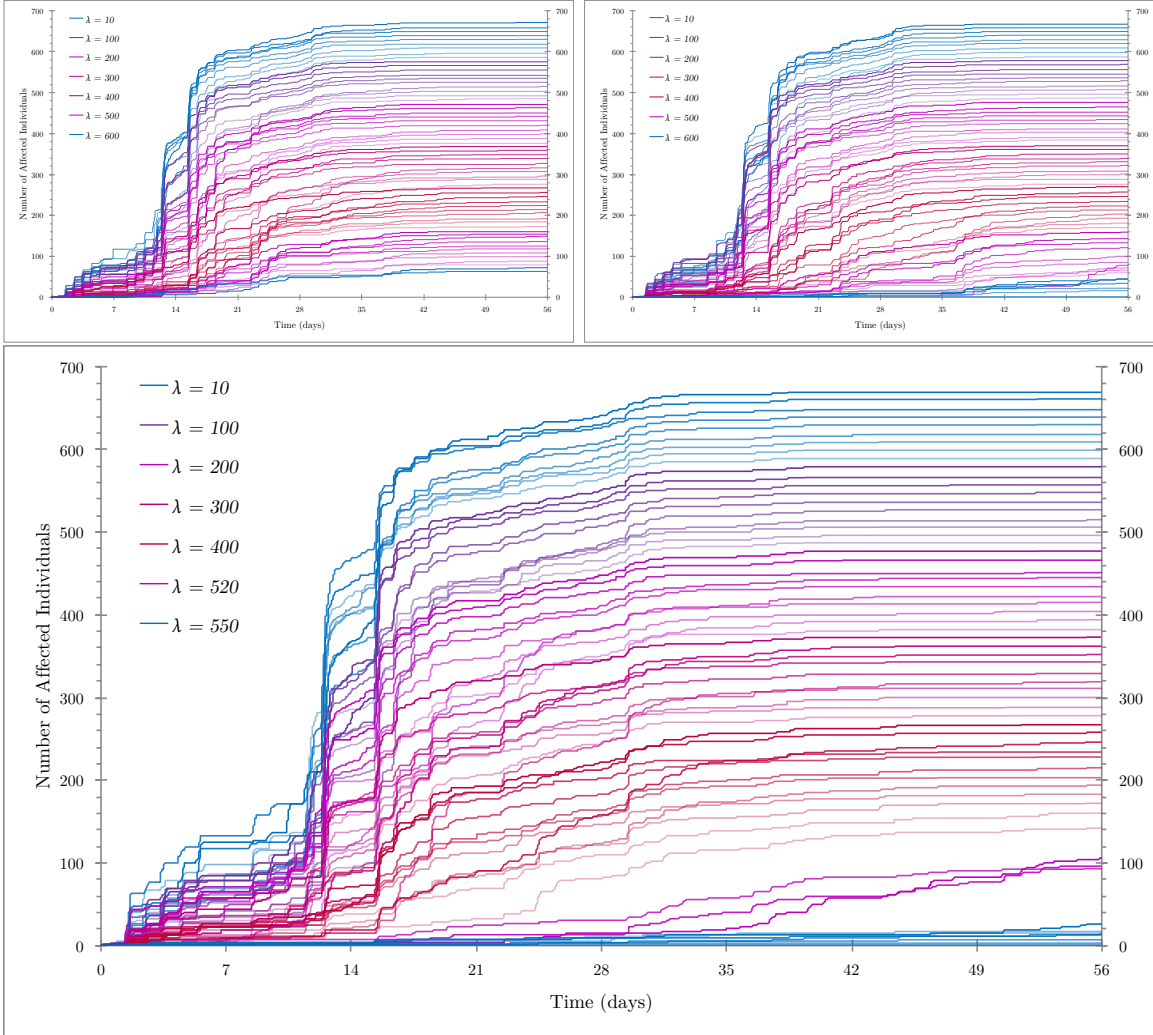


Figure 28: Affected Individuals when Vaccinated by Ring. Each curve is the median of five differently seeded simulations with λ vaccinations. Clockwise from top left: rings computed from one Monday, from one week, and from one month.

a graph can be computed in linear time with a streaming model; it can be computed as the graph is being parsed from the contact data. The density decomposition of the graph first requires that the entire graph be loaded into memory, and then its running time scales quadratically with the number of edges — of which there are very many, per Figure 8.

6.4 Quarantine Scenarios

Our final intervention introduces quarantining of infected individuals. As described in Section 4.4.5, we assume a perfect quarantine with which individuals are completely compliant, and we do not attempt to proactively quarantine anyone who has not yet exhibited the distinctive measles rash. Specifically, we introduce the quarantine state to two previous scenarios.

The results of enforcing a quarantine within the baseline measles scenario are shown in Figure 29. Because the initial seed will be quarantined as soon as they exhibit a rash, there is a 2–4 day window at the beginning of the simulation when they must infect others, else the infection can never spread. Once the initial seed infects their contacts, there is a clearly visible period of about 10 days when nothing happens. This is because the initial seed has moved to the quarantined state, and those they infected are still in the exposed state. As we saw with the baseline simulations in Figure 18, rapid growth is delayed at least until the first batch of exposed individuals becomes infectious. Those in that first batch are likewise limited to a 2–4 day window before their own quarantine begins — depending on how successfully they spread the disease, there may still be a significant number of uninfected individuals until after the second or third round of infections. Ultimately, in the majority of our simulations, a quarantine alone did not reduce the total number of individuals infected; it only changed the pattern by which that number was reached.

We then combined the quarantine strategy with a uniform vaccination probability of $p = 0.90$, which, of all our models, we felt would be a plausible setup and response to a measles outbreak. These results are plotted in Figure 30, and they are much more varied (sample standard deviation of final affected totals: $s = 17.4$). With a vaccination probability of $p = 0.90$, we expect that about 69 individuals would be

susceptible. The combination of a shorter infectious period with fewer susceptible targets means that at least three, often four “generations” of individuals had to be successively exposed before an outbreak approached its final total. Note the extended axis for time in Figure 30: we looped the Copenhagen dataset three times rather than two for these simulations to give the infection more time to propagate. We conclude that in a real-world scenario, a simple responsive quarantine is not adequate to contain an outbreak within such a densely connected local population, but it could delay the epidemic so that a more appropriate intervention could be attempted.

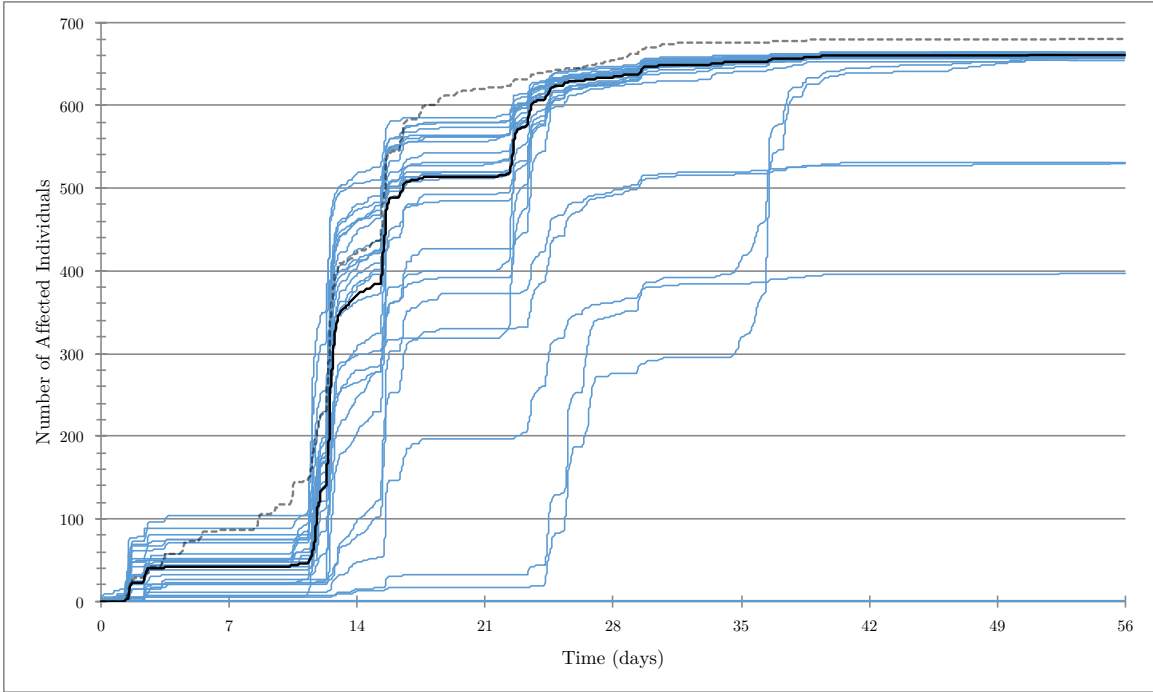


Figure 29: Affected Individuals in Quarantine Scenarios. Each curve is an average of five simulations with the same infection seed; the pointwise median of thirty curves is shown in black. The dashed line is the baseline median from Figure 18.

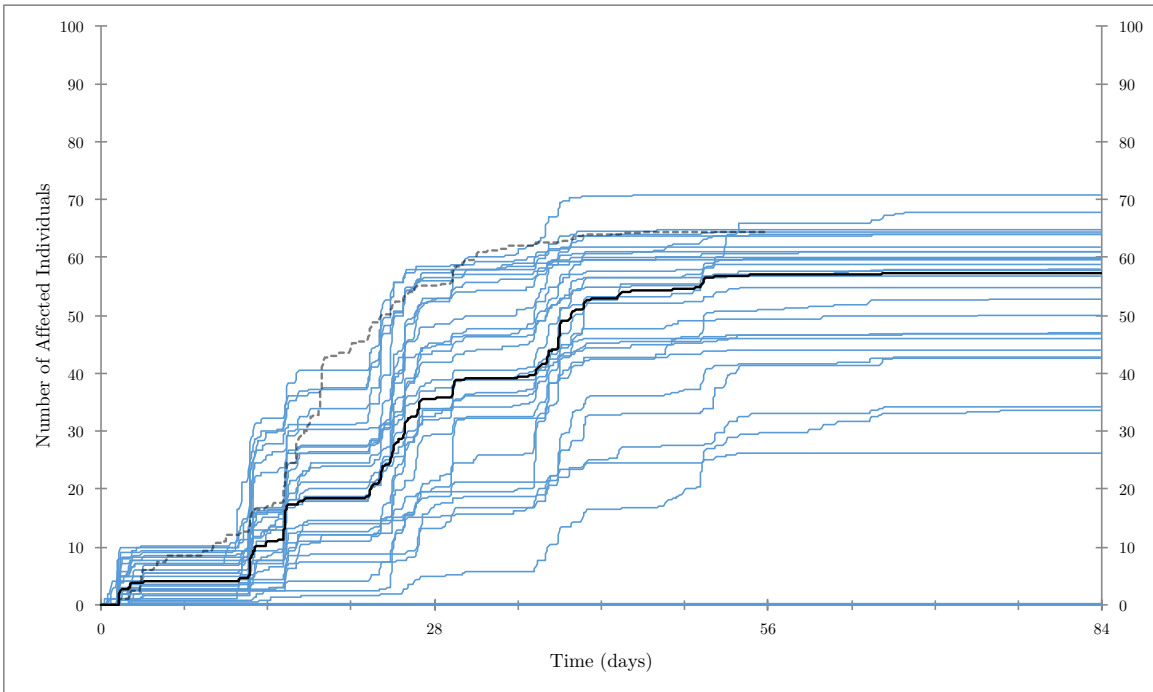


Figure 30: Affected Individuals in Vaccination-Quarantine Scenarios. Each curve is an average of five simulations with the same seed, where $p = 0.90$; the median of thirty curves is shown in black. The dashed line is the median from Figure 24.

Chapter 7

CONCLUSION

7.1 Future Work

7.1.1 Ineffective Vaccinations

In our intervention strategies, we attempted to contain measles, for which there exists a vaccine that is 97% effective (CDC (Centers for Disease Control and Prevention), n.d.-c). This is not the case for all diseases. As of 2017, for example, there was no vaccine for malaria, and the best candidate had between 26% and 40% efficacy (Phillips et al., 2017). Just as we applied existing graph metrics to target individuals in the event of limited vaccinations, similar approaches could be applied when simulating experimental vaccines that may not be so effective (though malaria itself does not spread through face-to-face contact, so it would not be realistic to simulate that particular disease with the Copenhagen dataset).

7.1.2 Additional Community Detection Algorithms

When developing our norovirus model, we incorporated a crude streaming algorithm to detect clusters of individuals. In a sense, our targeted vaccination strategies did the same thing, just with a different goal: to vaccinate key clusters rather than infect them. Our approaches are certainly not the only ways to detect clusters or other community structures. The iGraph library upon which our simulations are built, for example, already implements a number of community detection algorithms (The iGraph Core Team, 2015). We believe there is plenty of potential in applying such

algorithms, both to increase the realism of simulations and to increase the efficacy of interventions.

7.1.3 Combinations of Interventions

In our simulations, we found that neither a perfect quarantine nor extensive uniform vaccinations alone could contain an epidemic. However, combined, they can at least delay its spread, providing the opportunity to orchestrate an appropriate response. The natural next step is to explore other combinations of interventions. In particular, our quarantine policies were rather conservative. A more aggressive quarantine combined with targeted vaccinations should prove more effective.

7.1.4 Simultaneous Infections

Our epidemic simulations each employ a single compartmental model for a single disease. In the real world, it's possible for multiple outbreaks to occur simultaneously, and this may necessitate a change in interventions. In the case of measles, we could attempt to proactively treat individuals as soon as they display any symptoms, rather than waiting for the rash to appear. However, measles begins with a few cold- or flu-like symptoms (Rota et al., 2016) — what happens if an influenza outbreak occurs at the same time? If we are not careful, we might place both individuals infected with measles and individuals infected with the flu but susceptible to measles in the same hospital wing, for instance. Simultaneous infections should be simulatable with a few tweaks to our models.

7.1.5 External Contacts and Deaths

There are some individuals in the Copenhagen dataset who do not appear in the first snapshot, and so the number of individuals increases over roughly the first week, as seen in Figure 20. For the most part, however, our network does not grow. Real populations are, of course, typically not so isolated. Even if they are, their sizes could still change through the deaths of individuals. Our simulations are not so long for births to matter, but external contacts and deaths could be incorporated.

7.2 Summary of Contributions

To rectify the imprecision of existing techniques for mathematically modeling epidemics, we have simulated disease spreading on a high resolution social network, using a real-world dataset that captures individual human interactions with extremely fine granularity. Using these simulations, we implemented twelve strategies for intervening and either preventing or containing an infection. We showed that, by representing the network as a graph, we could apply existing graph metrics to inform more intelligent interventions. Notably, we showed that targeted vaccination based on the degree distribution or the density decomposition of the network could delay or prevent a measles-like epidemic using fewer vaccinations than existing estimates, which must assume a random distribution.

References

- Borradaile, G., Migler, T., & Wilfong, G. (2018). Density decompositions of networks. In S. Cornelius, K. Coronges, B. Gonçalves, R. Sinatra, & A. Vespignani (Eds.), *Complex networks ix* (pp. 15–26). Cham: Springer International Publishing.
- CDC (Centers for Disease Control and Prevention). (n.d.-a). *Fast stats - immunization*. Retrieved May 23, 2019, from <https://web.archive.org/web/20190523040139/https://www.cdc.gov/nchs/fastats/immunize.htm>.
- CDC (Centers for Disease Control and Prevention). (n.d.-b). *Measles | cases and outbreaks | cdc*. Retrieved May 16, 2019, from <https://web.archive.org/web/20190516082526/https://www.cdc.gov/measles/cases-outbreaks.html>.
- CDC (Centers for Disease Control and Prevention). (n.d.-c). *Measles | vaccination | cdc*. Retrieved May 16, 2019, from <https://web.archive.org/web/20190516074529/https://www.cdc.gov/measles/vaccination.html>.
- CDC (Centers for Disease Control and Prevention). (n.d.-d). *Norovirus | transmission | cdc*. Retrieved May 22, 2019, from <https://web.archive.org/web/20190522103408/https://www.cdc.gov/norovirus/about/transmission.html>.
- CDC (Centers for Disease Control and Prevention). (n.d.-e). *Pinkbook | measles | epidemiology of vaccine-preventable disease | cdc*. Retrieved May 14, 2019, from <https://web.archive.org/web/20190514200016/https://www.cdc.gov/vaccines/pubs/pinkbook/meas.html>.
- Dietz, K., & Heesterbeek, J. (2002). Daniel Bernoulli's epidemiological model revisited. *Mathematical Biosciences*, 180(1), 1 - 21. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0025556402001220> doi: [https://doi.org/10.1016/S0025-5564\(02\)00122-0](https://doi.org/10.1016/S0025-5564(02)00122-0)

- Duggan, A. T., Perdomo, M. F., Piombino-Mascali, D., Marciniak, S., Poinar, D., Emery, M. V., ... Poinar, H. N. (2016). 17th century variola virus reveals the recent history of smallpox. *Current Biology*, 26(24), 3407 - 3412. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0960982216313240> doi: <https://doi.org/10.1016/j.cub.2016.10.061>
- Duret, S., Pouillot, R., Fanaselle, W., Papafragkou, E., Liggans, G., Williams, L., & Van Doren, J. M. (2017). Quantitative risk assessment of norovirus transmission in food establishments: Evaluating the impact of intervention strategies and food employee behavior on the risk associated with norovirus in foods. *Risk Analysis*, 37(11), 2080-2106. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1111/risa.12758> doi: 10.1111/risa.12758
- Euler, L. (1741). Solvatio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum imperialis Petropolitanae*, 8, 128-140.
- Feng, Z., & Thieme, H. R. (1995). Recurrent outbreaks of childhood diseases revisited: The impact of isolation. *Mathematical Biosciences*, 128(1), 93-130. Retrieved from <http://www.sciencedirect.com/science/article/pii/002555649400069C> doi: [https://doi.org/10.1016/0025-5564\(94\)00069-C](https://doi.org/10.1016/0025-5564(94)00069-C)
- Fenner, F., Henderson, D. A., Arita, I., Jezek, Z., & Ladnyi, I. D. (1988). *Smallpox and its eradication* (No. 6). World Health Organization.
- Frías-Martínez, E., Williamson, G., & Frías-Martínez, V. (2011, Oct). An agent-based model of epidemic spread using human mobility and social network information. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing* (p. 57-64). doi: 10.1109/PASSAT/SocialCom.2011.142
- Galimberti, E., Barrat, A., Bonchi, F., Cattuto, C., & Gullo, F. (2018). Mining (maximal) span-cores from temporal networks. In *Proceedings of the 27th ACM international conference on information and knowledge management* (pp.

- 107–116). New York, NY, USA: ACM. Retrieved from <http://doi.acm.org/10.1145/3269206.3271767> doi: 10.1145/3269206.3271767
- Gates, B. (2015). The next epidemic — lessons from ebola. *New England Journal of Medicine*, *372*(15), 1381-1384. Retrieved from <https://doi.org/10.1056/NEJMp1502918> (PMID: 25853741) doi: 10.1056/NEJMp1502918
- Gemmetto, V., Barrat, A., & Cattuto, C. (2014, Dec 31). Mitigation of infectious disease at school: targeted class closure vs school closure. *BMC Infectious Diseases*, *14*(1), 695. Retrieved from <https://doi.org/10.1186/s12879-014-0695-9> doi: 10.1186/s12879-014-0695-9
- Gilbert, M. (1994). *The first world war*. Henry Holt and Company, Inc.
- Granell, C., & Mucha, P. (2018, 5 3). Epidemic spreading in localized environments with recurrent mobility patterns. *Physical review. E*, *97*(5). doi: 10.1103/PhysRevE.97.052302
- Johnson, N. P. A. S., & Mueller, J. (2002). Updating the accounts: Global mortality of the 1918-1920 “spanish” influenza pandemic. *Bulletin of the History of Medicine*, *76*(1), 105–115. Retrieved from <https://doi.org/10.1353/bhm.2002.0022>
- Kephart, J. O., & White, S. R. (1991, May). Directed-graph epidemiological models of computer viruses. In *Proceedings. 1991 ieee computer society symposium on research in security and privacy* (p. 343-359). doi: 10.1109/RISP.1991.130801
- Kermack, W. O., & McKendrick, A. G. (1927, August). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A*, *115*, 700-721. doi: <https://doi.org/10.1098/rspa.1927.0118>
- Krammer, F., Smith, G. J. D., Fouchier, R. A. M., Peiris, M., Kedzierska, K., Doherty, P. C., ... García-Sastre, A. (2018). Influenza. *Nature Reviews Disease Primers*, *4*(1), 3. Retrieved from <https://doi.org/10.1038/s41572-018-0002-y> doi: 10.1038/s41572-018-0002-y

- Matsuyama, R., Miura, F., & Nishiura, H. (2017, 03). The transmissibility of noroviruses: Statistical modeling of outbreak events with known route of transmission in japan. *PLOS ONE*, *12*(3), 1-16. Retrieved from <https://doi.org/10.1371/journal.pone.0173996> doi: 10.1371/journal.pone.0173996
- Matula, D. W., & Beck, L. L. (1983, July). Smallest-last ordering and clustering and graph coloring algorithms. *J. ACM*, *30*(3), 417–427. Retrieved from <http://doi.acm.org/10.1145/2402.322385> doi: 10.1145/2402.322385
- McEvedy, C. (1988). The bubonic plague. *Scientific American*, *258*(2), 118–123. Retrieved from <http://www.jstor.org/stable/24988987>
- Mills, C. E., Robins, J. M., & Lipsitch, M. (2004). Transmissibility of 1918 pandemic influenza. *Nature*, *432*(7019), 904–906. Retrieved from <https://doi.org/10.1038/nature03063> doi: 10.1038/nature03063
- Murray, J. D., Stanley, E. A., & Brown, D. L. (1986, November). On the spatial spread of rabies among foxes. *Proceedings of the Royal Society B*, *229*, 111-150. doi: <https://doi.org/10.1098/rspb.1986.0078>
- OED Online. (n.d.). *epidemic, adj. and n.* Retrieved May 27, 2019, from www.oed.com/view/Entry/63307. Oxford University Press.
- Phillips, M. A., Burrows, J. N., Manyando, C., van Huijsduijnen, R. H., Van Voorhis, W. C., & Wells, T. N. C. (2017, 08 03). Malaria. *Nature Reviews Disease Primers*, *3*, 17050 EP -. Retrieved from <https://doi.org/10.1038/nrdp.2017.50>
- Riedel, S. (2005, 01). Edward jenner and the history of smallpox and vaccination. *Proceedings (Baylor University. Medical Center)*, *18*(1), 21–25. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/16200144>
- Rota, P. A., Moss, W. J., Takeda, M., de Swart, R. L., Thompson, K. M., & Goodson, J. L. (2016, 07 14). Measles. *Nature Reviews Disease Primers*, *2*, 16049 EP -. Retrieved from <https://doi.org/10.1038/nrdp.2016.49>

- Sapieżyński, P. (n.d.). personal communication.
- Seidman, S. B. (1983). Network structure and minimum degree. *Social Networks*, 5(3), 269 - 287. Retrieved from <http://www.sciencedirect.com/science/article/pii/037887338390028X> doi: [https://doi.org/10.1016/0378-8733\(83\)90028-X](https://doi.org/10.1016/0378-8733(83)90028-X)
- Stopczynski, A., Pentland, A. S., & Lehmann, S. (2018). How physical proximity shapes complex social networks. *Scientific Reports*, 8(1), 17722. Retrieved from <https://doi.org/10.1038/s41598-018-36116-6> doi: 10.1038/s41598-018-36116-6
- Stopczynski, A., Sekara, V., Sapiezynski, P., Cuttone, A., Madsen, M. M., Larsen, J. E., & Lehmann, S. (2014, 04). Measuring large-scale social networks with high resolution. *PLOS ONE*, 9(4), 1-24. Retrieved from <https://doi.org/10.1371/journal.pone.0095978> doi: 10.1371/journal.pone.0095978
- The iGraph Core Team. (2015). *python-igraph*. Retrieved from <https://igraph.org/python/doc/igraph-module.html>.
- WHO (World Health Organization). (2017). Measles vaccines: WHO position paper — April 2017. *Weekly epidemiological record*(17), 205-228. Retrieved August 28, 2019, from <http://apps.who.int/iris/bitstream/10665/255149/1/WER9217.pdf>.