Estimating Network Structure and Propagation Dynamics for an Infectious Disease: Towards Effective Vaccine Allocation

by

Louis Y. Kim

B.S., Industrial and Enterprise Systems Engineering
University of Illinois at Urbana-Champaign, 2011

Submitted to the Sloan School of Management
in partial fulfillment of the requirements for the degree of

Master of Science in Operations Research

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2014

The author hereby grants to MIT and Draper Laboratory permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part.

Signature redacted

Author ................................................
Sloan School of Management
June, 2014

Certified by ........................................
Mark Abramson
Principal Member of Technical Staff
The Charles Stark Draper Laboratory
Thesis Supervisor

Certified by ........................................
Asuman Ozdaglar
Professor
Department of Electrical Engineering and Computer Science
Laboratory for Information and Decision Systems
Thesis Supervisor

Signature redacted

Accepted by ........................................
Dimitris Bertsimas
Boeing Professor of Operations Research
Co-director, Operations Research Center
Estimating Network Structure and Propagation Dynamics for an Infectious Disease: Towards Effective Vaccine Allocation

by

Louis Y. Kim

Submitted to the Sloan School of Management on June, 2014, in partial fulfillment of the requirements for the degree of Master of Science in Operations Research

Abstract

In the event of a pandemic influenza outbreak, such as the 2009-2010 H1N1 "Swine Flu" episode, it is crucial to effectively allocate limited resources in order to minimize the casualties. Design of effective resource allocation strategies requires good understanding of the underlying contact network and of the propagation dynamics.

In this thesis we develop a parameter estimation method that learns the network structure, among a family of graphs, and disease dynamics from the recorded infection curve, assuming that the disease dynamics follow an SIR process. We apply the method to data collected during the 2009-2010 H1N1 epidemic and show that the best-fit model, among a scale-free network and a small-world network, indicates the scale-free network.

Given the knowledge of the network structure we evaluate different vaccination strategies. As a benchmark, we allow the vaccination decisions to depend on the state of the epidemic and we show that random vaccination (which is the current practice), does not efficiently halt the spread of influenza. Instead, we propose vaccine allocation strategies that exploit the underlying network structure and provide a reduction in the number of infections by over 6 times compared to the current practice.

In addition, more realistic scenario involves random encounters between agents. To test this hypothesis, we introduced a dynamic network formation on top of the static network model. We apply the estimation method to the dynamic network model and show a small improvement in estimating the infection dynamics of the 2009-2010 H1N1 influenza.

Thesis Supervisor: Mark Abramson
Title: Principal Member of Technical Staff
The Charles Stark Draper Laboratory

Thesis Supervisor: Asuman Ozdaglar
Title: Professor
Department of Electrical Engineering and Computer Science Laboratory for Information and Decision Systems
Acknowledgments

I would not have been able to complete this work without the great support, encouragement, and help of so many people. First and foremost, I am deeply grateful to Mark Abramson and Stephan Kolitz at Draper Laboratory for their mentorship. From their dedicated support and invaluable coaching, I always found a genuine belief in my ability to solve challenging problems that I encountered while working on this thesis. I would also like to thank Asu Ozdaglar and Kimon Drakopoulos for their technical help and the very generous time they spent with me to advance this research. I am lucky to have worked with such amazing people at MIT and Draper Laboratory.

I have also had multiple conversations with Iain Dunning, Shannon Fast, Andrew Li, Miles Lubin, Allison O’Hair, Angie King, and Jonathan Paynter, as well as many other students at ORC, about my research. The open and cooperative environment at the ORC is structured exactly to encourage brainstorming new ideas with the help of many people. I am glad that I decided to pursue a Master’s degree here at ORC. Thank you to all my friends for your support!

I am forever grateful to my parents and my brothers for their unconditional love and support, not only during these past two challenging years, but throughout my entire life. I cannot expect any greater blessing. Thank you.

Finally, I want to thank the Charles Stark Draper Laboratory for the Draper Lab Fellow appointment that made my work here possible.
Contents

1 Introduction ........................................ 13
  1.1 Motivation .................................. 13
  1.2 Problem Statement ............................ 14
  1.3 Related Literature ............................. 15
    1.3.1 Epidemic Models ......................... 15
    1.3.2 Control of Epidemics .................... 16
    1.3.3 Network and Propagation Dynamics Inference 16
  1.4 Technical Approaches .......................... 17
  1.5 Experimentation ................................ 18
  1.6 Contributions .................................. 19
  1.7 Thesis Organization ........................... 19

2 Estimating Social Network Structure and Propagation Dynamics 21
  2.1 Introduction .................................. 21
  2.2 Methods ...................................... 22
    2.2.1 Data ..................................... 23
    2.2.2 Disease Dynamics ......................... 25
    2.2.3 Estimation Algorithm ....................... 28
      2.2.3.1 Small-World Network ................... 29
      2.2.3.2 Scale-Free Network ..................... 29
      2.2.3.3 Estimation Algorithm ................... 30
  2.3 Results ..................................... 32
    2.3.1 Estimation Algorithm Results ............... 32
    2.3.2 Discussion ................................ 36

3 Effective Control Policy ........................ 37
  3.1 Introduction .................................. 37
  3.2 Evaluating Effective Vaccine Allocation Strategies .......... 38
3.2.1 Methods ................................................................. 38
3.2.2 Results ................................................................. 39
   3.2.2.1 Vaccine Effectiveness ........................................ 39
   3.2.2.2 Effective Vaccine Allocation Strategies ............... 40
   3.2.2.3 Vaccine Effectiveness Under Varying Compliance Rates .... 51
   3.2.2.4 Sensitivity Analysis ....................................... 51
3.3 Discussion ............................................................. 52

4 Dynamic Network ...................................................... 55
   4.1 Introduction ....................................................... 55
   4.2 Dynamic Network Model and Parameter Estimation Results .... 56
      4.2.1 Dynamic Network Model ................................... 56
      4.2.2 Parameter Estimation on Dynamic Network Model ....... 57
      4.2.3 Estimation Results ....................................... 58
   4.3 Discussion ....................................................... 61

5 Conclusions and Future Research .................................. 63
   5.1 Conclusions ....................................................... 63
   5.2 Future Research ................................................ 64
List of Figures

2-1 The infection curve estimated from %ILI data and the number of vaccines administered during the observation period (October, 2009 - December, 2009) for Massachusetts. 24

2-2 The infection curve estimated from %ILI data and the number of vaccines administered during the observation period (October, 2009 - December, 2009) for New York State. ................................................................. 25

2-3 The infection curve estimated from %ILI data and the number of vaccines administered during the observation period (September, 2009 - December, 2009) for Montana. 26

2-4 The infection curve estimated from %ILI data and the number of vaccines administered during the observation period (September, 2009 - December, 2009) for North Dakota. ................................................................. 26

2-5 Small-world network generation steps [1]. ................................................................. 29

2-6 Flow diagram showing the estimation algorithm process. ........................................... 32

2-7 Massachusetts best fit curves generated by the algorithm using small-world network (green) and scale-free network (blue) compared to the estimated infection curve from data (black) and from using differential equations with “fully-mixed” assumption (red). 34

2-8 New York best fit curves generated by the algorithm using small-world network (green) and scale-free network (blue) compared to the estimated infection curve from data (black). ................................................................. 34

2-9 Montana best fit curves generated by the algorithm using small-world network (green) and scale-free network (blue) compared to the estimated infection curve from data (black). ................................................................. 35

2-10 North Dakota best fit curves generated by the algorithm using small-world network (green) and scale-free network (blue) compared to the estimated infection curve from data (black). ................................................................. 35

3-1 Plot of the expected infection curve for Massachusetts assuming no vaccine is available (black), compared to the expected infection when vaccine is randomly distributed (red). 40
3-2 Massachusetts scale-free network generated with estimated average degree of 2. Degree distribution is shown on top and probability plot of degree is shown on bottom.

3-3 Plot of expected infection curves for Massachusetts. Infection dynamics are simulated and compared under three cases: no vaccine distributed (black), vaccine randomly distributed (red), and distributed based on individual's degree (blue).

3-4 Comparison of expected infection curves under different cases: no vaccines distributed (black), vaccine randomly distributed (red), distributed based on individual's degree (blue), and distributed based on individual's number of healthy neighbors (purple).

3-5 Comparison of expected infection curves under different cases: no vaccine distributed (black), vaccine randomly distributed (red), distributed based on individual's degree (blue), distributed based on individual's number of healthy neighbors (purple), and distributed based on individual's number of infected neighbors (green).

3-6 Example network of 8 nodes (Top). Ordering of the vertices that generated the cutwidth of the network (Bottom).

3-7 Example network of 8 nodes in different states.

3-8 Resulting cutwidth when node 1 is removed.

3-9 Resulting cutwidth when node 4 is removed.

3-10 Resulting cutwidth when node 6 is removed.

3-11 The ranking of the susceptible individuals to receive vaccine in the order of the resulting cutwidth, from the lowest to the highest.

3-12 Comparison of expected infection curves under different cases: no vaccine distributed (black), vaccine randomly distributed (red), distributed based on individual's degree (blue), distributed based on individual's number of healthy neighbors (purple), distributed based on individual's number of infected neighbors (green), and distributed based on the cutwidth (light blue).

3-13 Percentage of infections averted by proposed vaccine distribution strategies under varying compliance rates.

3-14 Different vaccine distribution strategies' sensitivity to the number of vaccine doses available at each time period.

4-1 Flow diagram showing the estimation algorithm process.

4-2 Massachusetts best-fitting expected infection dynamics under static network model (in blue dotted lines with squares) and dynamic network model (in green dotted lines with diamonds). The real estimated infection curve is shown in black solid line with circles.
List of Tables

2.1 Populations of the states as of July, 2009 ........................................... 25
2.2 Data summary of four state’s infection and vaccines administered .......... 27
2.3 Result summary of the best fitting estimated parameters and error measures for the four states. ................................................................. 36

3.1 Summary of the effectiveness of different vaccine distribution strategies. ....... 44
3.2 Summary of the effectiveness of different vaccine distribution strategies, including the cutwidth-based vaccine distribution strategy. ......................... 50

4.1 Summary of Massachusetts parameter estimation results. ......................... 60
Chapter 1

Introduction

In this thesis, we describe a systematic approach to inferring social network characteristics and disease dynamics from real epidemic data. Based on the parameters estimated, we simulate the expected infection dynamics and provide pharmaceutical control methods that effectively halt the spread of the infectious disease. In this chapter we present the motivation for this research, state the problem, discuss the related literature, and provide an overview of the thesis by describing our modeling approach, experimentation plan, and main contributions. Lastly, we describe the organization of the thesis.

1.1 Motivation

According to the report of the U.S. National Intelligence Council’s 2020 Project, “Mapping the Global Future”, a global pandemic is the single most important threat to the global economic order [2]. In early 2003, the UN Food and Agriculture Organization reported that over one-third of the global meat trade was being embargoed as a result of mad cow disease, avian influenza, and other illnesses [2]. In addition to pandemic influenza’s threat to the economy, pandemic influenza has had a devastating impact on the world’s public health. The “Spanish Flu” pandemic of 1918-1919 spread worldwide within a few months, eventually infecting about one-third of the world’s population (about 500 million people), and in six months, some 50 million were dead [3]. There have been about three influenza pandemics in each century for the last 300 years, including the “Asian Flu” of 1957-58, the “Hong Kong Flu” of 1968-69, and the 2009-2010 “Swine Flu” pandemics [4]. While there have been continuous advances in medical and academic fields since the first recorded pandemic influenza event in 1918, the global community still has very little power to stop or mitigate a pandemic’s devastating force. A medical cure for influenza has not been identified. Various preventive measures such as vaccines, school and work closures, and good hygiene practices, and mitigating actions such as anti-viral drugs and self-isolation are the ways we can fight an outbreak [5].
The grave effects of pandemic influenza and the limited resources available to fight against an influenza outbreak motivate us to study and model the outbreak and spread of the disease. The modeling of infectious diseases is a tool which has been used to study the mechanisms by which diseases spread [6, 7, 8]. Infectious diseases spread via person-to-person contacts, therefore social contact networks play an important role in spread of epidemics. The social networks that accurately represent the human interactions will aid to model the infection spread effectively. Effectively modeling disease dynamics allows us to make predictions about diseases and enables us to evaluate various prevention and mitigation plans. This ability could have a significant effect on reducing the infection and mortality rates of a future epidemic. Therefore, it is crucial to model the epidemic dynamics as accurately as possible in order to evaluate and devise the most effective control policy to halt the spread of disease. This thesis will focus on modeling the spread of a specific disease, the 2009-2010 H1N1 influenza, and evaluate vaccination strategies.

Although we focus on epidemiology in this work, understanding spreading processes in complex networks and designing control strategies to contain them are relevant problems in many different areas, such as computer viruses [9, 10], viral marketing [11], and influence and opinion dynamics [12, 13], and modeling techniques and control methods introduced in this work can be applied in different settings.

### 1.2 Problem Statement

Motivated by the threat posed by epidemics, we analyze the problem of understanding the underlying network structure and disease dynamics in order to devise efficient control policies that will minimize the number of people affected.

Despite the vast amount of research on mathematical modeling of disease spread, there is still no widely accepted way to inform decision making in the case of a global catastrophe such as the 2009-2010 H1N1 "Swine Flu" pandemic [7]. Pandemic viruses are usually a mutated variation of an animal strain and are more easily transmitted than regular seasonal flu viruses [14]. The general population usually has little built-up immunity to the novel virus. As a result, a pandemic arises suddenly without much time for preparation, leaving government officials less time for vaccine production. The limited information and time during a pandemic make it difficult to accurately model the spread of influenza and to make effective decisions with regard to vaccine allocation and public policy. These decisions are resource-constrained in one way or another, including personnel, time, money, and available pharmaceuticals. How many vaccines to make, how to allocate and distribute vaccines across different states and cities, when and how to inform the public about the potential threat of the influenza outbreak, whether work and school closures will be necessary and, if so, when and for how long - are just some of the questions that decision makers have to answer.
Because the effectiveness of such decisions is contingent upon a whole set of known and unknown variables such as human behavior and disease characteristics, it is extremely difficult to assess how well alternative decisions and policies will perform, making the decision making process more challenging. Given the tremendous difficulty of and the extreme importance placed upon decision making such as choosing vaccination strategy in the case of a pandemic outbreak, we offer a quantitative approach to realistically model infection dynamics and evaluate different control strategies in order to determine effective strategies.

1.3 Related Literature

1.3.1 Epidemic Models

There are several epidemiological models for the spread of disease. One of the frequently used models is an SIS (susceptible-infected-susceptible) model, where an individual can be in one of the two possible states, susceptible or infected state. Spread of the common cold, which do not confer any long lasting immunity, and computer viruses are often studied using the SIS model. Another model that is often used is an SIR model. In the SIR model, additional state R represents that the individual has recovered from the disease and are permanently immune, so that they can never catch it again or pass it on to others. Studies of epidemic models can be largely divided into two: ones using the mean-field approximation to analytically tract the infection dynamics without networks [15, 16] and ones using the networks, without the approximation, to model the dynamics [17, 18].

Mean-field approximation Pastor-Satorras and Vespignani [16] modeled the propagation of computer viruses using the SIS model on scale-free networks. The study showed that the existence of hubs in the scale-free networks favor the spreading of the virus and lead to the absence of the epidemic threshold for constant curing rate. The absence of the epidemic threshold indicates that infections can proliferate on the scale-free networks, but the small spreading rate explains why statistically only a small percentage of viruses gives rise to a significant outbreak in computer community.

Chakrabarti et al. [15] proposed discrete-time stochastic SIS model to represent the infection spread. To get around the similar limitation of solving the Markov chain with exponentially increasing state space, they used the independence assumption, assuming the infection probabilities are independent of one another. These works used the mean-field approximation to analytically model the disease dynamics without networks.

Without mean-field approximation Berger et al. [18] modeled the SIS disease processes on the scale-free networks and showed how these processes differ from epidemics on other conventional
structure. In particular, they provided the epidemic probability for an infection beginning at a typical starting node as a function of infection rate. Draief et al. [17] used real networks to model the SIR disease dynamics and characterize how the size of the infection depends on the network topology.

1.3.2 Control of Epidemics

Controlling the contagion propagation is critical in order to protect the network from being overwhelmed by diseases and viruses. Main question that arises in controlling the epidemic is: How to distribute the limited resources in order to stop the spread?

Mean-field approximation  Pastor-Satorras and Vespignani [16] showed the absence of epidemic threshold in scale-free networks for constant curing rate and therefore the persistence of the viruses in the networks. Dezso and Barabasi [19] showed that by allowing the cure rate to be non-uniform, i.e. allocating curing resources proportional to the degree, the epidemic threshold can be restored, therefore making possible the eradication of a virus.

Recent studies by Preciado et al. [20, 21] proposed a convex optimization and Geometric Programming to solve optimal resource allocation problem using continuous-time stochastic SIS model. However, when similar approach was taken in our SIR model, due to an additional time-varying decision variable, the disease dynamic constraint turned out to be non-convex, making it hard to obtain the optimal vaccine allocation policy. Therefore, we rely on heuristic algorithms on our work to evaluate effectiveness of vaccine distribution strategies.

Without mean-field approximation  Borgs et al. [22] provided similar results as Dezso and Barabasi [19] that allowing non-uniform cure will eradicate the disease spread on the scale-free networks without using the mean-field approximation.

Drakopoulos et al. [23, 24] proposed dynamic policy based on the cutwidth for the rapid containment of a contagious process modeled as an SIS epidemic. To the best of our knowledge, there hasn’t been work on applying a cutwidth-based heuristic algorithm as a vaccination strategy in SIR models. We show in our simulation that the cutwidth-based vaccination performs better than several other policies previously implemented, such as distributing vaccines based on degree and on the number of infected neighbors.

1.3.3 Network and Propagation Dynamics Inference

The problem of inferring network structure was first studied by Adar & Adamic [25]. The study used Support Vector Machines combined with rich textual features on the data set consisting of thousands of blogs and URLs to predict the occurrence of edges between blogs. Recently, several
studies developed network inference algorithms to estimate the network structures and propagation dynamics. Some approaches infer only the network structure [26, 27], while others infer the transmission probability and propagation rates along with the network structure [28, 29].

Gomez-Rodriguez et al. [26] inferred network connectivity by using submodular optimization, called NETINF. Given the observation of the times when nodes were infected (i.e. blogs posting news articles) from the data set containing information diffusion in a set of millions of blogs and news articles, they found the near-optimal network that explained the observed infection times. Snowsill et al. [27] utilized text contents in addition to the times of the blog posts to determine the connectivity of the nodes (blogs) through detecting text reuses in the posts.

Meyers et al. [28] inferred the diffusion network by learning the infection probability between two nodes using a convex programming, called CONNIE. Du et al. [29] developed a flexible kernel method to infer the network and transmission rates. Both of these works inferred the network structures given the observation of the times when nodes became infected and assumed that the networks are static over time.

Most recent work by Gomez-Rodriguece et al. [30] used stochastic gradient descent method to infer networks and transmission rates, called NETRATE. This work is different than the other previous works in that the underlying networks can change over time and transmission rates are different across different edges.

Although the methods vary between these works, they are similar in that the required inputs to the estimation algorithm must at least have the information on nodes and the times when the nodes became infected/influenced. The main difference in our work is that we infer network structure characteristics and disease propagation dynamics from low-dimensional epidemic data, consisting only of state-wide aggregated temporal infection and vaccination data.

The most relevant work is by Groendyke et al. [31], where they estimated the parameters of a simple random network and a stochastic epidemic on that network using a Bayesian framework. However, this work requires recovery times of infected hosts to estimate the infection rate and the network structure, thus limiting the estimation to the small sample of individuals.

1.4 Technical Approaches

Our ultimate goal is to develop a decision support tool that helps government officials make better decisions on how to allocate the limited vaccine resources in order to achieve the greatest reduction in total infections. In working towards this goal, we employ the following road map:

- We first model the individuals of the U.S. states as nodes and the contacts/relationships among them as edges on a social network.
• We model and simulate the interactions between individuals in order to analyze how these interactions spread an infectious disease.

• We develop a method for parameter estimation to the data collected during the 2009-2010 epidemics from the Centers for Disease and Control in order to estimate the network structure and disease propagation parameters that can simulate the spread of H1N1 infections realistically.

• We evaluate the effectiveness of a vaccine and various vaccine distribution strategies. We propose and evaluate different policies:
  
  - We consider distributing vaccine doses based on each individual’s degree, number of healthy neighbors, and number of infected neighbors.
  
  - We propose a new policy inspired by recent works [23, 24]. The cutwidth of a graph \( G = (V, E) \) is the smallest integer \( k \) such that the vertices \( V(G) = \{v_1, \ldots, v_n\} \) of \( G \) can be arranged in a linear layout \( [v_1, \ldots, v_n] \) in such a way that, for every \( i = 1, \ldots, n-1 \), there are at most \( k \) edges with one endpoint in the set \( S = \{v_1, \ldots, v_i\} \) and the other in the set \( S = V - S = \{v_{i+1}, \ldots, v_n\} \) [32]. Our policy ranks individuals by the resulting cutwidth of the network when each individual is removed. The vaccine doses are then allocated in the order of individuals who will result in the lowest cutwidth when vaccinated. Note that this policy is different from the SIS epidemic curing policy in [23, 24] where all curing resources are allocated to the first infected node from the linear arrangement of nodes that generates the cutwidth.

  The purpose of this strategy is to vaccinate individuals to achieve the minimum possible infection-transmitting paths or contacts between infected and susceptible individuals in the network, reducing the infection spread as a result. We employ an integer programming formulation to find this ranking of susceptible individuals to receive the limited vaccine doses available in order to minimize the total number of infections.

1.5 Experimentation

For our experimentation, we first apply the parameter estimation algorithm to estimate the network structure and disease propagation parameters that simulate the dynamics of the 2009-2010 H1N1 influenza spread in the states of Massachusetts, New York, North Dakota, and Montana, based on the real data gathered at that time. We then analyze how different vaccine distribution strategies reduce the number of infections by simulating the expected infection dynamics on the scaled-down population of Massachusetts. Finally, we apply the parameter estimation method to the dynamic network model to test if using a more realistic network model will indeed result in a better approximation of the disease spread.
1.6 Contributions

This thesis makes three main contributions. First, it introduces a methodology for determining network structure and disease propagation parameters that can simulate the realistic spread of an infectious disease. Second, it presents heuristic and optimization-based vaccine allocation strategies, which effectively reduce the infections by the efficient utilization of limited resources. Lastly, it provides a method of modifying a static network model to reflect the dynamic nature of human contacts. Our main experimental result is that the proposed vaccine allocation strategies achieved significant reductions in the population affected by the infection, even under lower compliance rates.

1.7 Thesis Organization

In Chapter 2, we introduce a discrete-time stochastic SIR (Susceptible-Infected-Removed) model, which enables us to simulate interactions between individuals and examine how those interactions spread an infectious disease across a social network. We also describe the parameter estimation method that estimates social network structure and propagation dynamics of the 2009-2010 H1N1 influenza epidemic, based on the data collected during that time. Applying the parameter estimation algorithm to several U.S. states, we present the network structures and propagation dynamics that realistically simulate the spread of H1N1 influenza.

In Chapter 3, we analyze the effectiveness of the vaccine and different vaccine distribution strategies using the network structures and propagation dynamics parameters obtained. We present a detailed description and comparison of the vaccine allocation strategies we formulated. We also analyze how effective the suggested strategies are in mitigating the infection's spread under varying compliance rates and vaccine resource availabilities.

In Chapter 4, we describe a method to incorporate the dynamical nature of human interactions with a simple modification to the static network model. We apply the parameter estimation algorithm to the dynamic network model and analyze the ability of the model to realistically simulate the propagation of H1N1 influenza, compared to the estimation results using the static network model in Chapter 2.

In Chapter 5, we discuss our conclusions and provide recommendations for future research.
Chapter 2

Estimating Social Network Structure and Propagation Dynamics

2.1 Introduction

Diseases spread through human populations via contact between infected individuals (those carrying the disease) and susceptible individuals (those who do not have the disease yet, but can potentially catch it) [33]. These contacts form a social network, along which the disease is transmitted. Therefore, it has long been recognized that the structure of social networks plays an important role in understanding and analyzing the dynamics of disease propagation [34]. In this chapter, we present a methodology to estimate the structure of the underlying social network and the dynamics of an infectious disease. Better understanding of the social network and transmission parameters will help public officials devise better strategies to prevent the spread of disease.

Many previous studies of disease propagation assume that populations are “fully mixed,” meaning that an infective individual is equally likely to spread the disease to any other susceptible member of the population to which he belongs [8]. Larson et al. enriched the aforementioned models by incorporating different types of agents [35]. In these works, the assumption of “full mixing” allows one to generate nonlinear differential equations to approximate the number of infective individuals as a function of time, from which the behavior of the epidemic can be studied. However, this assumption is clearly unrealistic, since most individuals have contact with only a small fraction of the total population in the real world.

Building on this insight, a number of authors have pursued theoretical work considering network implications. These models replace the “fully mixed” assumption of differential equation-based models with a population in which contacts are defined by a social network [34, 36, 37, 38, 39, 40]. However, there is often limited data available to estimate the underlying network. One major strand
of work [41, 25, 26, 27, 28, 29] focused on estimating the network from the available data. These studies predicted the occurrence of edges between nodes given the infection times of the nodes. However, application of these studies was limited to the estimating networks of internet websites primarily due to the abundance of data in the area. In cases of influenza infections, individual-specific information such as infection times is often unavailable. We propose a method to infer the network characteristics and infection dynamics parameters from much lower-dimensional data, consisting only of state-wide aggregated temporal infection and vaccination data. More relevant work is by Groendyke et al. [30], where they estimated the parameters of a simple random network and infection dynamics using a Bayesian framework. However, this work requires recovery times of infected hosts to estimate the infection rate and the network structure, thus limiting the estimation to the small sample of individuals.

Another strand of work has employed large-scale experiments to map real networks by using various sources of data such as email address books, censuses, surveys, and commercial databases. However, this often requires extensive time and resources collecting, manipulating, and combining multiple data sources to capture large networks and estimate connections within those networks [9, 34, 42, 43, 44, 45]. In this work, we use much lower-dimensional data to infer the network characteristics, assuming that the network follows two of the widely studied network models, the scale-free or small-world models. In Chapter 4, we will also consider a dynamic network model, based on the idea that people often make different random contacts day-to-day in addition to more consistent contacts with close acquaintances.

The contribution of this chapter is the following: we develop a method to extract the network structure from the observed infection data. Specifically, our approach assumes a parameterized network model and disease process parameters to simulate an expected infection curve. Then, the algorithm greedily searches for the parameter values that will generate an expected infection curve that best fits the estimated real infection curve. We demonstrate that our suggested algorithm, assuming a scale-free network, closely estimates the network characteristics and disease dynamic parameters for the 2009 H1N1 influenza pandemic. Our results confirm that the network-based model performs better in estimating the propagation dynamics for an infectious disease compared to the differential equation-based models with the "fully mixed" assumption.

2.2 Methods

In this section, we describe a discrete-time stochastic multi-agent SIR (Susceptible-Infected-Removed) model, and propose a corresponding inference algorithm to fit the disease dynamics generated by the model to real H1N1 infection data. The inference algorithm learns the social network structure and the key disease spread parameters, such as the rate of infection and the rate of recovery, for
a given infectious disease. This enables us to make useful predictions about the contact network structure and disease propagation for similar types of diseases and allows us to devise appropriate control strategies. $\frac{N(t)}{H} \times \text{number of H1N1 cases}$

2.2.1 Data

We obtained data from state health departments, including the weekly percentage of all hospitalizations and outpatient visits resulting from influenza-like illness (%ILI; ILI is defined as fever [temperature of 100°F (37.8°C) or greater] and a cough and/or a sore throat without a known cause other than influenza) and the weekly number of vaccine doses administered over the 2009-2010 flu seasons [46, 47, 48]. Each point on the %ILI temporal curve represents the percentage of the total number of hospitalizations and outpatient visits that are specific to H1N1, defined as $F(t)$. We defined the number of H1N1-related hospitalizations and outpatient visits at each time as $N(t)$. We also obtained total estimated cases of H1N1, defined as $M$, and total estimated H1N1 related hospitalizations defined as $H$, from the Centers for Disease Control [47]. Since only the aggregated data for the United States was provided, we assumed that the same percentage of population was affected in individual states that we analyze in this paper. Using these data and the following assumptions, we estimated the number of H1N1 infections for each week, which we define as $I(t)$, as follows:

Assuming that the flu wave first grows, then declines after the peak of the infection, while the number of non-H1N1 hospitalizations remains relatively stable, we can estimate the number of non-H1N1 hospitalizations. Finally, the above allows us to estimate the number of H1N1-related hospitalizations during each period.

We define non-H1N1 related hospitalizations at week $t$ as

$$C(t) = C_0$$  \hspace{1cm} (2.1)

Total estimated H1N1 related hospitalizations and visits is represented as

$$H = \sum_{t=1}^{t=\text{max}} N(t)$$  \hspace{1cm} (2.2)

%ILI at week $t$ can be represented as the number of H1N1 related hospitalizations and visits divided by all hospitalizations

$$F(t) = \frac{N(t)}{N(t) + C_0}$$  \hspace{1cm} (2.3)

Solving for $N(t)$ in (2.3), we get

$$N(t) = \frac{C_0 F(t)}{1 - F(t)}$$  \hspace{1cm} (2.4)
Figure 2-1: The infection curve estimated from %ILI data and the number of vaccines administered during the observation period (October, 2009 - December, 2009) for Massachusetts.

Assuming that the cases of H1N1 infections, $I(t)$, are proportional to the number of H1N1 related hospitalizations and visits [35], $N(t)$, we have

$$I(t) = \frac{N(t)}{H} \times M$$  \hspace{1cm} (2.5)

We used the estimation method described above to estimate the infection curve, the plot of $I(t)$ over time, for the state of Massachusetts, in which the estimated true infection curve includes the effects of vaccines administered. Figure 1 shows the estimated temporal infection curve and the temporal curve of vaccines administered [49, 50].

Figures 2, 3, and 4 show the estimated true infection curves and the temporal curves of vaccines as administered for the states of New York, Montana, and North Dakota, respectively [51, 52, 53]. We have selected states where the data for vaccines administered were available since not all of the states had the estimated number of vaccines administered. Note that there were more vaccines available for the states, but the data we show reflect the number of doses actually administered. Relatively low proportion of the people in New York chose to get vaccinated, and this can be due to the various reasons such as busy work schedule, belief that the vaccination will not work effectively, etc. Not all states received vaccine doses in timely manner, for example, Montana received the vaccine as late as a month into the infection period. Table 1 summarizes the populations of the analyzed states according to the United States Census Bureau as of July, 2009 [54]. We assume a static network in this analysis where the size of the network stays constant over time. Table
New York: # Infected and # Vaccine Doses Administered

Figure 2-2: The infection curve estimated from %ILI data and the number of vaccines administered during the observation period (October, 2009 - December, 2009) for New York State.

Table 2.1: Populations of the states as of July, 2009

<table>
<thead>
<tr>
<th>States</th>
<th>Population (As of July, 2009)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Massachusetts</td>
<td>6,593,587</td>
</tr>
<tr>
<td>New York</td>
<td>19,541,453</td>
</tr>
<tr>
<td>Montana</td>
<td>974,989</td>
</tr>
<tr>
<td>North Dakota</td>
<td>646,844</td>
</tr>
</tbody>
</table>

2 summarizes all four states' infection and vaccines administered data. Note that the observation period differs between the states as Montana and North Dakota were affected by the outbreak earlier than Massachusetts and New York. For Massachusetts and New York, around 11% of each state's total population was infected during the observation period. For Montana and North Dakota, about 17% and 13% of each state's total population was infected. During each state's observation period, about 19%, 5%, 10%, and 19% of each state's total population was vaccinated for Massachusetts, New York, Montana, and North Dakota, respectively.

2.2.2 Disease Dynamics

We employ a variation of the susceptible-infective-removed (SIR) model first proposed by Kermack and McKendrick [55]. Individuals in the network, represented by nodes, are assigned one of the three states: the susceptible state (S) in which individuals are not infected but could become infected, the infected state (I) in which individuals are currently carrying the disease and can spread the disease
Figure 2-3: The infection curve estimated from %ILI data and the number of vaccines administered during the observation period (September, 2009 - December, 2009) for Montana.

Figure 2-4: The infection curve estimated from %ILI data and the number of vaccines administered during the observation period (September, 2009 - December, 2009) for North Dakota.
# Table 2.2: Data summary of four state’s infection and vaccines administered

<table>
<thead>
<tr>
<th>Week</th>
<th>Massachusetts</th>
<th>New York</th>
<th>Montana</th>
<th>North Dakota</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Infected</td>
<td>Vaccinated</td>
<td>Infected</td>
<td>Vaccinated</td>
</tr>
<tr>
<td>35</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>36</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>37</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>38</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>39</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>40</td>
<td>-</td>
<td>-</td>
<td>69,425</td>
<td>502</td>
</tr>
<tr>
<td>41</td>
<td>24,272</td>
<td>42,874</td>
<td>154,944</td>
<td>9,951</td>
</tr>
<tr>
<td>42</td>
<td>41,422</td>
<td>69,534</td>
<td>331,291</td>
<td>43,510</td>
</tr>
<tr>
<td>43</td>
<td>78,564</td>
<td>88,150</td>
<td>428,078</td>
<td>58,656</td>
</tr>
<tr>
<td>44</td>
<td>100,364</td>
<td>114,824</td>
<td>384,233</td>
<td>113,058</td>
</tr>
<tr>
<td>45</td>
<td>153,057</td>
<td>150,866</td>
<td>273,909</td>
<td>119,539</td>
</tr>
<tr>
<td>46</td>
<td>126,282</td>
<td>147,517</td>
<td>188,882</td>
<td>175,263</td>
</tr>
<tr>
<td>47</td>
<td>70,337</td>
<td>88,522</td>
<td>112,497</td>
<td>81,006</td>
</tr>
<tr>
<td>48</td>
<td>52,539</td>
<td>148,170</td>
<td>73,139</td>
<td>112,373</td>
</tr>
<tr>
<td>49</td>
<td>30,464</td>
<td>162,348</td>
<td>66,644</td>
<td>115,793</td>
</tr>
<tr>
<td>50</td>
<td>22,732</td>
<td>150,856</td>
<td>40,850</td>
<td>149,621</td>
</tr>
<tr>
<td>51</td>
<td>16,602</td>
<td>92,155</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Let $G = (V, E)$ denote an undirected graph with $n$ nodes, $m$ edges, and no self-loops. We denote by $V(G) = \{v_1, \ldots, v_n\}$ the set of nodes and by $E(G) \subseteq V(G) \times V(G)$ the set of undirected edges of $G$. If $\{v_i, v_j\} \in E(G)$, we call nodes $i$ and $j$ adjacent (or neighbors). We define the set of neighbors of node $v_i \in V$ as $\mathcal{N}_i = \{v_j \in V(G) : \{v_i, v_j\} \in E(G)\}$. The number of neighbors of $i$ is called the degree of node $i$. The adjacency matrix of an undirected graph $G$, denoted by $A = [a_{ij}]$, is an $n \times n$ symmetric matrix defined entry-wise as $a_{ij} = 1$ if nodes $i$ and $j$ are adjacent, and $a_{ij} = 0$ otherwise. We define $X_i(t) \in \{S, I, R\}$ to be the state of individual $i \in V$ at time $t$. We let

$$\eta_i(t) = \sum_{(j,i) \in E} 1_j$$

(2.6)

denote the number of infected neighbors the individuals $i$ has at time $t$, and $1_j$ is the indicator function defined as

27
Then, given that the individual $i$ is susceptible at time $t$, he will become infected at time $t+1$ with the following probability:

$$P(X_i(t+1) = I \mid X_i(t) = S, X(t)) = 1 - (1 - \beta)^{n_i(t)}$$

(2.8)

since with probability $(1 - \beta)^{n_i(t)}$ all infection attempts fail, assuming independence in infection attempts between neighbors. This is a commonly used assumption in simulation of spreading processes in complex networks [44, 45, 15].

Given that the individual $i$ is infected at time $t$, he will recover at time $t+1$ with the probability:

$$P(X_i(t+1) = R \mid X_i(t) = I, X(t)) = \delta$$

(2.9)

The above two stochastic transitions describe the disease dynamics. Given a network and set of initial infections, the disease propagation process can be simulated according to the described transition probabilities.

### 2.2.3 Estimation Algorithm

The estimation algorithm we introduce in this section uses the stochastic disease dynamic process above to simulate infections. The simulated results are compared to the real H1N1 infection data, and we optimize over the network and disease spread parameters to obtain a best-fit simulated curve. The purpose of the algorithm is to find network characteristics, such as degree distribution $k$ and disease spread parameters $\beta$ and $\delta$ that will help us make useful predictions about the network and how the disease spreads within the community.

Many real-world social networks, such as citation networks, online discussion groups, internet and router topologies, and sexual contact networks, are expected to exhibit small-world or scale-free properties [9, 40, 42, 43, 56, 57, 58, 59]. Because these models tend to represent many different types of social interactions, we tested both small-world and scale-free networks in our algorithm for the contact network as a reasonable starting point. The following subsections present background on two network models and indicate how we parameterize them to simulate the network generation in the estimation algorithm to be described.
2.2.3.1 Small-World Network

A small-world network is a graph in which most nodes are not neighbors of one another but also most nodes can be reached from every other node by a small number of steps [60]. Therefore, the small-world network has strong local clustering but short average path lengths. The length of a path, or a path length, is the number of edges traversed along the path from a starting node to a target node [58]. We can generate the small-world network with two parameters: average degree \( k \) and short-cut probability \( p \). We first create a ring lattice, where we place all the nodes in a ring and place edges between each node and \( k/2 \) neighbors on each side of the selected node. The right lattice will look like the one on the left side of Figure 5. After the ring lattice is created with average degree parameter \( k \), each and every edge will be visited and according to the short-cut probability \( p \), one end of the edge will be disconnected from the currently connected node and reconnected to a randomly selected node. The reconnecting node will be selected uniformly at random. Due to the short-cut connections, the resulting network now exhibits a short average path-length property. The network on the right side of Figure 5 represents the resulting small-world network after the long-range connections are established according to the short-cut probability \( p \).

2.2.3.2 Scale-Free Network

A scale-free network is a network whose degree distribution follows a power law, at least asymptotically. That is, the fraction, \( P(k) \) of nodes in the network having \( k \) connections to other nodes is given for large values of \( k \) by \( P(k) \sim k^{-\gamma} \), where \( \gamma \) is a parameter whose value is typically between 2 and 3, although occasionally it may be outside of the bounds [61, 62]. The most notable characteristic in a scale-free network is the existence of nodes with a degree that greatly exceeds the average degree.
of the network. A few very high degree nodes, often called “hubs”, are connected to much of the population, while most nodes have only a few connections.

We use an algorithm that adapts a preferential attachment mechanism invented by Barabasi and Albert [63]. Given the value of the average degree parameter $k$, we first generate a complete graph with $k + 1$ nodes, where each of the $k + 1$ nodes are connected to the rest of the $k$ nodes. We then bring the new nodes into the existing complete network, one by one, and each node connects to $k$ of the existing nodes with a probability that is proportional to the number of connections that the existing nodes already have. Formally, the probability $p_i$ that the new node is connected to existing node $i$ is

$$p_i = \frac{d_i}{\sum_j d_j}$$

where $d_i$ is the degree of node $i$ and the sum is made over all pre-existing nodes $j$. A scale-free network with $N$ nodes and average degree $k$ is then created once the $N$th node comes into the pre-existing network of size $N - 1$ and connects to $k$ nodes. Therefore, following this preferential attachment mechanism, existing nodes with high degrees tend to quickly accumulate more connections, while the nodes with small connections are unlikely to be chosen as the destination node for a new link, resulting in a power law distribution.

### 2.2.3.3 Estimation Algorithm

The following describes the estimation algorithm:

**Input**

The inputs to the algorithm include:

- A parameterized disease spread network structure, where nodes represent people and undirected edges represent contacts between people. In this chapter’s simulations, the network structure is assumed to be either scale-free or small-world, though the algorithm could be applied to other network structures. We will simulate other network models in Chapter 4.

- Initial values of the network parameters. For example, for the scale-free network, the average degree, $k_0$, and for the small-world network structure, the average degree, $k_0$, and the probability of a long-range contact, $p_0$.

- Initial values of the disease process parameters $\beta_0$ and $\delta_0$.

- Real temporal infection data to fit the model-generated expected infection dynamics.

- Data on vaccines administered (if administered).
Output

- The algorithm outputs network and infection parameters that are used to generate a simulated expected infection curve that fits the real data as closely as possible.

Procedure

Begin with the given initial values of the social network and disease spread parameters: $k_0$, $p_0$, $\beta_0$, and $\delta_0$. Let $\Delta k$, $\Delta p$, $\Delta \beta$, and $\Delta \delta$ be the amounts by which $k$, $p$, $\beta$, and $\delta$ are changed at each step in the optimization. Let $\epsilon_t$ and $\bar{\epsilon}_t$ each denote the number of infections for the real infection curve and the estimated expected infection curve at time $t$, respectively. We define the error, $\bar{E}$, between the simulated expected infection curve and the true infection curve as:

$$
\bar{E} = \sum_{t=1}^{t=\text{max. period}} |\epsilon_t - \bar{\epsilon}_t|
$$

(2.11)

Alternatively, we can look at least squares, but we use the absolute sum of the residuals defined above since the it provides a unit that is easy to comprehend.

Repeat the following steps until the error can no longer be reduced by changes to the parameters (we define the optimal output parameters as $k^*$, $p^*$, $\beta^*$, and $\delta^*$):

1. Given $k_0$, $p_0$, $\beta_0$, and $\delta_0$, search in all possible directions to find a direction that improves $\bar{E}$. That is, evaluate $\bar{E}$ at all possible combinations of $k$, $p$, $\beta$, and $\delta$, where $k \in \{k_0, k_0 + \Delta k, k_0 - \Delta k\}$, $p \in \{p_0, p_0 + \Delta p, p_0 - \Delta p\}$, $\beta \in \{\beta_0, \beta_0 + \Delta \beta, \beta_0 - \Delta \beta\}$, and $\delta \in \{\delta_0, \delta_0 + \Delta \delta, \delta_0 - \Delta \delta\}$. Evaluate $\bar{E}$ by doing the following for each set of parameters:

   (a) Generate the network according to the given network type (small-world or scale-free) and network characteristics ($k$, the average degree for a scale-free network; $\bar{k}$, the average degree and $p$, the short-cut probability for a small-world network).

   (b) Simulate $R$ realizations of the disease process. For each realization, initialize the disease simulation infection by assigning $N_1$ nodes to the infected states, where $N_1$ is the number of people infected at the beginning of the observation period in the data. We assume that the initial infected nodes are selected uniformly at random from among all the nodes. This is a commonly used assumption in epidemic simulation [45]. Update the disease states for each time period, according to the disease process parameters and the vaccine administration data. (We assume that those who receive vaccines in each time period are chosen uniformly at random.)

   (c) Generate an expected infection curve by averaging the number of infected individuals at each time period over the $R$ simulated realizations of the disease process.

   (d) Calculate $\bar{E}$. 

31
2. Determine which search direction resulted in the minimum error, that is, the values of $k$, $p$, $\beta$, and $\delta$ that achieved the lowest error. Update $k_0$, $p_0$, $\beta_0$, and $\delta_0$ to the values of $k$, $p$, $\beta$, and $\delta$.

The algorithm is summarized as a flow chart in Figure 6.

2.3 Results

We have applied our algorithm to data from the 2009 H1N1 and obtained network and infection parameters that approximate the dynamics of an infectious disease. We scaled down the population size by a factor of 10,000 uniformly at random in order to reduce the computation time for infection simulation. In our simulations, we set $R$, the number of realizations per set of parameters, equal to 50. This affects how well the simulated curve approximates the expected curve. We began our search with relatively large values of $\Delta k$, $\Delta p$, $\Delta \beta$, and $\Delta \delta$ (changes in average degree, long-range connection (short-cut) probability, infection probability, and recovery probability, respectively) and then manually decreased them as the error began to converge. Specifically, initially $\Delta k = 1$, $\Delta p$, $\Delta \beta$, $\Delta \delta = 0.1$. We narrowed the search by reducing $\Delta p$, $\Delta \beta$, and $\Delta \delta$ to 0.01. We implemented the algorithm in MATLAB, Version R2013a.

2.3.1 Estimation Algorithm Results

Figure 7 shows the resulting infection curves generated by the algorithm, compared to the estimated infection curve from data and from using differential equations with “fully-mixed” assumption for the state of Massachusetts. In addition to the error measure described in equation (10), we used the difference in total expected number of infections,

$$\tilde{E}_{Total} = \left| \sum_{t=1}^{t=\text{max. period}} e_t - \sum_{t=1}^{t=\text{max. period}} e_t \right|$$

(2.12)
and the difference in peak number of infections to compare the curves:

$$E_{Peak} = |\hat{\epsilon}_w - \epsilon_w|$$  \hspace{1cm} (2.13)

where \( w \) represents the time period of infection peak.

For the small-world network, the estimated parameter values were 10 for average degree \((k)\), 0 for short-cut probability \((p)\), .21 for infection probability \((\beta)\), and .89 for recovery probability \((\delta)\). The error measured \((\hat{E})\) was close to 21 infections, which is 28% of total number of infections. Compared with the data-generated infection curve, the simulated expected infection curve for the small-world network had 1% lower expected total number of infections and a 30% lower expected peak infections. Overall, the small-world network model did not provide a good fit to the estimated infection curve from data.

On the other hand, the best-fit infection curve generated using a scale-free network fits the data well. The estimated parameter values were 2 for average degree \((k)\), .45 for infection probability \((\beta)\), and .69 for recovery probability \((\delta)\). For preliminary face validation on the recovery probability, two clinical studies on small samples of H1N1 patients have revealed that 38% and 76% of the sampled patients recovered from H1N1 within a week respectively [64, 65]. Given that it is hard to come across large sample data on H1N1 patients’ recovery times to accurately validate the estimated recovery probability, we used small sample studies for preliminary check on the estimated recovery probability. Measured error was about 11 infections, constituting 16% of the total infections. Compared to the estimated real infection curve from data, we measured a 1.9% difference in the expected total number of infections and a 2.3% difference in the expected peak infections. This result is a significant improvement over the estimation under the “fully-mixed” assumption, which had a measured error of 26 infections (36% of total infections), a 25.7% difference in the expected total number of infections, and a 10% difference in the expected peak infections. Figure 7 below compares the infection curves generated from real data to the expected infection curves generated using “fully-mixed” assumption, estimated small-world network, and estimated scale-free network.

Figures 8, 9, and 10 compare the resulting infection curves for the state of New York, Montana, and North Dakota, respectively. Table 3 the figures summarizes the best fitting estimated parameters and error measures for the states that we analyzed. Error measures are represented as a percentage of total real infections (for \(\hat{E}\) and \(\hat{E}_{Total}\)) or real peak infections (for \(\hat{E}_{Peak}\)), respectively. Resulting parameters varies from state to state, however, scale-free network tend to fit the estimated real-data well compared to small-world network or differential equations with “fully-mixed” assumption generated curve.

33
Figure 2-7: Massachusetts best fit curves generated by the algorithm using small-world network (green) and scale-free network (blue) compared to the estimated infection curve from data (black) and from using differential equations with “fully-mixed” assumption (red).

Figure 2-8: New York best fit curves generated by the algorithm using small-world network (green) and scale-free network (blue) compared to the estimated infection curve from data (black).
Figure 2-9: Montana best fit curves generated by the algorithm using small-world network (green) and scale-free network (blue) compared to the estimated infection curve from data (black).

Figure 2-10: North Dakota best fit curves generated by the algorithm using small-world network (green) and scale-free network (blue) compared to the estimated infection curve from data (black).
Table 2.3: Result summary of the best fitting estimated parameters and error measures for the four states.

2.3.2 Discussion

Understanding the network structure and the disease dynamics on the network has important implications both for refining epidemic models and for devising the control measures necessary in order to effectively utilize resources to prevent the spread. The spread of infection is often complex to analyze due to the lack of information about the contact network on which it occurs. This chapter showcases a methodology to learn network and disease propagation parameters of an infectious disease, HINI influenza. The findings for H1N1 give us useful insight into the infection dynamics of similar diseases and assist in analyzing the effects of different vaccination policies. In Chapter 3, we will use the results in this chapter to analyze the effects of different vaccination policies and suggest effective control policies in order to help prevent infection and minimize the number of people affected by the influenza.
Chapter 3

Effective Control Policy

3.1 Introduction

Epidemic outbreaks such as H1N1 influenza showcase how susceptible large communities are to the spread of such infections. The occurrence of such pandemic outbreaks raises the question of intervention strategies that can mitigate the effects of the disease on the population. There are two main categories of interventions: pharmaceutical and non-pharmaceutical. For pharmaceutical intervention to H1N1, antiviral medication such as Oseltamivir (Tamiflu) can be used for treatment purposes, if taken within the first 48 hours of occurrence of influenza-like symptoms. Some of this medication has been shown to reduce the complications and symptoms and shorten illness by one to two days [66]. Antiviral medication is generally not used for preventive purposes for anyone but the medical professionals who are likely to come across many infectious individuals in the course of their daily activities. The pharmaceutical intervention that is most used for massive distribution is vaccines. These are also not 100% effective, but they significantly lower the susceptibility and infectivity of a person. Non-pharmaceutical interventions include a person’s behavioral changes to help lower his or her chances of contracting influenza and to prevent the spread of the virus to others. There are a number of studies analyzing the impact of non-pharmaceutical interventions on the spread of influenza-like illnesses [5, 67, 19] - for example, how implementing human behaviors such as covering coughs, using hand sanitizers, and self-isolation by telecommuting to work will affect the epidemics of infectious diseases.

In this work, we focus on the impact of pharmaceutical interventions, specifically vaccines. To devise effective pharmaceutical control policy, we need to be able to evaluate the effect of vaccines and different vaccine allocation strategies. Good understanding of the contact networks on which the infection spreads and of the propagation dynamics of the disease are necessary to evaluate the effect of vaccination strategies. In Chapter 2, we have successfully learned the contact network and the
disease dynamics parameters. Given the latter, we can use the model to study the effect of different control methods. In this chapter, we will use heuristics to analyze different vaccine distribution policies and suggest efficient vaccine allocation strategies.

3.2 Evaluating Effective Vaccine Allocation Strategies

We showed the data on the number of vaccine doses administered at each time period in Chapter 2. Note that there were more doses available, but the number administered reflects the doses actually used at each time period. Therefore, we use this number as the constraint on the vaccine intervention resources. We will use simulations to evaluate the effectiveness of different vaccination policies as well as studying their performance in terms of total number of people infected by the influenza. After the effective vaccine allocation strategies are identified, we analyze their efficacies for different compliance rates. We also analyze different allocation strategies' sensitivity to the number of doses available at each time period.

3.2.1 Methods

In Chapter 2, the parameter estimation method indicated a scale-free network for the 2009 H1N1 influenza outbreak. We use the disease process described in Chapter 2 with the estimated network parameters and disease dynamics parameters to simulate the expected infection curve under different vaccine allocation policies. We assume that the baseline vaccination strategy, the strategy that was used during the 2009 H1N1 outbreak, is to vaccinate individuals uniformly at random among the susceptible individuals at each time period. The number of individuals vaccinated at each time period is constrained by the number of doses administered according to the data. The following summarizes the simulation input, output, and procedure.

Input

The inputs to the algorithm include:

- The underlying network structure, where nodes represent people and undirected edges represent contacts between people. We use a scale-free network, built using the preferential attachment mechanism as explained in 2.3.2. It is possible to use other parameterized network structures such as a small-world network.

- The best-fitting network parameter, the average degree, \( k^* \), obtained from the estimation algorithm in 2.3.

- The best-fitting disease process parameters \( \beta^* \) and \( \delta^* \) obtained from the estimation algorithm.

- Data on vaccine doses administered (if administered).
A selected vaccine distribution strategy.

**Output**

- The algorithm outputs a simulated expected infection curve under the specified vaccine allocation scheme.

**Procedure**

1. Generate the network according to the given network type and network characteristics ($k^*$, the average degree for a scale-free network; $k^*$, the average degree and $p^*$, the short-cut probability for a small-world network).

2. Simulate $R$ realizations of the disease process. For each realization, initialize the disease simulation infection by assigning $N_1$ nodes to the infected states, where $N_1$ is the number of people infected at the beginning of the observation period in the data. We assume that the initial infected nodes are selected uniformly at random from among all the nodes. This is a commonly used assumption in epidemic simulation [68]. Update the disease states for each time period, according to the disease process parameters and the vaccine administration data. (Those who receive vaccine doses at each time period are chosen according to the specified distribution strategy, and once a susceptible individual is vaccinated, he or she will immediately become immune to the disease and the individual’s state will change to the removed state.)

3. Generate an expected infection curve by averaging the number of infected individuals at each time period over the $R$ simulated realizations of the disease process.

**3.2.2 Results**

We conducted the simulations in MATLAB. We used the parameters found for the state of Massachusetts, which were 2 for the average degree, .45 for infection probability, and .69 for recovery probability. We reduced the size of the population by a factor of 1000 in order to reduce the computation load. A scale-free network of 6594 nodes was used.

**3.2.2.1 Vaccine Effectiveness**

We used the simulation method described above to evaluate different vaccination strategies. First, we simulated the disease-spreading process assuming no vaccine is distributed so that we could evaluate the effectiveness of vaccination when we compare the result with the disease-spreading process without vaccination. Subsequently, we simulated the disease dynamics according to the current vaccine distribution scheme, which is to distribute vaccine in a random manner. Fifty instances of disease dynamics were simulated, and these instances were averaged to obtain an expected infection...
Effectiveness of Vaccines Under Random Distribution

Figure 3-1: Plot of the expected infection curve for Massachusetts assuming no vaccine is available (black), compared to the expected infection when vaccine is randomly distributed (red).

For the state of Massachusetts, the simulation results in Figure 1 show that by randomly distributing doses, the total expected number of infections is reduced by 15%.

3.2.2.2 Effective Vaccine Allocation Strategies

We tested the following four vaccine distribution strategies:

- Distribute vaccine based on each individual’s number of connections (degree).

- Distribute vaccine based on each individual’s number of healthy neighbors at the time of vaccination.

- Distribute vaccine based on each individual’s number of infected neighbors at the time of vaccination.

- Distribute vaccine based on a cutwidth inspired ranking

Note that the last three policies are state-dependent strategies, requiring the knowledge of the states of all individuals in the network at the time of vaccination.

Vaccine Distribution Based on the Degree From Chapter 2, we determined that the scale-free network well represents the contact network on which the H1N1 spreads. This knowledge of the network structure helps to devise efficient vaccine allocation strategies. Many studies have shown that, since scale-free networks have some nodes with a very large number of connections compared
to the average degree, removing these nodes will effectively halt the dynamics of spreading processes [69, 70, 71]. Indeed, the estimated scale-free network for Massachusetts (scaled down by a factor of 1000), which was generated using the principle of preferential attachment, in our case has a few nodes with connections over 100 times the average connections. Figure 2 shows the degree distribution (top) and probability plot of the degree (bottom) for Massachusetts’ estimated scale-free network. Over 80% of the population have degree less than or equal to the average degree of 2, while 0.01% of the population have over 25 connections. These well-connected individuals are more likely to contract infections, and once they are infected, they can rapidly spread infections in the population. Therefore, if these popular individuals are vaccinated and don’t contract infections, infections cannot spread as rapidly, thus greatly slowing the spread of H1N1 influenza.

We implemented this strategy of distributing doses according to the number of individuals’ connections and simulated the infection dynamics for the state of Massachusetts. At each time period, the network and the infection states were analyzed, and $v(t)$ most connected susceptible individuals were selected to get vaccinated, where $v(t)$ represents the number of doses available at
Expected Infections Curves Under Different Vaccination Schemes

Figure 3-3: Plot of expected infection curves for Massachusetts. Infection dynamics are simulated and compared under three cases: no vaccine distributed (black), vaccine randomly distributed (red), and distributed based on individual’s degree (blue).

time period \( t \). The number of doses available at each time period remains the same, constrained by the number of doses distributed according to the data. Figure 3 shows the resulting expected infection curve under this vaccine allocation scheme, compared to the infection curves under the random vaccine distribution case and under the case assuming no vaccine is available. When doses are distributed according to the individual’s number of contacts, the total expected number of infections was reduced by 92% compared to when there were no vaccination efforts. This allocation strategy reduces the total expected infections by about six times more when compared to the random allocation strategy.

Vaccine Distribution Based on the Number of Healthy Neighbors

Based on the previous result (that the vaccine distribution according to the individual’s degree on the scale-free network effectively reduces the total number of infections as much as six times compared to the random distribution strategy), we can modify the strategy to develop different, and possibly more efficient, strategies. One such modification of the degree-based vaccine distribution strategy is to distribute doses according to the number of healthy neighbors. The degree-based vaccination can lose its effectiveness when a susceptible-highly-connected individual to be vaccinated is already surrounded by infected neighbors. In this situation, vaccinating that individual will only benefit the person and will not have any preventive impact on the further infections since the person does not have any susceptible neighbors to transmit the infection. Therefore, vaccinating the individuals with high
numbers of healthy neighbors may effectively slow the spread of the infections. Figure 4 compares the resulting expected infection curve under this allocation strategy to the expected infection curves under previously mentioned cases. The total expected number of infections was reduced by 91% compared to when there was no vaccine distributed. In this specific example, the allocation strategy resulted in similar effectiveness as distributing vaccine based on the individual’s degree. Since the total expected infections under these two strategies only accounted for about 1% of the total population, the special case of a susceptible-highly-connected individual being surrounded by infected individuals is unlikely to happen, thus causing the two strategies to behave similarly.

**Vaccine Distribution Based on the Number of Infected Neighbors** Another modification of the degree-based distribution strategy is to allocate doses according to each individual’s number of infected neighbors. Those surrounded by many infected neighbors are likely to get infected and further spread infections to the susceptible neighbors. By allocating vaccine to those who have many infected neighbors, we effectively block the likely passages of the infection spread, reducing the propagation of the influenza. Figure 5 below compares the expected infection dynamics under this strategy to the expected infection curves of previously discussed cases. The total expected infections were reduced by 97% compared to when there was no vaccine distributed. This allocation strategy reduces the total expected infections by 6.6 times when compared to the random allocation strategy, showing an improvement from the strategies distributing vaccine based on the degree and on the
Figure 3-5: Comparison of expected infection curves under different cases: no vaccine distributed (black), vaccine randomly distributed (red), distributed based on individual’s degree (blue), distributed based on individual’s number of healthy neighbors (purple), and distributed based on individual’s number of infected neighbors (green).

<table>
<thead>
<tr>
<th></th>
<th>No vaccines</th>
<th>Random</th>
<th>Degree</th>
<th>Healthy</th>
<th>Infected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total infected</td>
<td>1046</td>
<td>812</td>
<td>86</td>
<td>99</td>
<td>28</td>
</tr>
<tr>
<td>% infection averted</td>
<td>-</td>
<td>15%</td>
<td>92%</td>
<td>91%</td>
<td>97%</td>
</tr>
<tr>
<td>Improvement</td>
<td>-</td>
<td>-</td>
<td>6.2</td>
<td>6.1</td>
<td>6.6</td>
</tr>
</tbody>
</table>

Table 3.1: Summary of the effectiveness of different vaccine distribution strategies.

number of healthy neighbors. Table 1 summarizes the effectiveness of different vaccine distribution strategies. Total expected number of infections, % of infections averted by allocating vaccine doses, and improvement over random distribution strategy are summarized.

**Vaccine Distribution Based on a Cutwidth-Inspired Ranking** The cutwidth of a graph $G = (V, E)$ is the smallest integer $k$ such that the vertices $V(G) = \{v_1, \ldots, v_n\}$ of $G$ can be arranged in a linear layout $[v_1, \ldots, v_n]$ in such a way that, for every $i = 1, \ldots, n-1$, there are at most $k$ edges with one endpoint in the set $S = \{v_1, \ldots, v_i\}$ and the other in the set $\bar{S} = V - S = \{v_{i+1}, \ldots, v_n\}$ [32]. We will denote the cutwidth of a graph $G$ as $\text{CW}_f(G)$ where $f$ represents the particular ordering of vertices that generated the cutwidth. The cutwidth of a node, $\text{CW}_f(v_i)$, is the number of edges with one endpoint in the set $S = \{v_1, \ldots, v_i\}$ and the other in the set $\bar{S} = \{v_{i+1}, \ldots, v_n\}$ for the ordering $f = [v_1, \ldots, v_i, \ldots, v_n]$. Figure 6 is an example of a graph with the cutwidth of 5.
Top figure shows the network, and the bottom figure demonstrates the ordering of the vertices that generated the cutwidth.

We use the concept of the cutwidth in our vaccine allocation strategy. It was shown in [23, 24] that the cutwidth based policies achieve sublinear expected extinction time in some cases (resilient networks). The intuition is that by allocating resources to guide infections to occur in the order that generates the cutwidth, infection transmitting paths are minimized, therefore reducing the rate at which the infection spreads. Motivated by this intuition, we propose here a cutwidth based policy for our model. We rank individuals by the resulting cutwidth of the network when each individual is removed. The vaccine doses are then allocated in the order of individuals who will result in the lowest cutwidth when vaccinated. The purpose of this strategy is to vaccinate individuals to achieve...
the minimum possible infection-transmitting paths or contacts between infected and susceptible individuals in the network, reducing the infection spread as a result. We modify the cutwidth minimization integer program formulation proposed by Luttamaguzi et al. [72] to find the cutwidth from the linear arrangement of the susceptible individuals. The sequence below explains how we apply the cutwidth-based vaccine allocation strategy in our epidemic simulation:

1. At each time period $t$, remove individuals in the removed state.

2. Place all infected individuals in the front of the cutwidth ordering.

3. For each and every susceptible individual who have infected neighbors, remove the individual (change the infection state from $S$ to $R$) and find the cutwidth ordering among the remaining susceptible individuals and the resulting cutwidth using the integer programming formulation described below.

4. Rank the susceptible individuals in the order of resulting cutwidth, from the lowest to the highest.

5. Vaccinate the first $v(t)$ susceptible individuals in the cutwidth ranking ($v(t)$ represents the number of doses available at time period $t$).

An intuition behind the cutwidth-based vaccination strategy in our model is that we remove nodes that will minimize the possible infection transmitting paths between infected and susceptible individuals. Infection travels through edges connecting infected and susceptible individuals, and the rate at which infection spreads is proportional to the cut between infected and susceptible individuals. Therefore, by removing nodes that minimizes the cutwidth, we reduce the infection spreading rate. This is different than vaccinating individuals that are connected to the most infected individuals because the cutwidth also takes into account what will happen in the future.

**Integer Programming Formulation**

$$
\begin{align*}
\min & \quad b \\
\text{s.t.} & \quad \sum_{k \in H} x_i^k = 1 \quad (3.1) \\
& \quad \sum_{i \in H} x_i^k = 1 \quad (3.2) \\
& \quad y_{i,j}^{k,i} \leq x_i^k \quad (3.3) \\
& \quad y_{i,j}^{k,j} \leq x_j^k \quad (3.4) \\
& \quad x_i^k + x_j^l \leq y_{i,j}^{k,i} + 1 \quad (3.5)
\end{align*}
$$
\[
\sum_{(i,j) \in \mathcal{E}} \left( \sum_{k \leq c < l} y_{i,j}^{k,l} + \sum_{l \leq c < k} y_{i,j}^{k,l} \right) + \sum_{k > c} \left( \sum_{i \in H} d_i x_i^k \right) \leq b, \quad \forall c \in \{1, \ldots, |H| - 1\} \tag{3.7}
\]

\[x_i^k \in \{0, 1\} \tag{3.8}\]

\(H\) denotes the set of susceptible individuals in the network at time period \(t\). \(x_i^k\) are the binary decision variables with indices \(i \in H\) and \(k \in \{1, 2, \ldots, |H|\}\) specifying whether node \(i\) is placed in position \(k\) in the ordering. This binary variable takes on value 1 if and only if \(i\) occupies the position \(k\) in the ordering; otherwise \(x_i^k\) takes on value 0. Constraints (3.2) and (3.3) together ensure that each node is assigned to one position, and two nodes are in different positions. Consequently, constraints (3.2), (3.3) and (3.8) together imply that a solution of the formulation is an ordering. \(y_{i,j}^{k,l}\) indicates whether both \(x_i^k = 1\) and \(x_j^l = 1\). Therefore, \(y_{i,j}^{k,l} = x_i^k x_j^l\), which is accomplished by the linear constraints (3.4)-(3.6).

Constraint (3.7) reflects, for each position \(c\) in the ordering, the number of edges whose origin is placed in any position \(k\) (\(1 \leq k \leq c\)) and destination in any position \(l\) (\(c+1 < l \leq |H|\)), where \(\mathcal{E}\) represents the set of edges between susceptible nodes and \(d_i\) represents the number of edges connecting node \(i\) and the infected nodes. Because the cutwidth problem consists of minimizing the maximum number of cutting edges in any position \(c \in \{1, \ldots, |H| - 1\}\) of the ordering, the objective function \(b\) must be a value greater than or equal to this quantity.

Figures 7-11 show the example of the vaccine distribution strategy based on the cutwidth-inspired ranking. Figure 7 shows an example network at a time when some individuals are in infected and removed states. Figures 8-10 show the resulting cutwidth obtained by solving the integer programming formulation when each susceptible individual connected to the infected neighbor(s) is removed. Figure 11 shows the ranking of the susceptible individuals in the order of the resulting cutwidth, from the lowest to the highest. Available vaccine doses will be distributed following the ranking.

We evaluate the effectiveness of the cutwidth-based vaccine distribution strategy by simulating and the expected infection dynamics under different vaccine allocation strategies. For demonstration purposes, we simulated the infection spread on the scale-free network of 10 nodes with average degree of 2 and two initial infections. Infection rate of .69 and recovery rate of .45 were used for this simulation. Figure 12 compares the resulting infection curves, and Table 2 summarizes the results. Total expected number of infections, % of infections averted by allocating vaccine doses, and infection extinction time are summarized. In the simulation, the cutwidth-inspired-ranking-based vaccine distribution strategy performed better than all other proposed strategies in mitigating the infections. A good improvement was observed in the infection extinction time. While the infections persisted until five to six time periods into the infection spread under the other strategies, the cutwidth-based distribution strategy effectively eradicated the infections after three time periods.
Figure 3-7: Example network of 8 nodes in different states.

Figure 3-8: Resulting cutwidth when node 1 is removed.

Figure 3-9: Resulting cutwidth when node 4 is removed.
Figure 3-10: Resulting cutwidth when node 6 is removed.

Figure 3-11: The ranking of the susceptible individuals to receive vaccine in the order of the resulting cutwidth, from the lowest to the highest.
Although the vaccine distribution strategy based on the cutwidth-inspired ranking performs well, there are computational disadvantages to it. The integer programming formulation requires setting up $O(n^4)$ constraints in order to solve for the cutwidth of the network, where $n$ is the size of the network. Moreover, the formulation needs to be solved multiple times at each time of the epidemic. When the network size gets large, the epidemic simulation under this strategy can take a very long time.

<table>
<thead>
<tr>
<th></th>
<th>No vaccines</th>
<th>Random</th>
<th>Degree</th>
<th>Healthy</th>
<th>Infected</th>
<th>Cutwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total expected</td>
<td>8.66</td>
<td>5.60</td>
<td>4.36</td>
<td>4.58</td>
<td>4.12</td>
<td>3.80</td>
</tr>
<tr>
<td>infections</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% infection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>averted</td>
<td></td>
<td>35.3%</td>
<td>49.7%</td>
<td>47.1%</td>
<td>52.4%</td>
<td>56.1%</td>
</tr>
<tr>
<td>Infection</td>
<td>9</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>extinction time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.2: Summary of the effectiveness of different vaccine distribution strategies, including the cutwidth-based vaccine distribution strategy.
3.2.2.3 Vaccine Effectiveness Under Varying Compliance Rates

Although the suggested strategies are effective in greatly reducing the spread of H1N1 influenza, it is not always possible to force all selected individuals to get vaccinated. In this section, we analyze how effective the suggested vaccine distribution strategies perform under varying compliance rates. For example, the degree-based vaccination strategy described above tries to allocate vaccine to $v(t)$ most connected individuals ($v(t)$ represents the number of doses available at time period $t$), however, only a small number of $v(t)$ individuals may comply with getting vaccinated. We tested the effectiveness of the vaccine distribution strategies when varying fractions of targeted individuals are randomly chosen to get vaccinated. Figure 13 shows how much different strategies reduce infections by distributing doses efficiently under varying compliance rates. Results indicate that even if only 4% of the targeted individuals decide to get vaccinated, all proposed strategies can achieve reductions in infections similar to the random distribution strategy.

3.2.2.4 Sensitivity Analysis

Figure 14 depicts how different vaccine distribution strategies are sensitive to the number of vaccine doses available at each time period. The proposed distribution strategies that target the most infectious individuals resulted in a significant improvement by using the first couple of doses. The distribution strategy based on the individual's number of infected neighbors reduced the infections by 90% with only 5 doses at each time period. The allocation strategies based on the degree and on
the individual's number of healthy neighbors reduced the infections by 90% with 30 doses at each time period. After the most infectious individuals were vaccinated and over 90% of the infections were prevented, the improvement in reductions is negligible.

On the other hand, the random distribution strategy resulted in much slower improvement as the available doses increase. In the scale-free network, majority of the individuals have small number of connections, and vaccinating these non-influential individuals will not slow the propagation effectively. Since only a small fraction of the individuals are influential, it is expected that many doses of vaccine are needed to vaccinate the most infectious individuals by random distribution. As many as 1400 doses are needed in order to reduce the infections by 90%.

3.3 Discussion

It is imperative to have a good understanding of the contact network on which the infectious disease spreads and its propagation dynamics in order to analyze the dynamics of the infection and evaluate control strategies. Given the results from Chapter 2, we are able to realistically simulate the infection dynamics of H1N1 influenza. In this chapter, we focused on the impact of pharmaceutical interventions on the propagation of the influenza. Specifically, we evaluated the effectiveness of vaccine against the spread of H1N1 influenza. Furthermore, we showed that using the knowledge of the contact network and infection states, vaccine can be more efficiently utilized. By distributing vaccine according to the number of connections each individual has, the total expected number of

Figure 3-14: Different vaccine distribution strategies’ sensitivity to the number of vaccine doses available at each time period.
infections can be reduced by 6 times as much compared to a random allocation scheme for the case of the state of Massachusetts. Moreover, distributing the vaccine based on the individual’s number of infected neighbors successfully slowed the propagation of infections and reduced the total expected number of infections by as much as 6.6 times compared to the random distribution strategy. These proposed vaccine distribution strategies were shown to reduce infections as much as the random distribution strategy under a 4% compliance rate. This result confirms that good understanding of a contact network structure is essential to evaluate and devise efficient and effective control policies.

The results from this chapter provide government officials and policy makers useful insights for their infection prevention efforts. Although implementing the selective vaccination strategies described in this chapter can be difficult from the public policy point of view, having the tools to estimate the benefits of such strategies and the additional resources required for the current strategy to be as effective as other strategies will aid the officials in determining the best course of resource investment to fight infectious diseases.
Chapter 4

Dynamic Network

4.1 Introduction

Many real-world complex systems can be represented as networks, where the entities in these systems represent the nodes or vertices and links or edges that connect the vertices. We encounter such networks in almost any application domain such as biology [73], computer science [12, 13, 74], epidemiology [5, 36] and sociology [12, 13]. Networks are especially useful to represent social ties and interactions, and understanding of these networks allows us to study the propagation dynamics of ideas, rumors, infectious diseases, and other entities that can spread by way of social connections. In previous chapters, we used social network analysis, agent-based modeling, and parameter estimation to realistically model the spread of the H1N1 influenza in various states. In simulating how the H1N1 influenza spreads, we assumed that the underlying social network remains static, meaning that the number of nodes does not change over time. Moreover, the edges connecting nodes, representing the contacts between the nodes, remain fixed over time. Indeed, most social network analysis in the past concentrated on static networks. However, most real networks evolve through time: changes of network structure can happen if some nodes and/or edges appear and disappear [75, 76]. For our case, the contact network will become dynamic if we consider births and deaths as well as change in contacts as individuals travel and change locations through their daily activities. In this chapter, we introduce a dynamic component to the network model we discussed in previous chapters to demonstrate that the dynamic nature of the contact networks can be modeled.
4.2 Dynamic Network Model and Parameter Estimation Results

4.2.1 Dynamic Network Model

There are many factors that make the underlying social network on which the H1N1 influenza spreads dynamic. For example, births and deaths will create new nodes and remove existing nodes. Also, when individuals make new connections, new edges will be created between the corresponding pairs of individuals. Susceptible individuals can contract the flu infection through person-to-person contact with infected individuals. Therefore, we used the network representation of the social connections to model the person-to-person contacts upon which the influenza can spread. Because we assumed that the network is static, individuals in the states and their connections did not change over time. In our dynamic network model, we continue to assume that there are no births and deaths, so the sizes of the networks will remain the same, which is equivalent to the estimated population sizes of the U.S. states at the beginning of the 2009 H1N1 flu season, as shown in Chapter 2. This is not an excessively unrealistic assumption since the number of nodes can stay relatively stable when the birth and death rates are similar during the flu observation period. When the edges remain static, this signifies that people make person-to-person contacts with the same connections repeatedly during the flu season. Realistically, people do have frequent contacts with people they do not know by chance in public places such as transit stations, restaurants, and workplaces, and these random contacts can cause the individuals to contract infectious diseases.

In order to reflect these random contacts in representing the network, we add new edges on the static network at the beginning of every time interval on our discrete time model; these edges last only for a single time period. The edges from the static network can be interpreted as the relationships between close acquaintances such as family, friends, classmates, and co-workers. The close acquaintances make contact repetitively, so the edges representing contacts between these persons remain intact throughout the entire observation period. New edges that form for a short time interval can be thought of as the random encounters in public places, and this adds a dynamic component to the static network that we previously dealt with.

We add a parameter to our previous scale-free static network model to generate the dynamic network described above. We denote this parameter as $r$. As described in Chapter 2, a scale-free network can be generated with a single parameter $k$, representing the average degree. We first generate the scale-free network, we visit each and every node, and each node will add a new edge to every other node in the network with the probability $r$. Next, this new network becomes the underlying network on which the disease process propagates for a single time period. Finally, we go back to the original scale-free network and repeat the random process of adding new edges to generate the network for the next time interval. As a result of this random process, the contact
network will have a different set of new edges reflecting random encounters at each time period.

With these two parameters, \( k \) and \( r \), a dynamic network can be generated, and along with the disease dynamic parameters introduced in Chapter 2, \( \beta \) and \( \delta \), we can simulate the H1N1 influenza propagation on the dynamic network described above. In the following section, we will apply the parameter estimation algorithm that was previously introduced in Chapter 2 on the dynamic network.

### 4.2.2 Parameter Estimation on Dynamic Network Model

Now that we can simulate the disease propagation process, we can apply the parameter estimation algorithm described in Chapter 2 on the dynamic network disease propagation model. The algorithm applied to this new dynamic network model will find the best-fitting parameters that simulate the expected infection dynamics that closely mimic the real estimated infection curve from the data introduced in Chapter 2. The following recapitulates the parameter estimation algorithm applied on the dynamic network model:

#### Input

The inputs to the algorithm include:

- A parameterized disease spread dynamic network structure, where nodes represent people and undirected edges represent contacts between people. We use the scale-free network initial average degree, \( k_0 \), as the underlying static network representing contacts among close acquaintances. The initial random contact parameter, \( r_0 \), is used to represent unpredicted and sporadic contacts, which cause the network to be dynamic in terms of edges.

- Initial values of the disease process parameters \( \beta_0 \) and \( \delta_0 \).

- Real temporal infection data to fit the model-generated expected infection dynamics.

- Data on vaccines administered (if administered).

#### Output

- The algorithm outputs network and infection parameters that are used to generate a simulated expected infection curve that fits the real data as closely as possible.

#### Procedure

Begin with the given initial values of the social network and disease spread parameters: \( k_0, r_0, \beta_0, \) and \( \delta_0 \). Let \( \Delta k, \Delta r, \Delta \beta, \) and \( \Delta \delta \) be the amounts by which \( k, r, \beta, \) and \( \delta \) are changed at each step in the optimization. Let \( e_r \) and \( e_t \) each denote the number of infections for the real infection curve.
and the estimated expected infection curve at time t, respectively. We define the error, $\hat{E}$, between the simulated expected infection curve and the true infection curve as:

$$\hat{E} = \sum_{t=1}^{t=\text{max period}} |e_t - \hat{e}_t|$$  \hspace{1cm} (4.1)

Repeat the following steps until the error can no longer be reduced by changes to the parameters (we define the optimal output parameters as $k^*$, $r^*$, $\beta^*$, and $\delta^*$):

1. Given $k_0$, $r_0$, $\beta_0$, and $\delta_0$, search in all possible directions to find a direction that improves $\hat{E}$. That is, evaluate $\hat{E}$ at all possible combinations of $k$, $r$, $\beta$, and $\delta$, where $k \in \{k_0, k_0 + \Delta k, k_0 - \Delta k\}$, $r \in \{r_0, r_0 + \Delta r, r_0 - \Delta r\}$, $\beta \in \{\beta_0, \beta_0 + \Delta \beta, \beta_0 - \Delta \beta\}$, and $\delta \in \{\delta_0, \delta_0 + \Delta \delta, \delta_0 - \Delta \delta\}$. Evaluate $\hat{E}$ by doing the following for each set of parameters:

   (a) Generate the scale-free network according to the selected average degree, $k$. Generate new edges representing random encounters according to the random encounter probability, $r$.

   (b) Simulate R realizations of the disease process. For each realization, initialize the disease simulation infection by assigning $N_1$ nodes to the infected states, where $N_1$ is the number of people infected at the beginning of the observation period in the data. We assume that the initial infected nodes are selected uniformly at random from among all the nodes. This is a commonly used assumption in epidemic simulation [45]. Update the random encounters according to $r$ on the underlying scale-free network. Update the disease states for each time period, according to the disease process parameters and the vaccine administration data. (We assume that those who receive vaccines in each time period are chosen uniformly at random.)

   (c) Generate an expected infection curve by averaging the number of infected individuals at each time period over the R simulated realizations of the disease process.

   (d) Calculate $\hat{E}$.

2. Determine which search direction resulted in the minimum error. Update $k_0$, $r_0$, $\beta_0$, and $\delta_0$ to the values of $k$, $r$, $\beta$, and $\delta$ that achieved the lowest error. The algorithm is summarized as a flow chart in Figure 1.

4.2.3 Estimation Results

It is noteworthy that the new dynamic network disease propagation model extends the previous static network disease dynamic model in a simple manner, adding one additional parameter that makes the network dynamic by considering random encounters at each time period. This allows us to test if implementation of the dynamic network disease dynamic model drastically improves
upon the static network disease model. As Equation (2) indicates, because the dynamic network disease propagation model keeps the parameters from the static network disease propagation model in addition to the new parameter, the parameter estimation algorithm applied on the dynamic network model will perform at least as well as the static network model. If the dynamic network model does not improve upon the static network model, then the parameter estimation applied on the dynamic network model will result in the same best-fitting parameter values as the algorithm applied on the static network model, $k^*$, $\beta^*$, and $\delta^*$, and zero probability for the random encounter parameter, $r^*$.

\[
\begin{align*}
\min_{k, \beta, \delta} \bar{E} & \quad \geq \quad \min_{k, \beta, \delta} \bar{E} \\
\text{s.t. parameter estimation algorithm} & \quad \text{s.t. parameter estimation algorithm}
\end{align*}
\]

If the parameter estimation algorithm applied on the dynamic network model results in non-zero values for $r^*$, this indicates that the better fit was obtained from the dynamic network disease propagation model compared to the static network model, suggesting that the dynamic network model can simulate more realistic expected infection dynamics.

We applied the algorithm on the dynamic network disease propagation model for the state of Massachusetts and compared the parameter estimation result on the static scale-free network disease propagation model. Figure 2 below compares the best-fitting expected infection curves under static and dynamic network models. The real estimated infection curve is shown in black solid lines with circles. Note that the population size was reduced by a factor of 10,000 to make the computation faster.

Although it is hard to tell by inspecting the curves, the best-fitting expected infection curve generated using the dynamic network model showed a slight improvement from the fit obtained using the static network disease propagation model. The resulting parameter values were 2 for average degree ($k$), .00003 for random contact probability ($r$), .43 for the infection probability ($\beta$),
and .67 for the recovery probability ($\delta$). Measured error, the absolute sum of the differences in the infections, $\bar{E}$, constituted about 13% of the total infection, and this is about a 20% improvement from the 16% error measured from the static network disease spread model. The error on the total expected number of infections, $\bar{E}_{Total}$, was 1.5%, improved from 1.9%. The error on the expected infections at the peak, $\bar{E}_{Peak}$, was 2.33%, compared to 2.27%. Table 1 below summarizes the results.

<table>
<thead>
<tr>
<th>Massachusetts</th>
<th>Static</th>
<th>Dynamic</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>$r$</td>
<td>-</td>
<td>.00003</td>
</tr>
<tr>
<td>$\beta$</td>
<td>.45</td>
<td>.43</td>
</tr>
<tr>
<td>$\delta$</td>
<td>.69</td>
<td>.67</td>
</tr>
<tr>
<td>$E$</td>
<td>16%</td>
<td>13%</td>
</tr>
<tr>
<td>$\bar{E}_{Total}$</td>
<td>1.9%</td>
<td>1.5%</td>
</tr>
<tr>
<td>$\bar{E}_{Peak}$</td>
<td>2.3%</td>
<td>2.3%</td>
</tr>
</tbody>
</table>

Table 4.1: Summary of Massachusetts parameter estimation results.
4.3 Discussion

In this chapter we introduced a dynamic network model that extends the previously used network model that was static. Although the majority of the research on network analysis in the past has dealt with static networks where the number of nodes or edges forming the network does not change over time, the need for using dynamic networks is being widely recognized, as most real-world networks evolve through time. Whether the networks under analysis are representing internet router systems, social contact structures, or human DNA systems, changes in the systems can happen over time, and static networks are not suitable to reflect these changes in the systems. For the network representing the social contacts for which the infectious disease can spread, birth and death processes as well as change in contacts between individuals all reflect changes in the system, and thus require change in the network structure representing the system.

Previously we have shown that the scale-free network well reflected the contact network on which the H1N1 influenza spreads. We extended this static network by introducing an additional parameter reflecting the random encounters people have in public areas that can cause the spread of the influenza virus. By extending the static model, we were able to test to see if the dynamic network disease model can simulate the disease dynamic more realistically compared to the static network model. The parameter estimation algorithm applied on the dynamic network disease propagation model showed that the dynamic network model including the effects on random encounters simulated the H1N1 dynamics in Massachusetts slightly better, compared to the static network disease spread model we introduced previously.

We showed a way to make the static network model more realistic by considering random encounters at each time period throughout the observation period. However, there are many other ways to make network models dynamic and more realistic. For example, changes in the number of nodes can be implemented to reflect the birth and death processes in the social network. The edges can appear and disappear as close acquaintances move around geographically and cause changes in person-to-person contact frequency. This work provides a foundation for expanding static social network models to incorporate the dynamic aspects of contact networks. We hope that future research will strengthen the capability to model dynamic network systems.
Chapter 5

Conclusions and Future Research

In this chapter we offer some conclusions based upon the models, results, and analysis. We also suggest some future research directions.

5.1 Conclusions

This work provides a tool for realistic epidemic simulations based on actual data, which enables the capabilities to analyze vaccine effectiveness and the efficiency of various vaccine allocation strategies. Such capabilities will support government officials to make better decisions on how to allocate the limited vaccine resources in order to achieve the greatest reduction in casualties from infectious diseases.

In Chapter 2, we described the data collected and the approximations used to estimate the real infection dynamics. We presented the parameter estimation algorithm that estimates the contact network structure and disease propagation dynamics, which can simulate the expected infection spread that is close to the real infection dynamics. Results of the algorithm applied to the 2009-2010 H1N1 infection data from the four U.S. states (Massachusetts, New York, North Dakota, and Montana) showed that the best-fit model indicates a scale-free network.

Based on this finding, in Chapter 3 we proposed several vaccine allocation strategies that exploited the special network structure, namely the power-law degree distribution, and provided significant reduction in infections. Moreover, in our simulations, the recommended strategies proved to be as effective as random distribution strategies under compliance rates as low as 4%. From these results, we can clearly see the potential power of this tool to guide decision making in vaccine allocation to fight infectious diseases.

Chapter 4 presented a methodology to modify the static network model to incorporate the dynamic nature of human interactions. While the static network model used in previous chapters was limited to representing contacts that occur on a regular basis, such as contacts between family
members, friends, and co-workers, the dynamic network model was able to capture more irregular random encounters that can happen in public places. In our experiment, some improvement was observed in estimating the real infection dynamics using the dynamic network model. This suggests the possibilities for investing in future research to develop improved models that would better guide decision makers in determining efficient vaccine distribution strategies.

5.2 Future Research

Future research should include several additions and modifications to the social network model, the parameter estimation algorithm, and efficient control policy.

One can examine alternative disease processes, such as SEIR (Susceptible-Exposed-Infected-Removed), MSIR (Maternally immune-Susceptible-Infected-Removed), and SIS (Susceptible-Infected-Susceptible). Depending on the disease being studied, some epidemiology models are more accurate than others. Some studies modeling disease process as the SIS epidemic proposed methods for the optimal resource allocation. However, optimal vaccine distribution problem for the SIR model is known to be NP-hard, therefore guiding many studies on the SIR epidemic to rely on heuristic algorithms. Exploring different epidemic models will bring helpful insights into designing effective control strategies.

One can model and examine the effect of human behaviors on the infections in addition to the effect of pharmaceutical intervention. There are many studies that focus on the pharmaceutical intervention to stop the disease spread. Increasing number of studies examine the effect of human behaviors on the disease spread. Modeling and analyzing the effect of both the pharmaceutical and non-pharmaceutical intervention strategies will provide additional intuition for controlling the epidemics.

One can examine more efficient parameter estimation algorithms. Because greedily searching through all possibilities of parameter values bring computational burden, we had to factor down the size of the network. Implementing other methods by works that focus on network inference (discussed in introduction) or applying good pruning rules can refine search directions and enable more efficient search on larger networks.

One can examine alternative tools to develop near-optimal control policies. For the SIR model where the optimal vaccination problem is hard to obtain, areas of control theory and mathematical programming may be further explored in an effort to develop near-optimal control strategies. Relaxation to the NP-hard problem to obtain the bounds on the optimal solution will help evaluating effectiveness of heuristic algorithms.

The increasing availability of data and growing need for stronger modeling and predictive capabilities are driving ever greater expansion in the areas of social network analysis and biosurveillance.
This study provides a framework for the future examination of many different diseases and effective intervention strategies to prevent casualties. We hope that this work serves as a foundation for much more research to come.
Bibliography


