THE INFLUENCE OF SOCIAL NETWORK GRAPH STRUCTURE ON DISEASE DYNAMICS IN A SIMULATED ENVIRONMENT

Tina V. Johnson, M.S.

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

December 2010

APPROVED:

Armin R. Mikler, Major Professor
Samuel F. Atkinson, Committee Member
Suhasini Ramisetty-Mikler, Committee Member
Philip H. Sweany, Committee Member
Xiaohui Yuan, Committee Member
Ian Parberry, Interim Chair of the Department of Computer Science and Engineering
Costas Tsatsoulis, Dean of the College of Engineering
James D. Meernik, Acting Dean of the Robert B. Toulouse School of Graduate Studies

Johnson, Tina V. The Influence of Social Network Graph Structure on Disease Dynamics in a Simulated Environment. Doctor of Philosophy (Computer Science and Engineering), December 2010, 91 pp., 19 tables, 33 illustrations, bibliography, 84 titles.

The fight against epidemics/pandemics is one of man versus nature. Technological advances have not only improved existing methods for monitoring and controlling disease outbreaks, but have also provided new means for investigation, such as through modeling and simulation. This dissertation explores the relationship between social structure and disease dynamics. Social structures are modeled as graphs, and outbreaks are simulated based on a well-recognized standard, the susceptible-infectious-removed (SIR) paradigm. Two independent, but related, studies are presented. The first involves measuring the severity of outbreaks as social network parameters are altered. The second study investigates the efficacy of various vaccination policies based on social structure. Three disease-related centrality measures are introduced, *contact, transmission,* and *spread centrality*, which are related to previously established centrality measures *degree, betweenness,* and *closeness*, respectively. The results of experiments presented in this dissertation indicate that reducing the neighborhood size along with outside-of-neighborhood contacts diminishes the severity of disease outbreaks. Vaccination strategies can effectively reduce these parameters. Additionally, vaccination policies that target individuals with high centrality are generally shown to be slightly more effective than a random vaccination policy. These results combined with past and future studies will assist public health officials in their effort to minimize the effects of inevitable disease epidemics/pandemics.

ACKNOWLEDGMENTS

It is difficult to describe in words the gratitude that I feel toward Dr. Armin R. Mikler for his tremendous help and support over the past several years. When I initially spoke with him about enrolling as a PhD student at UNT, I was completely amazed by his knowledge and enthusiasm. More than four years later, I am still completely amazed. His dedication to his job, field of study, and particularly to his students is remarkable. Thank you, Dr. Mikler, for your patience, guidance, and friendship. You are truly an inspiration to me.

I would like to thank Dr. Susie Ramisetty-Mikler for serving as a committee member and for the additional assistance she provided as I was writing this dissertation. Through e-mail and personal meetings, she helped me through some of the more difficult portions of this work. Her input was invaluable. Additionally, I want to express gratitude toward committee members, Dr. Samuel F. Atkinson, Dr. Philip H. Sweany, and Dr. Xiaohui Yuan. To all my committee members, thank you for your time and support.

Many people have provided help and encouragement and I would be remiss if I did not include the following people in this acknowledgement: Terry Griffin, not only for his help with GraphViz, but also for his friendship and company on the drive to and from UNT. Dr. Ranette Halverson and other faculty members at Midwestern State University for their encouragement and support. Members of the CeCera Research Group for being a fantastic group of friends and a wonderful resource for help. Dr. Thomas Böhme for graph theory insights and discussions and Dr. Sumihiro Suzuki for assistance with statistical analyses.

Most important, I want to thank my family, Adam, Tonya, Kerri, Cullin, and many extended family members. This journey has not been easy and I would not have finished if not for the people who matter the most in my life. Thank you for your continued love and encouragement.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

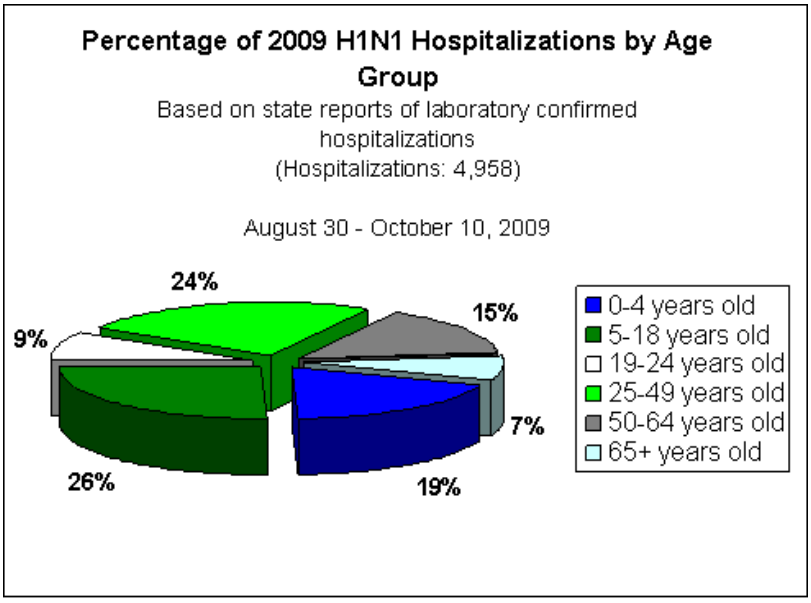
INTRODUCTION

Naturally occurring epidemics/pandemics have always plagued mankind, however, increased population coupled with decreased travel time in the modern era have amplified the cause for concern over such events. Additionally, the cultural environment today is much more diverse than ever before. An outbreak that once would have affected only a small portion of society might now impact the entire world. Moreover, public health experts agree that future epidemics/pandemics are inevitable [64]. We should not ask, "When will it happen?" but rather "How will we deal with it when it happens?" The recent emergence of the H1N1 virus, also known as the swine flu, increased public awareness regarding the serious nature of a pandemic event. In fact, as of 17 October 2009, the World Health Organization (WHO) reports more than $414,000$ laboratory confirmed cases and nearly $5,000$ deaths worldwide attributed to the H1N1 virus [2]. In the United States alone, the Centers for Disease Control and Prevention (CDC) reports $4,958$ laboratory confirmed hospitalizations and 292 deaths as a result of the virus during the time period of 30 August 2009 through 10 October 2009 [1]. The breakdown of the US cases by age is shown in Figure 1.1. Disease dynamics, i.e. how, where, and to whom a disease will spread, are unpredictable. Emerging viruses do not necessarily follow the same pattern as previous outbreaks of a similar nature.

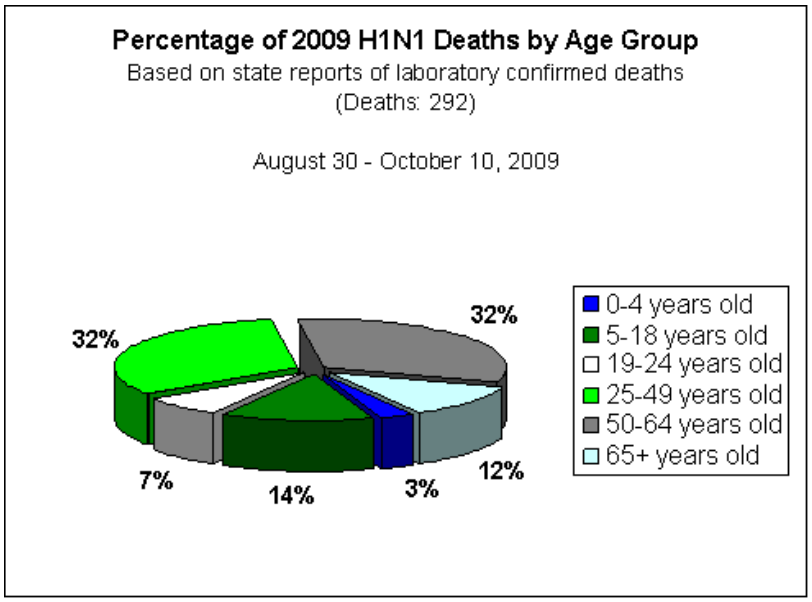
Among other concerns, there is often a shortage of vaccines. Shortages may be caused by an interruption in supply, an increase in demand, or a lack of funding [41]. In the case of the 2009 H1N1 swine flu, delays were attributed to problems in production, packaging, and distribution, along with the challenge of producing the seasonal flu vaccine at the same time [35]. Regardless of the reason, a lack of vaccine for an entire population presents the dilemma of who should receive the available supply. Several options can be considered when vaccine priority decisions are made. A policy that is best for the individual may not be optimal for the entire population [34]. Highest priority groups for the H1N1 vaccine were health care workers and people who were at

risk of severe complications if infected. This included pregnant women, young children, people who lived with or cared for children under six months old, and children ages 5 to 18 with chronic medical conditions [28]. A long-standing policy for seasonal influenza includes vaccination of the elderly, even though schoolchildren and working adults are known to transfer disease at a higher rate due to having a higher contact rate [34]. A recent study conducted by researchers at Yale University School of Medicine and Clemson University found that consideration of transmission is an important factor when developing a vaccination policy. Further, the study concluded that previous and new CDC recommendations are suboptimal based on five outcome measures: total infections averted, total deaths averted, years of life lost, contingent valuation (an assumption of life value based on age), and economic costs [53]. The controversy over who should be eligible to receive a vaccine when the supply is limited is not one that is easily resolved.

On the positive side, evolution of technology, such as real-time surveillance, has provided access to unprecedented resources that can be used to fight the spread of infectious diseases. Even the ability to quickly and efficiently disseminate information plays a vital role in preventing an outbreak from getting out of control. However, information alone is insufficient to adequately prepare for the emergence of new and unknown infectious diseases. Simulation and modeling tools are needed to better understand disease dynamics and prepare for unseen types of epidemics, thereby improving methods for disease control. Development of such tools requires cooperation and coordination among the government, public health agencies, and universities. It further necessitates a collaborative approach by experts in the fields of biology, medicine, sociology, epidemiology, technology, computer science, etc. The joint effort of these entities and individuals are crucially important to compensate for the favorable disease environment that has been created through the natural progression of mankind. The interest of public welfare is at stake.



(a)



(b)

FIGURE 1.1. CDC reported H1N1 related hospitalizations (a) and deaths (b) in the United States [1]

1.1. Disease Definitions and Concepts

Terms associated with disease and disease spread may not be clearly understood by the general public. Misconceptions that all diseases are infectious or that all infectious diseases are communicable are common. To appropriately model diseases, it is important to understand basic terminology used in the field of epidemiology.

DEFINITION 1.1. Disease is an interruption, cessation, or disorder of body functions, systems, or organs [54].

Infectious diseases are caused by an invasion of biological agents, collectively referred to as *pathogens* that include bacteria, viruses, or parasites. Pathogens have the ability to enter, survive, and multiply within a host. If the pathogens additionally have the ability to transfer from a host to another agent, the disease is considered *communicable*. The transmission of a communicable disease can be *vertical*, host to offspring, or *horizontal*, host to peer. Horizontal transmission may occur through direct contact, may be air-borne, food-borne, or water-borne, or may require a vector, as with Malaria. Both infectious and noninfectious diseases can be classified as either *acute*, sudden onset with a relatively short duration, or *chronic*, less severe but much longer lasting [54].

The models developed for this research were designed to simulate infectious, communicable, acute diseases. While not restricted to influenza, the models discussed herein emphasize influenza-like illnesses. Four stages of progression, shown in Figure 1.2, are generally associated with this type of disease. The first stage describes the time period prior to the point of infection. This is the *susceptible* stage. The second stage, *presymptomatic*, encompasses a *latent* state and an *incubation* period. The latent state is the time beginning when an individual is first infected until they themselves are able to infect others. The incubation period describes the time between the point when infection occurs and the moment when symptoms emerge. The third stage is that of *clinical disease* which begins when symptoms first appear. The final stage is the *removed* state, which is the result of recovery or death.

1.2. Disease Simulation

When a disease is introduced into a population, certain conditions must be met in order for the disease to transmit and successfully spread. Both disease and population parameters influence the course of the potential outbreak. To simulate an infectious outbreak it is important not only to use a valid disease model, but also to recognize the essential role of the underlying social network. In the initial stage of an outbreak, the majority of the population is susceptible to the disease. As the disease spreads, the number of individuals who are susceptible decreases and the number

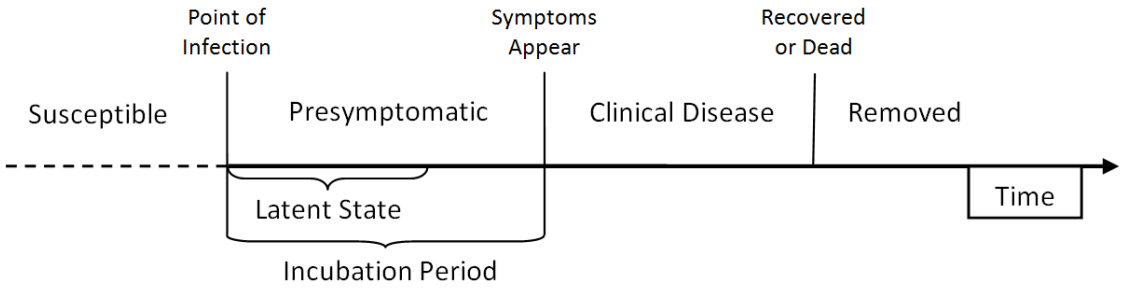


FIGURE 1.2. Disease progression within a host

of those who become infected/infectious increases. Assuming that the disease is acute rather than chronic, the progression of the outbreak eventually results in a decrease in the number of infectious individuals and an increase in the number of those who have recovered from the illness or are otherwise removed, such as through natural immunity or death. The movement of the population from states of susceptible, infectious, and removed forms a basis for modeling disease spread.

The susceptible-infectious-removed (SIR) paradigm and its counterparts, such as susceptible-infectious-susceptible (SIS) and susceptible-latent-infectious-removed (SLIR), are recognized standards for modeling many infectious diseases. The SIR model (discussed more thoroughly in Section 2.2) was first introduced by Kermack and McKendrick in a 1927 paper titled "A Contribution to the Mathematical Theory of Epidemics" [46]. The basic SIR model can be modified as necessary to more accurately represent the particular disease under question.

Just as the disease model is important to the simulation, so is the underlying social network. Social networks are complex and graph models used to mimic these networks may vary. Connections between individuals, and thus disease contacts, are precarious. The research presented in this dissertation explores the effect of graph structure on the dynamics of disease spread in a simulated environment.

1.3. Disease Dynamics

Is it possible to precisely measure the severity of an epidemic or pandemic? What gauge can be used to determine that an outbreak at a particular time and place is more destructive than one at another time and/or location? Because parameters change from one occurrence to another, it may be impossible to make an entirely valid comparison between two distinct outbreaks. There are, however, indicators that are widely accepted as epidemiologic quantifiers. Even though these standards may not provide a completely unbiased account for comparison, they do provide a metric for classification.

One quantifier often referred to in disease-related literature is the basic reproduction number, R_0 [6, 7, 19, 39, 67]. R_0, as formally defined in Section 2.3, is the expected average number of secondary infections by a single infectious individual in a completely susceptible population. It is an epidemic threshold that is measured at the beginning of an outbreak at a time when the majority of the population is susceptible. R_0 provides an indication how quickly an infection will spread. Because R_0 is based on secondary infections, larger values of R_0 suggest a higher probability that an outbreak will progress into an epidemic or pandemic. After an epidemic/pandemic has run its course, the duration of the outbreak and the total number and proportion of individuals infected can also be considered. In a simulated environment, these values can be measured and compared from one outbreak to another.

1.4. Social Networks and Graph Theory

Graphs are exceptionally useful tools for analyzing social networks [79]. In the study of graph theory, graphs are represented by a set of vertices and a set of edges such that the edges represent an association between two vertices [82]. In a social network, the vertices represent individuals or groups of individuals and the edges represent some sort of connection between two people or two groups. There are many advantages to using graphs to analyze social networks, including an established vocabulary, mathematical operations, and the ability to use and prove theorems about graphs that can be transferred to the social structure [79].

If the social structure is already known, a corresponding graph can be constructed based on the existing data. An example of this is Padgett's Florentine families [79]. This network consists of sixteen families where the edges represent marriages between pairs of families. Historical data allows the creation of a representative graph. On the other hand, when a graph is developed for simulation purposes, the exact nature of the structure may be unknown because the network usually emerges as a function of a random sequence. The simulated graphs constructed for this research range from entirely random to completely ordered.

1.5. Problem Statement

The foundation of disease modeling is dependent upon the design of the underlying social network. The fundamental premise of the research presented in this dissertation is that network structure and disease outbreaks are tightly coupled. Results presented here reveal that changes in social structure affect several aspects of disease spread, including the basic reproduction number, the outbreak duration, and the proportion of individuals who become infected. Further, it is demonstrated that intervention strategies within an established social structure affect these figures. In particular, the following research questions are addressed:

(i) In a simulated environment, how does the particular social network structure predict R_0, the proportion of the population that becomes infected, and the epidemic/pandemic duration?

(ii) How does the vaccination of key individuals in an established social network, as identified by centrality measures, affect the progression and outcome of an epidemic in terms of R_0, the proportion of the population that becomes infected, and the epidemic/pandemic duration?

(iii) Which vaccination strategies are the most effective for specific social network structures in a simulated environment as measured by a reduction in outbreaks affecting greater than 20% of the population and a reduction in the proportion of the population infected?

Although the results from simulated environments are not likely to completely transfer to real life, the insight gained from such research can certainly help direct investigations in applied settings. The questions addressed here are designed to promote interest in graph theory as it applies to disease spread through social networks, particularly as an approach that can be used to prevent or impede an epidemic/pandemic. Targeted vaccination policies are explored at the theoretical level in this research in expectation that the results will have relevance in practice.

1.6. Overview

This chapter has introduced key concepts and provided the motivation for the research presented herein. The remainder of this dissertation is structured as follows: Chapter 2, where most of the significant literature is reviewed, establishes the necessary background in the areas of epidemiology, disease models, the basic reproduction number, and graph theory. Historical information in the field of epidemiology is presented, highlighting several of the main contributors to this area of interest, followed by an overview of an established disease model, SIR (susceptible-infectious-removed). Next, the basic reproduction number, R_0 is formally defined and discussed. The remainder of Chapter 2 focuses on graph theory concepts and definitions. Chapter 3 presents and discusses the experimental results related to the analysis of graph structure and outbreak severity. The findings in Chapter 3 relate to Research Question i. Chapter 4 illustrates the importance of key individuals in a disease outbreak. Vaccination methods are simulated to address Research Questions ii and iii. Chapter 5 presents the main perspectives of this study and summarizes the research results.

CHAPTER 2

BACKGROUND

2.1. Epidemiology

Epidemiology is the study of health-related states in an effort to prevent and control health problems [54]. A primary focus of epidemiology is to determine the cause of disease and the means by which disease can spread. This assumes that diseases are not randomly distributed, but rather afflict specific individuals or populations who are at risk [38]. The traditional triangle of epidemiology, as shown in Figure 2.1, demonstrates that communicable diseases involve an agent, a host, and an environment. The *agent* is the underlying cause of the disease, the *host* is the organism that carries the disease, and the *environment* is composed of the surroundings and conditions that make it possible for the disease to propagate over time [51, 54].

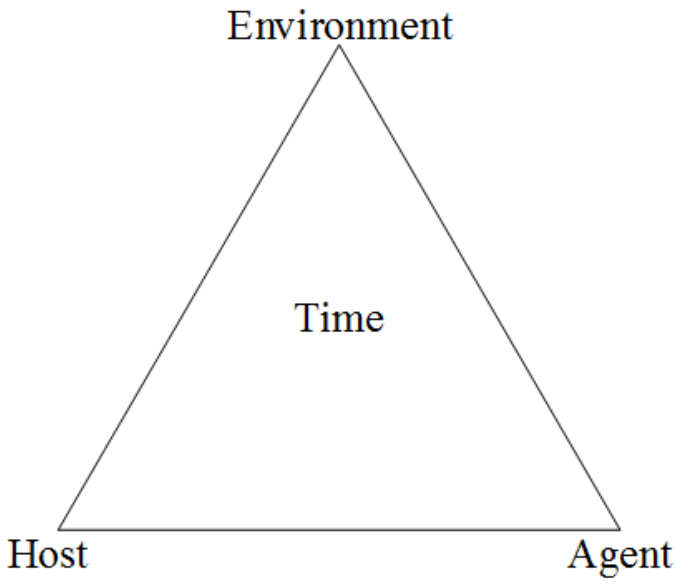


FIGURE 2.1. Epidemiologic triangle [54]

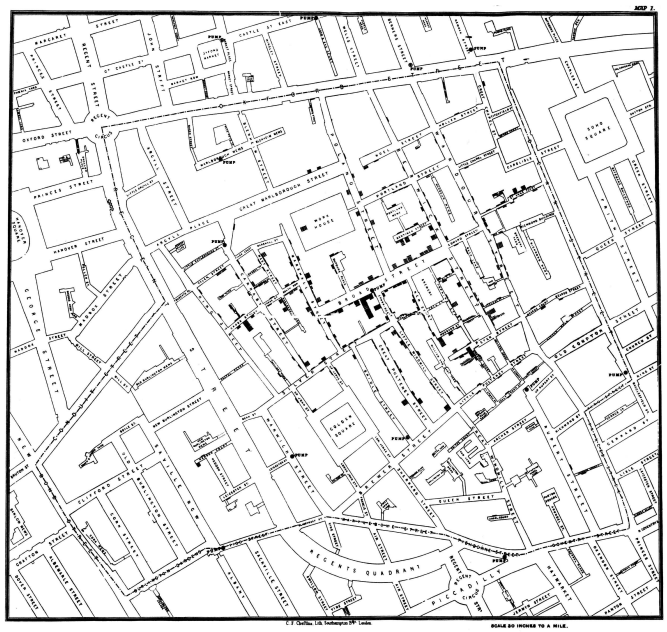
Epidemiology dates back to the time of Hippocrates (460-377 BC), who is considered to be the father of modern medicine and the first epidemiologist [37, 54]. Many of his aphorisms are still in use today, such as “As to diseases, make a habit of two things—to help or at least to do no harm.” Possibly the most important contribution that Hippocrates made to the field of epidemiology is that of observation. He believed that as time passed, physicians would be able to predict the diseases that would likely affect the local population and when those diseases could be expected.

John Graunt (1620-1674) added to the field of epidemiology and demographics by studying death records in London in 1603 [16, 54]. He was the first to estimate life expectancies and thereby establish the area of vital statistics to the field of epidemiology. Graunt developed a systematic technique for understanding diseases and causes of death that contributed to the modern methods that are still in use today.

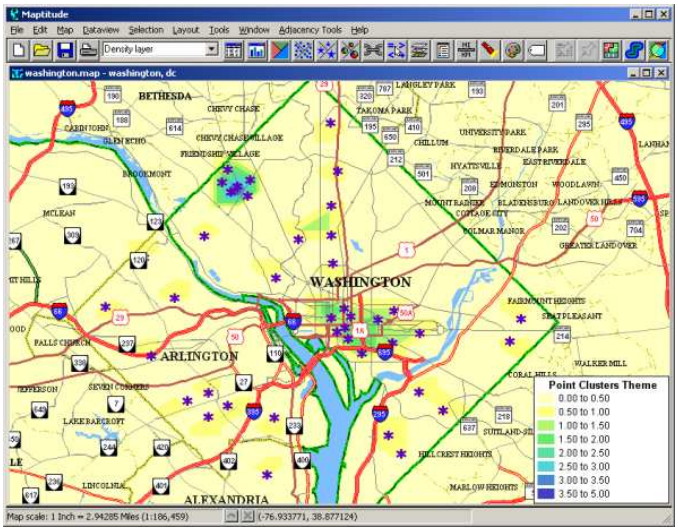
Thomas Sydenham (1624-1689), sometimes referred to as the English Hippocrates, made major contributions to the field of medicine by classifying fevers and identifying diseases along with methods of treatment [31, 54]. Sydenham, like Hippocrates, endorsed an empirical approach to medicine. He recognized that science is, and always will be, incomplete. This emphasizes the need to rely on common sense in addition to pure factual knowledge. He is credited with stating, “Investigate first, explain afterwards if you like; but remember that nature is always something very much greater than all your explanations” [31].

Another respected physician who made a major contribution to the field of epidemiology was John Snow (1813-1858) [54]. In the mid-1800’s, a large cholera outbreak occurred in London. By spot mapping the cholera cases onto a map of London streets, Snow was able to discern a pattern in the outbreaks. This pattern led to the conclusion that the cholera infections could be attributed to the water being drawn from the Broad Street pump. John Snow’s study was instrumental in demonstrating the importance of tracking diseases by spacial data [49]. Just as with Hippocrates, Snow’s success came largely due to careful observation and record-keeping, sound epidemiologic practices that are still relevant today. Current technology is capable of generating similar spot maps that can be used to track disease outbreaks. The use of dynamic graphics, implemented with

global information systems (GIS) software, allows a point pattern analysis to be mapped onto a case histogram [61]. The histogram can then display the number of cases, both by spacial and temporal occurrences. This system of disease tracking is comparable to that of Snow's, as depicted in Figure 2.2.



(a)



(b)

FIGURE 2.2. (a) Snow's cholera map [21]; (b) Maptitude GIS system [3]

The list of contributors to the field of epidemiology is long and varied. The few mentioned here, along with many others, established the foundation of this discipline. Researchers used resources that were currently available to learn about diseases and epidemics. Although the methods previously established are still valuable, technological advances have provided new tools that earlier scientists could have never imagined. Continued development in the field of epidemiology will undoubtedly rely heavily on the use of technology. In addition to analyzing historical data, disease outbreaks can now be studied theoretically through simulation. Computational epidemiology is a relatively new domain that is certain to become a core component of epidemiologic research.

2.2. Disease Models

To better understand difficult computational systems that model disease outbreaks, it is helpful to first look at a widely accepted elementary model known as SIR (susceptible-infectious-removed) [6, 46]. Initially in the SIR model, the majority of the population falls in the susceptible category. As the disease spreads, individuals move from susceptible to infectious and from infectious to removed, as represented in Figure 2.3. The bell-shaped curve shown in Figure 2.4 demonstrates the rise and fall of the number of individuals in the infectious group over the course of an epidemic. Attributes of the graph are indicative of the severity of the outbreak. The duration of an epidemic is measured from the initial infectious case until there are no longer any infectious individuals. The basic reproduction number, as discussed in Section 2.3, is measured at the beginning of the outbreak. The area under the curve is directly related to the total number of infectious individuals, however, the area must be divided by the infectious period to obtain an accurate estimate. The basic SIR model makes the following assumptions:

- The population density remains constant. Births, deaths, and immigration are ignored.
- The population mixes homogeneously. That is, contacts between any two individuals are equally likely to occur.
- An individual moves directly from the susceptible state into the infectious state.
- Once an individual enters the removed state, they remain in that state.

These assumptions form a solid foundation for disease modeling, however, many models are based on the SIR paradigm to develop more complicated systems that include additional parameters and relationships. It is quite common to find models that include one or more of the modifications listed below:

- The population mixes in a non-homogeneous manner. Contacts among individuals are based on demography and/or geography.
- Additional states are incorporated, such as latent, exposed, or symptomatic.
- Individuals are allowed to become susceptible again after they have recovered from the illness.

FIGURE 2.3. SIR schematic

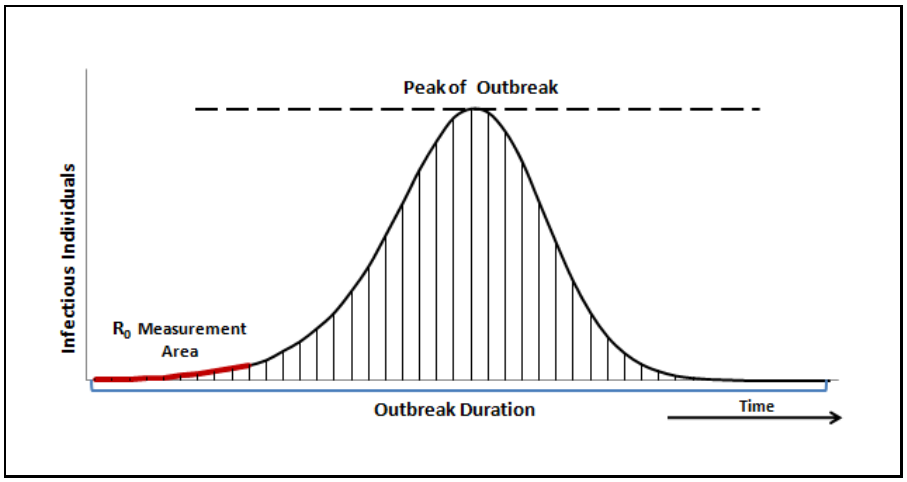


FIGURE 2.4. Epidemic curve

The SIS model is a modification of SIR as shown in Figure 2.5 (a). This model is appropriate when the disease under investigation is such that infected individuals recover, but do not develop immunity to the disease. The SIS model is a modification of SIR through the elimination of the removed state. Alternately, the SIR model can be extended through the addition of one or more

states. Figure 2.5 (b) illustrates the SEIR model which includes an *exposed* or *latent* state. SEIR is a suitable model for infectious diseases in which individuals enter a latent stage before becoming infectious. Both SIS and SEIR are valid adaptations of the SIR model [5, 36, 42, 48].

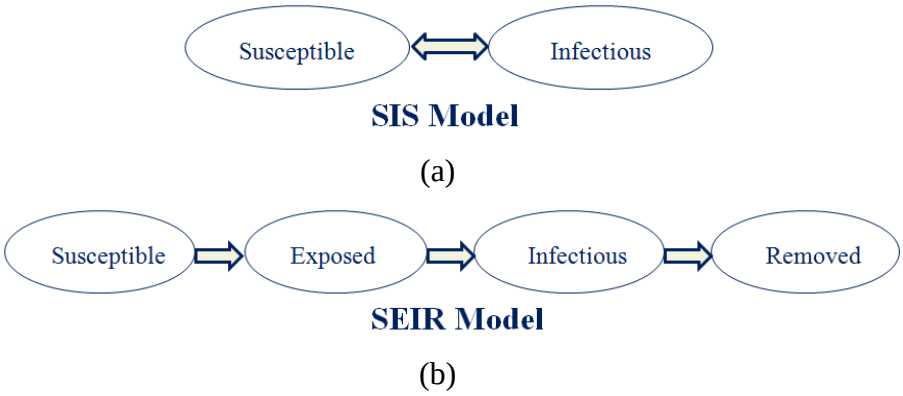


FIGURE 2.5. Variations of the SIR model

2.2.1. Computation Models in Epidemiology

The spread of an infectious disease through a population depends, to a large extent, on randomness. The chance occurrence between an individual who is infectious with one who is susceptible, the probability that the disease will transfer, and further propagation to other members of the population lead to an increasingly intractable set of events. A Monte Carlo simulation is a statistical sampling technique that is applicable to problems that have high levels of uncertainty [55, 56]. Monte Carlo methods are well-suited to model disease spread and have been used extensively in this area [11, 24, 13, 59, 58].

Implementation of computational models vary; common methods include cellular automata, agent-based systems, contact models, and hybrid approaches. Cellular automata models (CA) consist of a grid of individual cells, a set of cell states, a neighborhood definition, and a set of transition rules [69]. Each cell in a CA begins in a particular state and may change state depending on the transition rules and the state of neighboring cells. Agent-based models are designed with interacting, autonomous agents that each follow a set of rules [50, 68]. The agents can display specific behaviors and are able to react to their environment and/or other agents in the system. Agent-based

models are often computationally expensive and may require parallelization to work efficiently. Contact models focus on the probability of contacts between individuals. In a contact network, vertices represent one or more individuals and edges represent contacts between individuals [57]. Hybrid models combine one or more computational models.

Mikler et al. developed a global stochastic cellular automata model (GSCA) which addresses two of the limitations of the basic SIR model, that of homogeneous mixing and the lack of a latent period [59]. The implementation of GSCA overcomes the problem of neighborhood saturation found in classical CA models by allowing global contacts in addition to typical neighborhood contacts. The GSCA has been used to model influenza, conjunctivitis, and the common cold. A hybrid approach contact simulator, the global stochastic contact model (GSCM), was also developed at the University of North Texas [58]. Like the GSCA, the GSCM includes a latent period and it also incorporates a symptomatic state which allows a behavioral change to occur once an individual realizes that they have become infected. Increased contacts between individuals resulted in an elevation of the number of infected individuals and a decrease in outbreak duration in simulations conducted with the GSCM. Both the GSCA and the GSCM were designed to incorporate geographic and demographic dimensions of the population under study.

EpiSims, developed in 2004 by researchers at the University of Maryland and the Los Alamos National Laboratory in New Mexico, is another example of a SIR-based computational model [27]. EpiSims is an agent-based system that simulates a population of individuals, each following a specific daily schedule. The social interaction network is represented as a bipartite graph consisting of a set of nodes which represents people and a set of nodes which represents locations. Transfer of disease is only possible between susceptible and infectious individuals when contacts are made at a particular location during the same time frame.

EpiSimdemics, an SEIR model developed at Virgina Tech as an extension of EpiSims, broadened the scope to include large, realistic social networks by making adjustments to parallelize the code [10]. The input data set for EpiSimdemics is approximately 100GB and includes data from the U.S. Census for demographics, NAVTEQ for road network information, Dun and Bradstreet

(a commercial database) for business, the National Household Transportation Survey for scheduling individuals, and the Digest of Education Statistics for school locations and enrollment. The algorithm used to simulate an outbreak is a simple discrete event simulation, which implies that the system only changes state when an event occurs. The system is composed of both people and locations. An example of an event is a particular person leaving a specific location at a given time. If an infectious person and a susceptible person are at the same location at the same time, there is a possibility for transfer of the disease. The model includes realistic states in addition to the fundamental SEIR, such as vaccinated and asymptomatic. EpiSimdemics has been used for real studies by the Department of Homeland Security, the Department of Defense, and the Department of Health and Human Services.

Closely related to EpiSims and EpiSimdemics, researchers at Virgina Tech have more recently developed EpiFast, a parallel agent-based SEIR model [13]. EpiFast has many of the same features as EpiSimdemics, but it executes much faster as the model is significantly less complex. Unlike EpiSimdemics, in which state changes can occur every hour, EpiFast measures discrete time steps by the day. EpiFast also increases speed by using a pre-constructed people-people contact network, i.e. social network. Intervention strategies include vaccination, individual behavioral changes, and facility closures. On similar networks, EpiFast was shown to execute ten times faster than EpiSimdemics .

2.3. The Basic Reproduction Number

Individuals who become infected at the onset of an outbreak play a key role in the progression of a disease. For an epidemic or pandemic to occur, the rate of increase in the number of newly infected individuals at the beginning of the outbreak must exceed an epidemic threshold referred to as the basic reproduction number, R_0. The value of R_0 provides an indication of the transfer of specific disease pathogens as well as the conduciveness of environmental conditions.

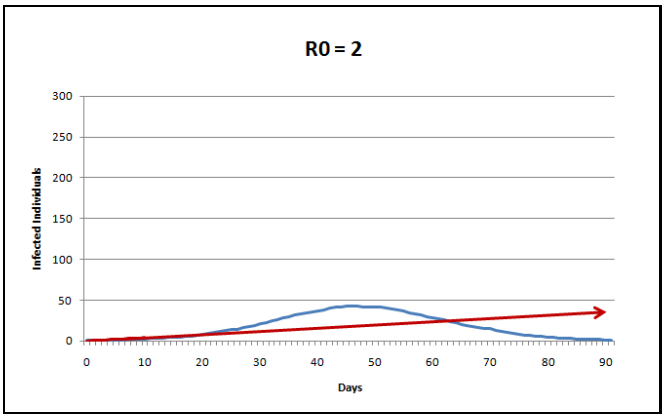
DEFINITION 2.1. The basic reproduction number, R_0, is defined as the average number of expected secondary cases produced from a single primary infectious case in a completely susceptible population.

R_0 is an established epidemiologic indicator used to estimate the probability that an infectious disease will create an epidemic or pandemic [6]. $R_0 > 1$ indicates that an epidemic or pandemic is likely to occur because, on average, every infectious person will transfer the disease to more than one other person; therefore, the disease will continue to spread. A value of $R_0 < 1$ suggests that the disease spread cannot be maintained and should die quickly with relatively few individuals infected.

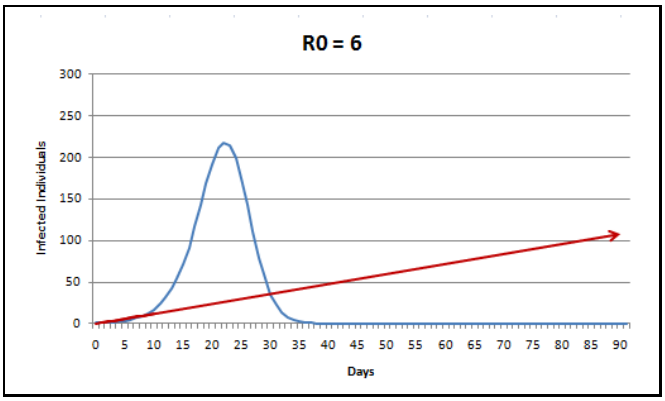
It is unlikely that the primary infectious case and the secondary infections resulting from that case can be identified, however, estimates of R_0 are generally based on data collected near the beginning of an outbreak since the majority of the population is susceptible at that time. R_0 provides an indication of how quickly a disease will spread throughout a population and is related to a trendline based on initial data from an outbreak. The graphs in Figure 2.6 were created by simulated outbreaks with estimated R_0 values of 2, 6, and 10. Each outbreak curve is accompanied by a linear graph $y = \frac{R_0}{d}x$. In this linear equation, d represents the infectious period and x represents a single day of the outbreak. Because an infectious individual has the potential to create secondary infections over d days, dividing by d normalizes the trendline to the outbreak curve which measures infected individuals *per day*. Note that as the R_0-related slope increases, the outbreak curve becomes taller and the duration of the outbreak decreases. As R_0 increases, the disease spreads more rapidly throughout the population which results in an increase in infectious individuals (a taller peak) and a decrease in the time it takes the epidemic to run its course (a shorter duration).

2.3.1. The Importance of Understanding R_0

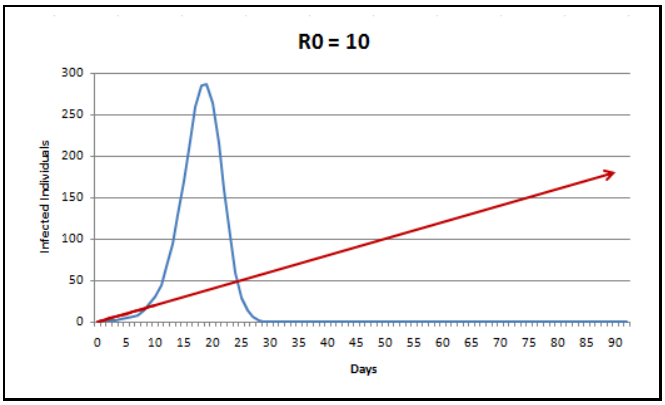
From a public health point of view, a clear understanding of R_0 is beneficial in determining a course of action when a disease is introduced into a susceptible population. A valid estimate of R_0 provides an indication of the force of a specific disease. The R_0 estimate can be used to guide implementation of accepted intervention strategies to prevent the outbreak from progressing into an epidemic or pandemic. Estimates of R_0 have been calculated for past outbreaks of malaria, tuberculosis, SARS, and Spanish influenza (see Table 2.1) [20, 19, 67, 71, 77].



(a)



(b)



(c)

FIGURE 2.6. Outbreak graphs and R_0-related slope lines.

The 2002-2003 SARS outbreak was kept under control largely due to early diagnosis and patient isolation [20]. Chowell et al. used a variation of the SIR model (SEIJR, which includes exposed and diagnosed individuals) along with regional and global data to determine the effects that model parameters have on R_0. In another article, Chowell and others studied the Spanish

TABLE 2.1. R_0 Estimates of Past Outbreaks

Disease	R_0 **Estimate**	**Author**
Malaria	$1 - 3,000+$	Smith [71]
Tuberculosis	$1.10 - 31.26$	Sanchez [67]
SARS	$0.24 - 2.47$	Chowell [20]
Spanish influenza	$1.20 - 7.50$	Vynnycky [77]

flu outbreak in Geneva, Switzerland [19]. An epidemic model and hospital records were used to estimate R_0 for the first and second waves of the pandemic spread.

In the paper, "Appropriate Models for the Management of Infectious Diseases," Wearing et al. stress the significance of accurately determining the latent and infectious periods in mathematical models [81]. This paper suggests that common methods for determining these two parameters are often incorrect, resulting in an underestimate of R_0 and thus misguided efforts to control an outbreak.

Farrington and Whitaker recognize the significant role of medical intervention in lowering the *effective* reproduction number, $R_e(t)$ [29]. Two sets of serological studies were compared, one with data from 1987, prior to the introduction of the measles, mumps and rubella (MMR) vaccine and one in 1996, "post-vaccination". The results show a marked decrease in the estimate of both R_0 and $R_e(t)$.

Based on survey data from two military ships and five Maryland communities, White and Pagano estimate the effective reproductive number each day of the 1918 influenza outbreak applying two distinct likelihood methodologies [83]. The first method presented by White and Pagano, models R_i (the effective reproduction number on day i) parametrically as a function of time. The second method, described by Wallinga and Teunis [78], is expressed as a probability, p_{ij}, that case i was infected by case j, accounting for the time difference between the initial onset of symptoms for both cases. The first method, based on four parameters, can be generalized to other settings. The second method produces results that follow the same pattern as the epidemic curve. Estimates for the Maryland communities range from 1.34 to 3.21. The average estimate for the two ships is

slightly higher at 4.97. This higher value of R_0 may be attributed to the close living quarters on the ships resulting in more frequent contacts.

2.3.2. Deriving R_0 Mathematically

The fundamental equations of the SIR model can be used to derive R_0 mathematically. The established differential equations below represent the movement from susceptible to infected to removed. The constant α is a probability that describes the likelihood of disease transfer. The constant γ is the removal rate, which is the reciprocal of the average number of days in the Infected state. The three SIR differential equations are defined as follows:

(1) $$\Delta S = -\alpha S_t I_t$$

(2) $$\Delta I = \alpha S_t I_t - \gamma I_t$$

(3) $$\Delta R = \gamma I_t$$

The SIR equations correspond directly to Figure 2.4. The negative sign in Equation (1) indicates that as the disease spreads, the number of susceptibles decline. Likewise, the number of removed individuals, Equation (3), increases. The number of infected individuals initially increases and then decreases following a bell-shaped curve. Equation (2) provides the basis for the calculation of R_0. If the rate of infection is faster than the rate of removal ($\Delta I > 0$), for some time, t, an epidemic occurs. Factoring γI_t from Equation 2, the change in infected individuals over time becomes:

(4) $$\Delta I = \gamma I_t \left(\frac{\alpha S_t}{\gamma} - 1 \right)$$

It is now evident that if $\frac{\alpha S_t}{\gamma} > 1$, the number of infected individuals will increase. The mathematical definition of R_0 is taken directly from Equation 4. Because R_0 is measured at the beginning of the outbreak ($t = 0$), the definition is as follows:

(5) $$R_0 = \frac{\alpha S_0}{\gamma} = (\alpha S_0)\left(\frac{1}{\gamma}\right)$$

In Equation (5), S_0 represents the initial population of susceptibles and αS_0 represents the number of new infections per infected individual. This value is then multiplied by the average duration of infectivity, $\frac{1}{\gamma}$, because an infectious individual can continue to infect others as long as they remain infectious.

2.3.3. Experimental Expected R_0

Based on the mathematical definition of R_0 shown in Equation 5, an expected value of R_0 can be derived to validate computational models. In the mathematical equations of Section 2.3.2, αS_t represents the probability of disease transfer from infected individuals to susceptible individuals at time t. The computational model used in this research replaces αS_0 with a Contact Rate (CR) multiplied by a Transmission Rate (TR). The number of days infectious (DaysI) is equivalent to $\frac{1}{\gamma}$. Equation 6 demonstrates the equivalence between the mathematical value of R_0 and the experimental expected value.

(6) $$R_0 = (\alpha S_0)\left(\frac{1}{\gamma}\right) = (CR)(TR)(DaysI)$$

Because R_0 is a measure of secondary infections, it can be concluded that an infectious individual will infect others based on the transmission rate of the disease, how many contacts are made in a day, and the length of time the individual is infectious. The Ordinary Differential Equations (ODE) that describe an outbreak can be compared with a computational model based on the contact rate, transmission rate, and number of days infectious using Equation 6 as a basis. To construct the comparison, α and γ are calculated as shown in Equations 7 and 8. The daily ODE values of SIR are then calculated using a spreadsheet. The chart in Figure 2.7 is a comparison using the infectious column from an ODE spreadsheet and the daily infectious count from a comparable simulation.

TABLE 2.2. Comparable Disease Parameters for ODE and Computational Model

Simulation	ODE
$N = 500$	$S_0 = 499$
$CR = 20$ $TR = 0.03$	$\alpha = 0.0012$
$DaysI = 4$	$\gamma = 0.25$

The simulated outbreak was averaged over 100 independent simulations. The parameters used for this comparison are shown in Table 2.2.

(7) $$\alpha = \frac{(CR)(TR)}{S_0}$$

(8) $$\gamma = \frac{1}{DaysI}$$

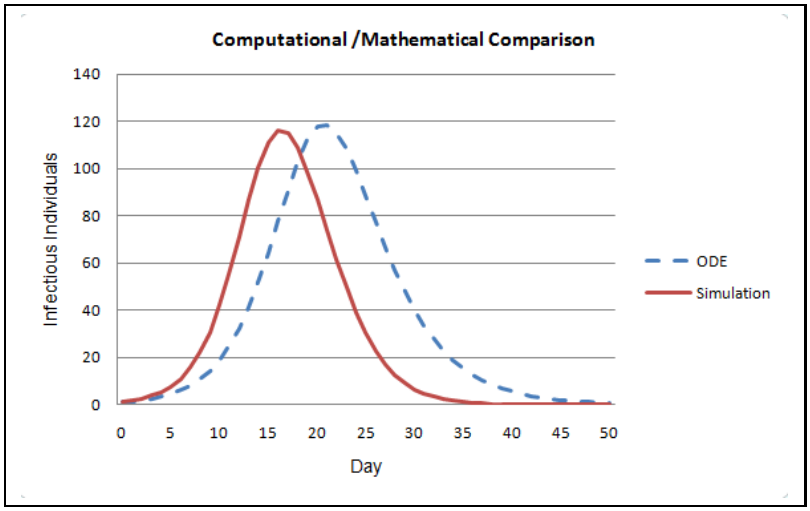


FIGURE 2.7. Comparison between ODE and computational model

Figure 2.7 illustrates that the infectious curve of the ODE lags behind that of the simulation. This is likely attributed to the fact that the infectious count is reduced each day by γ in the ODE, whereas each infectious individual in the simulation remains in the count until after they have been in the infectious state for DaysI. This causes a discrepancy that becomes apparent early in the oubreak comparison.

2.4. Graphs

Many real-world systems naturally correlate to graph structures. Examples include the Internet and World Wide Web topology [4, 9, 84], mobile devices [18, 75], road networks [25, 45], biological systems [30, 73, 76], and a host of other domains. Graph structures are inherently well suited to describe social networks and have been used extensively in this field [15, 17, 23, 65, 66, 79]. A major benefit to using graphs is the ability to reduce a complex system to a simplified model of entities and relationships.

DEFINITION 2.2. A graph is formally described as a tuple $G = (V, E)$ in which V is a set of vertices and E is a set of edges [82].

The vertices of a graph, also referred to as nodes, represent a set of entities, such as people, locations, or objects. The edges in a graph represent a relationship between two nodes. On a map, the nodes could correspond to cities and the edges, roads. In a biological system, a graph may symbolize proteins and protein interactions, metabolic networks, or various other life structures. In a social network, the nodes of a graph are individuals and the links between individuals represent some sort of relationship. For the research presented herein, the links represent contacts in the social network and disease transmission in the outbreak graph.

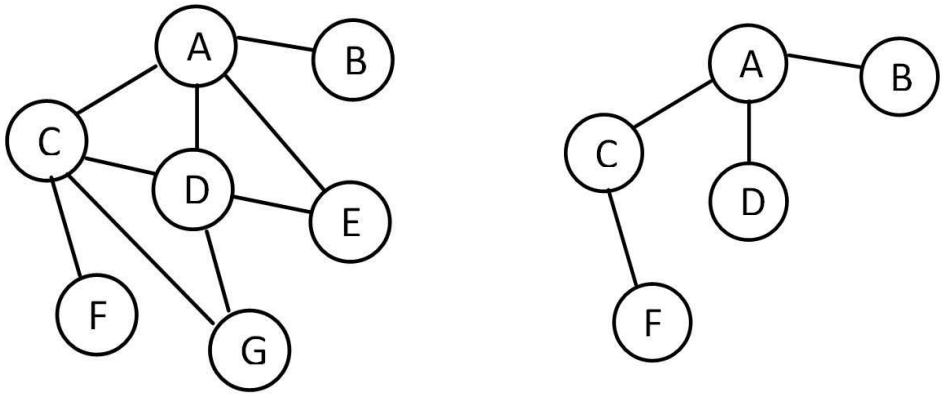
2.4.1. Graph Theory Concepts

Two vertices in a graph are *adjacent* if an edge exists between them. An *adjacency matrix*, A_{ij} of a graph is an n by n matrix representation of a graph of size n in which each entry in the matrix represents a value describing the relationship between nodes i and j. In a non-weighted graph, each adjacency matrix entry is either a 1 or a 0. A value of 1 indicates that node i is adjacent to node j and a value of 0 signifies that the two nodes are not adjacent. Table 2.3 illustrates the adjacency matrix of the graph in Figure 8(a). A *loop* is an edge that connects a vertex to itself. Edges that connect the same pair of vertices are referred to as *multiple edges*. A graph that contains no loops or multiple edges is a *simple graph*. If numerical values are assigned to the edges, the graph is considered to be a *weighted graph*. The edge weight may refer to cost, distance, or any

other relationship between nodes. A *subgraph* of G is a graph G', such that $V(G') \subseteq V(G)$ and $E(G') \subseteq E(G)$. Figure 2.8 illustrates the relationship between a graph and a subgraph. The graphs shown in Figure 2.8 are *undirected graphs*, indicating that the relationship between two nodes is identical, such as *"is a relative of"* or *"lives in the same neighborhood"*. A *directed graph*, or *digraph* on the other hand, suggests a directional relationship, such as "is the child of" or "passes the disease to". Undirected graphs and digraphs are both useful tools, but serve different purposes. In regard to disease simulation, a social network is an undirected graph, but the transmission of disease from one person to another inherently implies direction. The difference between an undirected graph and a directed graph is illustrated in Figure 2.9.

TABLE 2.3. Adjacency matrix of graph in Figure 2.8(a)

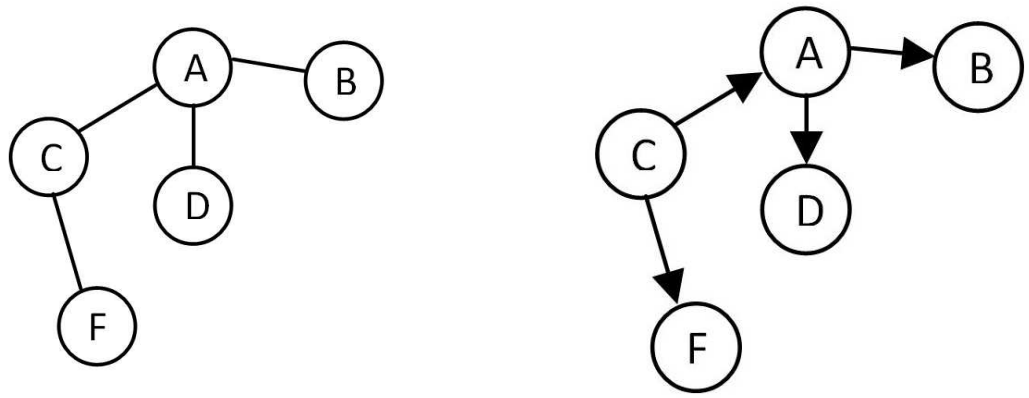
	A	B	C	D	E	F	G
A	0	1	1	1	1	0	0
B	1	0	0	0	0	0	0
C	1	0	0	1	1	0	1
D	1	0	1	0	1	0	1
E	1	0	0	1	0	0	0
F	0	0	1	0	0	0	0
G	0	0	1	1	0	0	0



(a) (b)

FIGURE 2.8. (a) Graph, G; (b) G', a subgraph of G

The *degree* of a node denotes the number of incident edges. In a digraph, there are both an *in-degree* and an *out-degree* indicating edges coming in and edges going out, respectively. Closely



(a) (b)

FIGURE 2.9. (a) Undirected graph; (b) directed graph

related to the degree of a node is the *density* of a graph. The number of edges, m, in a graph with n nodes, excluding self-loops, is bounded by Equation (9). The density of a graph, $D(G)$, is the ratio of the number of edges present, m, to the maximum number of edges possible as shown in Equation (10). The density of a graph ranges from 0, if no edges are present, to 1, if the maximum number of edges are present.

(9) $$m \leq \frac{n(n-1)}{2}$$

(10) $$D(G) = \frac{m}{n(n-1)/2} = \frac{2m}{n(n-1)}$$

Another important aspect of graphs involves paths between pairs of nodes. The shortest path between two nodes is referred to as the *geodesic* path. The largest geodesic distance between a given node and all other nodes in a graph is known as the *eccentricity* of the node. The largest geodesic distance between any two vertices in a connected graph is called the *diameter*. The diameter can also be described as the largest eccentricity of all nodes in a graph.

2.4.2. Random Graphs, Ordered Lattices, Hypergraphs, and Small-World Graphs

In 1951, Solomonoff and Rapoport described structures referred to as *random nets* [72]. In 1960, Erdös and Rényi continued the investigation of *random graphs* [26], as shown in Figure 2.10. Two distinct methods for building random graphs were described. One begins with a fixed number of vertices, n, and a fixed number of edges, m. The edges are randomly selected out of the $\frac{n(n-1)}{2}$ that are possible. Using this technique, there are $\binom{\frac{n(n-1)}{2}}{m}$ equiprobable random graphs that can be constructed. The alternate definition is one in which the number of vertices are fixed, but the edges are selected randomly with probability p. The number of edges using this technique is a random variable. Therefore, to develop a graph with an average of m edges, the value of p should be set to $\frac{m}{\binom{n}{2}}$. For example, to construct random graph with an average of 5 edges in a graph with 10 nodes, the probability p that an edge exists between two vertices is $\frac{1}{9}$ as calculated below:

$$\frac{5}{\binom{10}{2}} = \frac{5}{45} = \frac{1}{9}$$

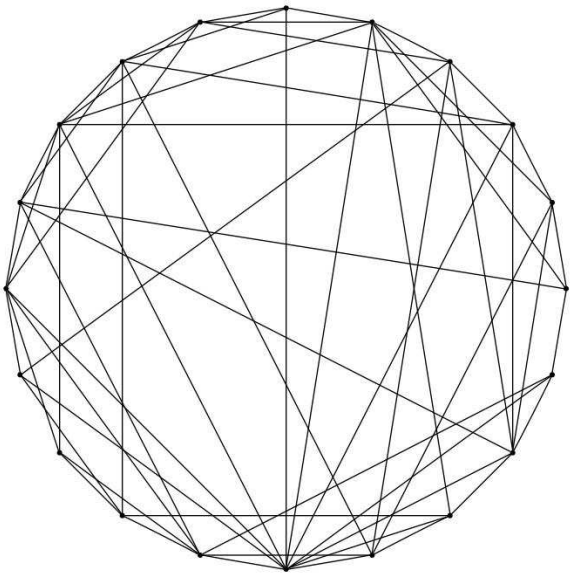


FIGURE 2.10. Random graph example

Contrary to a random graph, a completely ordered graph (also called a regular graph or a lattice) is a graph in which each node is linked to k of its immediate neighbors. To visualize such a graph, it is helpful to think of the vertices in aligned in a circular fashion as demonstrated in Figure 2.11.

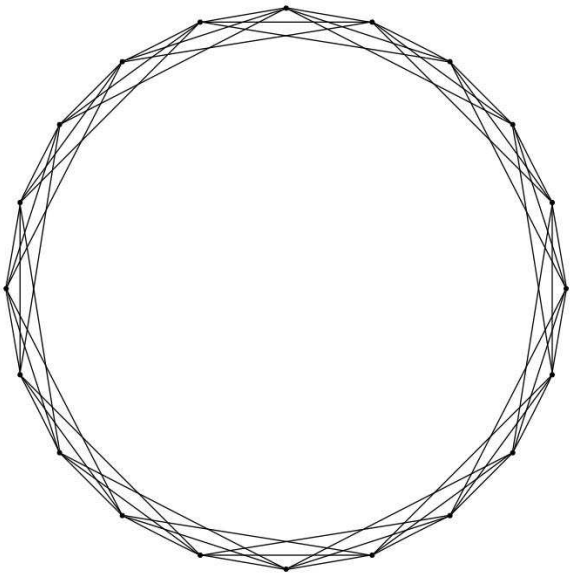


FIGURE 2.11. Ordered graph example

Although regular graphs represent contacts with neighbors, other interactions that are known to exist in real social networks are not accurately represented. Such interactions are those that frequently occur in daily life when meetings occur between two people who are not normally in the same social circle. This includes contacts that take place while in a grocery store, on vacation, riding in public transportation, standing in line at an event, or a number of other similar situations. Even in very large social networks, the small-world effect theorizes that any two people are connected by a relatively short chain of intermediate contacts [60]. A small-world graph is a model based on the small-world effect. It is a structure that falls between a random graph and an ordered lattice, exhibiting the clustering behavior of an ordered graph while maintaining the small-world property observed in random graphs. First introduced in the mid-1950's, small-world graphs gained scientific popularity after a publication by Watts and Strogatz in 1998 [70]. Since that time, many researchers have explored the properties and applications of small-world graphs [22, 43, 47, 52, 60].

A small-world graph can be easily constructed from an ordered lattice by rerouting some of the edges [80]. Each edge may be rerouted to another vertex based on some probability, p. A value of $p = 0$ results in a completely ordered graph and a value of $p = 1$ creates a random graph. A

small-world graph results when $0 < p < 1$. Figure 2.12 is an example of a small-world graph. Note that when an edge is rerouted, loops and multiple edges are prohibited resulting in a simple graph.

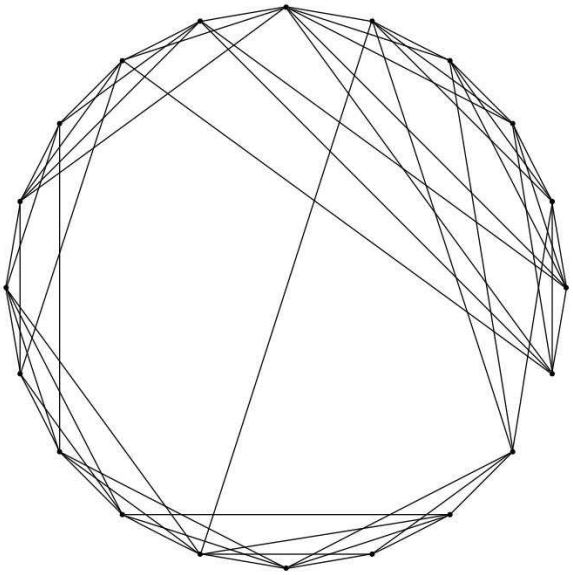


FIGURE 2.12. Small-world graph example

A hypergraph, as shown in Figure 2.13, is another way to represent a social network. Hypergraphs consider ties among subsets of individual nodes; that is, edges can connect any number of vertices rather than joining only two nodes [44]. Thus, the formal description of the graph changes. The graph can now be described as $H = (V, E^h)$ in which E^h refers to a set of hyperedges. Each hyperedge is a subset of the vertex set. Hypergraphs are appropriate for affiliation networks, or membership networks, in which the connection among individuals may represent those who belong to the same social group or club [79].

The graphs under study for this investigation cover a wide spectrum of those that could be considered for social network representation. The random graph provides a small degree of separation between any two nodes, but does not display the clustering effect typically found in social networks. Conversely, the ordered lattice captures the clustering effect, but does not maintain the small degree of separation. Both the small world network and the hypergraph more accurately depict the characteristics that are likely to be found in true social networks.

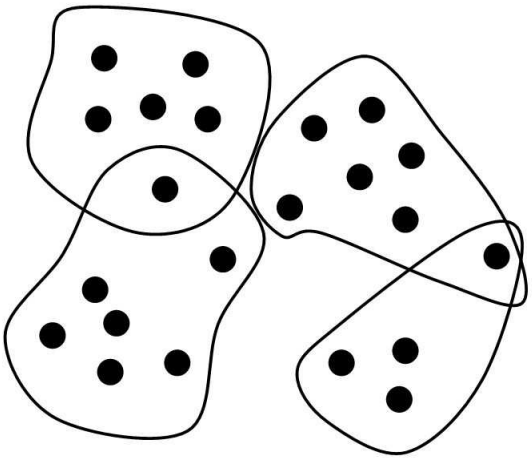


FIGURE 2.13. Hypergraph example

2.4.3. Centrality Measures

Point centrality, also referred to as node centrality, is used to determine which nodes are the most *important* in a graph [33]. Importance, of course, is relative to the purpose of the graph. For the purpose of this research, importance refers to the ability to transfer a disease. Of four centrality measures outlined by Wasserman and Faust in "Social Network Analysis" [79], Degree, Closeness, and Betweenness were implemented for this research. Although there are other measures of centrality, these three were selected to represent the structure of a graph. The initial software for this research was validated using the centrality indices for Padgett's Florentine families as shown in [79]. A description and example of each of these centrality measures is outlined below.

Degree Centrality. Degree centrality is the most straightforward to compute because it is simply a count of the number of edges incident to a node. An individual with more connections to other individuals may be deemed more important. Degree centrality can be calculated for a point in a graph of size n as shown in Equation (11). $C_D(i)$, the degree centrality of node i, is the sum of all adjacent nodes as indicated by A_{ij}, the adjacency matrix. In Figure 2.14, it is easily observed that Node *4* is of degree 8 and could be regarded as the most important node in the network. However, a degree of size 8 in a much larger graph might be comparatively small. It is common to normalize centrality measures to produce a value that is independent of the graph size. The largest possible degree of a node in a graph of size n is $n - 1$. Therefore, Equation (12) can be used to calculate

a degree centrality that is normalized to the size of the graph. Since the network displayed in Figure 2.14 has 11 nodes, each vertex degree is divided by 10 to give the relative degree centrality, $C_D^{'}(i)$, as shown in Table 2.4.

In the context of disease spread, an individual who has a high degree centrality makes contacts with more individuals in the population. If an individual with high degree centrality becomes infected with a disease there is greater opportunity for the disease to propagate. The degree centrality of individuals infected early in an outbreak is of even greater importance as this will have an effect on the value of R_0 and may determine whether of not an epidemic emerges.

$$C_D(i) = \sum_{j=1, j\neq i}^{n} A_{ij} \tag{11}$$

$$C_D^{'}(i) = \frac{\sum_{j=1}^{n} A_{ij}}{n-1} = \frac{C_D(i)}{n-1} \tag{12}$$

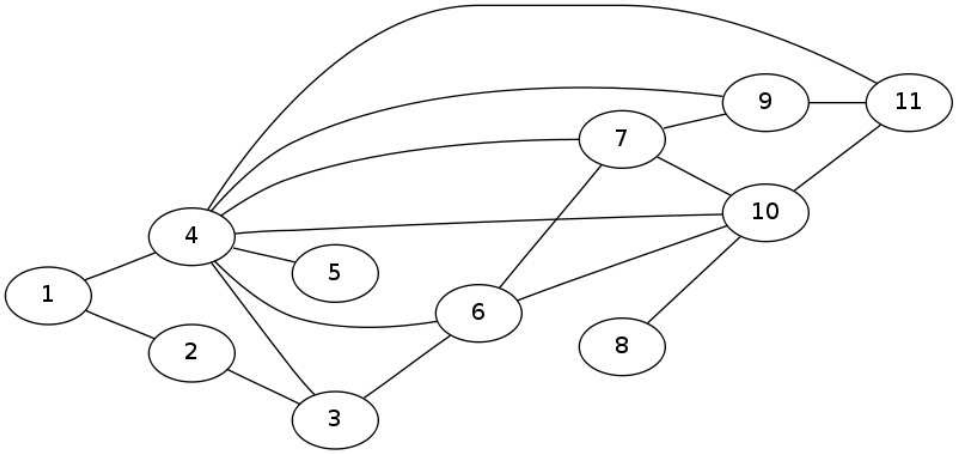


FIGURE 2.14. High level of degree centrality illustrated by Node *4*

Closeness Centrality. Closeness centrality is dependent upon the geodesic path from one vertex to another. It is a measure of the distance from a particular node to every other node in the graph followed along a geodesic network path. A large closeness value indicates that a vertex can

TABLE 2.4. Relative degree centrality measures for vertices in Figure 2.14

v	$C_D'(v)$	v	$C_D'(v)$
1	0.2	7	0.4
2	0.2	8	0.1
3	0.3	9	0.3
4	0.8	10	0.5
5	0.1	11	0.3
6	0.4		

quickly influence other nodes in the network. Unlike degree centrality, closeness centrality takes into account indirect as well as direct connections. It is reasonable to expect that the duration of an outbreak will be influenced if individuals with a high degree of closeness become infected, as these individuals are tightly connected to the rest of the population. In Figure 2.15, although several other nodes have a higher degree centrality, Node *5* is at most two hops from any other node in the network, making it the most important node based upon closeness centrality (see Table 2.5). Closeness centrality of a node, $C_C(i)$, is calculated as shown in Equation (13). In this equation, $d(i,j)$ refers to the geodesic distance from node i to node j. Some formulas for closeness do not take the reciprocal of the summation of the distances, however when a node is a greater distance away, the centrality should decrease. Therefore, the geodesic distances should be weighted inversely. The maximum closeness value is obtained when a node is directly connected to every other node in the network. In a network of size n, the maximum closeness is $\frac{1}{n-1}$. Thus, a relative closeness centrality, $C_C'(i)$, is calculated by multiplying by $n-1$ as shown in Equation (14).

(13) $$C_C(i) = \frac{1}{\sum_{j=1, i \neq j}^{n} d(i,j)}$$

(14) $$C_C'(i) = \frac{n-1}{\sum_{j=1, i \neq j}^{n} d(i,j)} = (n-1)C_C(i)$$

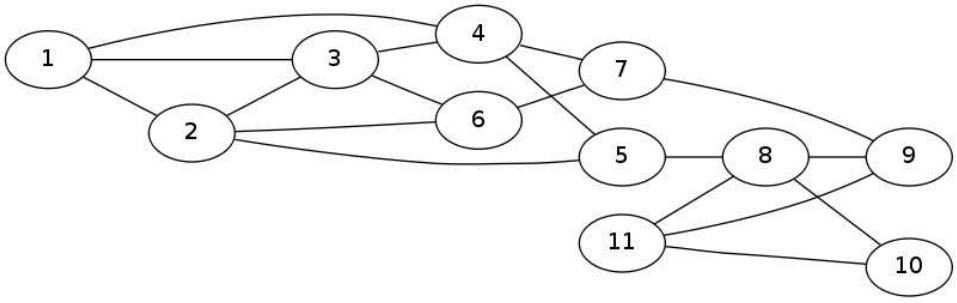


FIGURE 2.15. Closeness centrality illustrated by Node *5*

TABLE 2.5. Relative closeness centrality measures for vertices in Figure 2.15

v	$C_C^{'}(v)$	v	$C_C^{'}(v)$
1	0.43	7	0.56
2	0.53	8	0.53
3	0.46	9	0.50
4	0.56	10	0.37
5	0.59	11	0.42
6	0.48		

Betweenness Centrality. Similar to closeness centrality, betweenness centrality is based on network paths. A node has a high *betweenness centrality* if it falls on a large proportion of network paths when all paths are considered. Nodes with a high betweenness centrality assert more control over the flow of information across a network. Note that in Figure 2.16, Node *6* falls on every path that connects the left side of the graph to the right side of the graph. If this node is removed from the network, the graph will become disconnected. Node *6*, therefore, is central to the flow of information in this graph and has the highest betweenness measure as shown in Table 2.6.

Individuals or groups in a population who possess a relatively high betweenness centrality provide important links by which a disease can spread. If these individuals or groups can be identified, it may be possible to use preventative strategies, such as vaccination or quarantine measures, to effectively prevent a disease from reaching large portions of a population. These strategies could essentially disconnect the population graph and prevent further spread.

Freeman suggests a method for calculating the betweenness of a point that incorporates the probability that the point will lie on a randomly selected geodesic path [32]. To determine the *partial betweenness* of point i on a path that connects points s and t such that $s \neq i \neq t$, let g_{st}

represent the number of geodesic paths from s to t and $g_{st}(i)$ the number of geodesic paths from s to t that contain i. Now the probability, $b_{st}(i)$, that point i lies on a randomly selected path from s to point t is shown in Equation (15).

To consider the overall betweenness centrality of point i which includes all geodesic paths in the network, $C_B(i)$, the sum of all partial betweenness values is calculated as shown in Equation (16). Since $C_B(i)$ is simply a count, the relative potential based on the size of the network is not taken into consideration. A relative betweenness value, $C_B'(i)$ as shown in (18), can be derived by expressing this value as a ratio of $C_B(i)$ to the maximum betweenness value possible in a network of size n. The maximum betweenness value, $maxC_B(i)$ as shown in Equation (17), occurs when a node *i* falls on every geodesic path connecting all nodes not including *i*.

(15) $$b_{st}(i) = \left(\frac{1}{g_{st}}\right)(g_{st}(i)) = \frac{g_{st}(i)}{g_{st}}$$

(16) $$C_B(i) = \sum_{s=1}^{n} \sum_{t=s+1}^{n} b_{st}(i)$$

(17) $$maxC_B(i) = \frac{[n(n-1)]}{2} - [n-1] = \frac{n^2 - 3n + 2}{2}$$

(18) $$C_B'(i) = \frac{2C_B(i)}{n^2 - 3n + 2}$$

Information Centrality. Information centrality is based on the same concept as betweenness centrality, but also considers the degrees of the nodes along each network path. When betweenness centrality is calculated, it is assumed that two geodesic paths are equally likely to be "chosen" and therefore the probability of each geodesic path is identical. It may be presumed that vertices along the path which have a high degree are more likely to be on a "chosen" geodesic path. An even

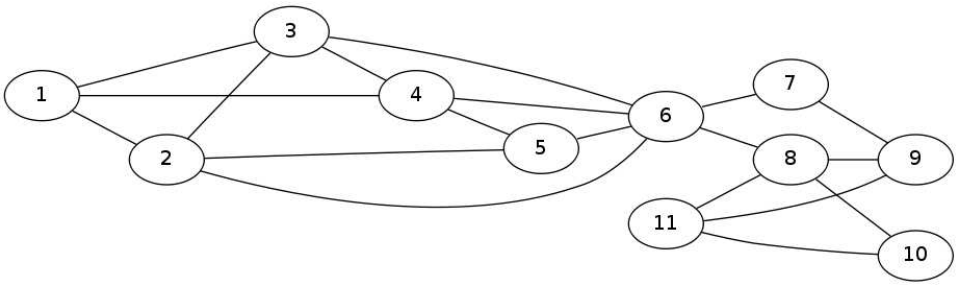


FIGURE 2.16. Betweenness centrality illustrated by Node 6

TABLE 2.6. Relative betweenness centrality measures for vertices in Figure 2.16

v	$C_B^{'}(v)$	v	$C_B^{'}(v)$
1	0.01	7	0.07
2	0.06	8	0.36
3	0.05	9	0.05
4	0.06	10	0.00
5	0.01	11	0.02
6	0.59		

greater generalization is considered in determining information centrality. It is possible that a *non-geodesic* path is of greater significance than *all other paths*. Information centrality takes this point under consideration. *All* paths, both geodesic and non-geodesic, are weighted when information centrality is calculated. Information centrality was not implemented in this research.

2.5. Summary

Building computational models for simulating disease spread is challenging, at best. The underlying framework of an outbreak model must emulate complex behaviors without becoming too computationally expensive. A graph theoretical approach allows this social environment to be represented in a simple format, i.e. nodes and edges. The nodes of a graph represent the individuals in a population and the edges in the graph correspond to relationships among individuals. This basic construct creates a foundation for disease simulation.

The SIR infectious disease model is very compatible with a graph-based social network. This research explores two methods for implementing a simulated disease outbreak. The first technique, implemented in Chapter 3, generates the social network simultaneously as the disease proliferates.

Contacts are established between individuals who may be susceptible, infectious, or removed. The disease transfers from an infectious individual to one who is susceptible with a probability based on the transfer rate. The second approach, implemented in Chapter 4, creates a static social network and then simulates a disease spread on the predefined graph.

Dynamically generating a social network may arguably be more realistic because true social contacts are not restricted or predetermined. This is a reasonable choice for investigating how an outbreak manifests itself if no intervention strategies are employed. The exploration in Chapter 4, however, focuses on targeted vaccination. This cannot be accurately tested in dynamically created social networks because the node attributes must be known in advance of vaccination. Nodes that have certain properties in one graph may not exhibit the same properties in another graph. Thus, it is impossible to create the contact graph concurrently with an outbreak. A concession is to create and save the contact graph first, vaccinate specific individuals, and then allow the outbreak to evolve.

CHAPTER 3

GRAPH STRUCTURE AND OUTBREAK SEVERITY

As previously discussed in Section 2.4.2, small-world graphs are considered acceptable models for representing social networks. In addition to the size of a small-world network, there are two parameters that affect the structure of the graph. The first is the neighborhood size, k, and the second is the probability of a random contact, p. The neighborhood size represents a *contact* group. The definition of *contact*, however, changes depending on the situation being modeled by the small-world graph. Even in the same area of research, the size of k varies. For example, a contact required for the transfer a sexually transmitted disease is not equivalent to a contact necessary for the transfer of influenza. The value of k, therefore, depends on context. Similarly, the probability of a random contact, p, is a conditional parameter. The purpose of the experiments presented in this chapter is to explore outbreak variation as a result of changes in these two parameters.

3.1. Simulation Method

Based on the SIR model with an additional latent state, each experiment has a number of static parameters as shown in Table 3.1. The parameters selected for these experiments result in a large portion of the population becoming infected in most cases. This is intentional and is not a fallacy of the simulator. Each experiment begins with the primary case in the *infectious* state and all other individuals in the *susceptible* state. Contacts are made each day until the predetermined number of contacts for the entire population is reached. The simulation continues as long as there are individuals in either the infectious or latent state. A social network is created dynamically as the simulation progresses; an edge is created between two nodes every time a contact occurs. Generated contact graphs range from ordered to random. An ordered graph, $p = 0$, is one in which contacts are *only* allowed within a restricted neighborhood of size k. A random graph, $p = 1$, is one in which contacts are made randomly between any two individuals in the network.

Initial experiments are conducted on a small network graph of size 30. N nodes are labeled $0 - 29$ and node 29 is linked back to node 0. This population size provides modeling capabilities that are limited to situations in which a small group of individuals is predominantly self-contained, such as in a nursing home or on a ship. However, a benefit to such a small population is that it allows visualization of the contact and infectious graphs, which is not viable on larger population sets. For this set of experiments, the neighborhood size increases from $k = 2$ to $k = 10$ in increments of 2 and from $k = 14$ to $k = 30$ in increments of 4. For each value of k, the probability of a random contact increases from $p = 0.0$ to $p = 1.0$ in increments of 0.1. For each unique value of k and p, the simulation is repeated 100 times.

Similar experiments are conducted on a larger population ($N = 500$) to gauge whether the results are scalable. Due to the substantial number of simulations (10 values of k, 10 values of p, 100 simulations each), larger population sizes are beyond the scope of this research. The neighborhood size increases from $k = 2$ to $k = 10$ in increments of 2 and from $k = 20$ to $k = 100$ in increments of 20. As with the previous experiments, for each value of k, the probability of a random contact increases from $p = 0.0$ to $p = 1.0$ in increments of 0.1, and for each value of k and p, there are 100 executed simulations. The complexity of the graphs created using a population of size 500 makes it impractical to include visual representations.

3.2. Results

Figure 3.1 illustrates a representative contact graph and the resulting outbreak graph with $k = 6$ and $p = 0$. Note that not all of the nodes $(0 - 29)$ are included in the contact graph. This signifies that during this particular simulation there are some individuals that never make any contacts. Since $p = 0$, all contacts are made within the neighborhood of size six, three to the left and three to the right. Because the graphs are not weighted, an edge between two nodes signifies one or more contacts. The contact graph depicted in Figure 3.1 (a) illustrates that Node 3 makes at least one contact with Node 1 and the related infectious graph in Figure 3.1 (b) indicates that the disease transfers from Node 3 to Node 1. It is not evident, however, how many contacts are made before or after the transfer. Any contacts that occur after disease transfer are useless in terms of disease

propagation since an individual can only be infected one time. In contrast, Figure 3.1 illustrates a representative contact graph and the resulting outbreak graph with $k = 6$ and $p = 1$. For this particular simulation, the resulting contact graph is a near complete graph. In a random graph, such as this, contacts are not as likely to be repeated. This implies that there is a higher probability that the disease will transfer to more individuals in the population. Another point regarding Figure 3.1, is that the neighborhood size is essentially irrelevant. A value of $p = 1$ indicates that every contact is random. Each individual is allowed to make contact with any other individual in the population, thereby effectively eliminating the boundaries of a neighborhood.

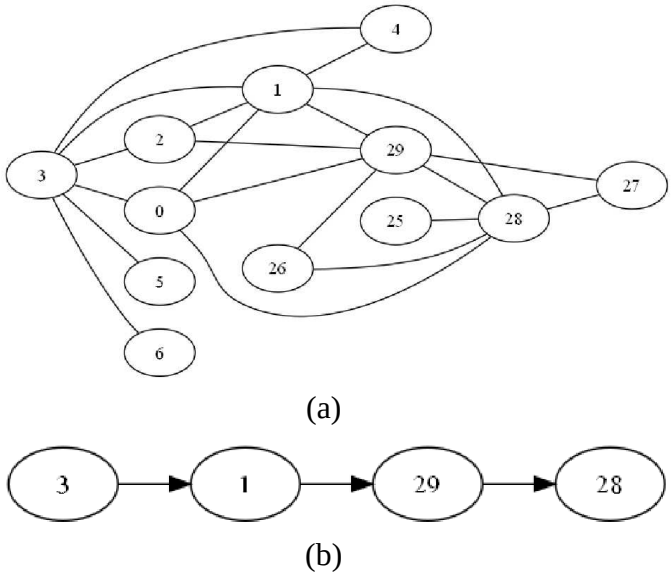
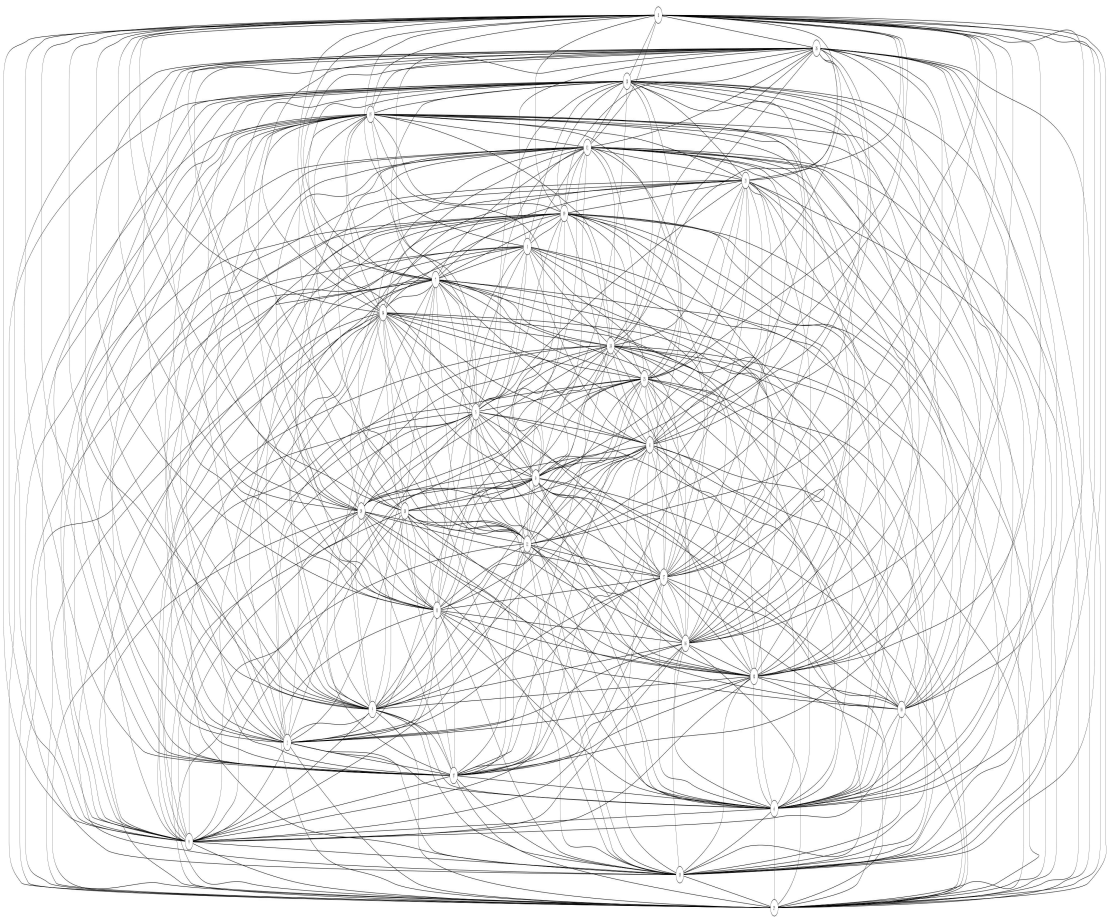


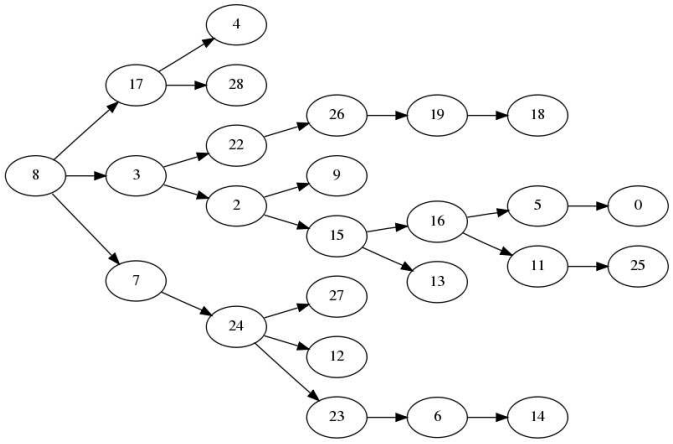
FIGURE 3.1. Example of an ordered graph with k = 6, p = 0. (a) Contact graph; (b) Resulting outbreak graph

3.3. Duration

With a population of size 30, the minimum average duration is 19 days which occurs with the smallest neighborhood size and no random contacts, $k = 2$ and $p = 0$. This coincides with the minimum percent infected of 18%, which is an indication that the duration of an outbreak is relatively short when few individuals become infected. The maximum duration is 34 days which occurs when $k = 10$ and $p = 0$. Although the shortest duration aligns with the fewest infected, the longest duration does not align with the most infected. Initially, the duration increases as



(a)



(b)

FIGURE 3.2. Example of an ordered graph with $k = 6$, $p = 1$. (a) Contact graph; (b) Resulting outbreak graph

more individuals become infected, however, there appears to be a point at which concurrency of secondary infections reduces the duration. Therefore, there is not a direct relationship between outbreak duration and the proportion of the population infected.

Figure 3.3 reveals that limiting the neighborhood size below 6 has a significant impact on the duration when $p \leq 20\%$. Larger values of k have a fairly consistent duration regardless of the value of p. This is likely due to the fact that many of the simulations with $k < 6$ and $p \leq 20\%$ are unable to sustain an outbreak. During these simulations, secondary infections are limited by the small neighborhood size and the disease dies out quickly. However, as k and/or p increases, the chance for secondary infections rises, resulting in an increase in the number of sustained outbreaks.

With a population of size 500, the minimum average duration is 26 days which occurs when $k = 2$ and $p = 0$. As with the smaller population, the shortest duration coincides with the fewest number of infected individuals. The maximum duration is 165 days and occurs when $k = 20$ and $p = 0$. Consistent with experiments on a population of size 30, the longest duration does not align with the most infected. Figure 3.4 (a) demonstrates that for neighborhood sizes of $k = 2$ through $k = 10$ in which all contacts are made within the neighborhood ($p = 0$), the outbreak either does not occur or does not last long when it occurs. As p increases the duration of the outbreak peaks and then falls. A reasonable explanation for the peak is as follows: As outside contacts increase ($p > 0$), there is an increase in the number of simulations that actually produce a significant outbreak. This increase in significant outbreaks brings up the average duration. As the number of random contacts is further increased, there is a rise in concurrent secondary infections. The rise in concurrent secondary infections allows the disease to move more rapidly throughout the population, causing a decrease in the duration.

Figure 3.4 (b) illustrates that neighborhood sizes $k = 20$ through $k = 100$ on average are able to sustain an outbreak. Similar to the fall after the peak in the smaller neighborhoods, as the random contacts increase, the disease spreads more quickly. The average duration of all simulations for neighborhood sizes $k = 20$ through $k = 100$ is 69.2 days with a standard deviation of 18.25.

Considering only values of p from 0.5 to 1, the average duration is 62.6 days with a standard deviation of 2.46. This indicates that the duration becomes much more stable as p increases.

The duration of an outbreak by itself does not appear to be a reliable severity indicator. An extended time period does not necessarily signify that a large proportion of the population will become infected. For example, the longest duration of 165 only infects 38% of the population, whereas the average over all simulated runs is 70%. Likewise, a reduced time frame may indicate that the outbreak subsides before it has a chance to take hold or it may indicate that the disease rapidly spreads throughout the population infecting many individuals. This is demonstrated by two simulations that both have a duration of 64 days. One, which has a neighborhood size of 10 with 0 random contacts, infects only 7% of the population. The other, which has a neighborhood size of 80 with 90% random contacts, infects 82% of the population. Without additional information, such as the proportion infected, the duration of an outbreak does not accurately reflect the severity.

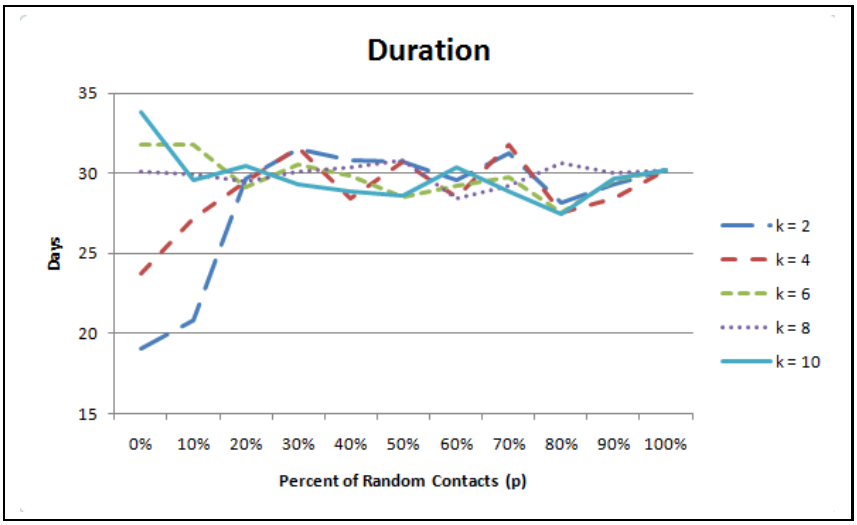
3.4. R_0

The value of R_0 is measured according to the strict definition by averaging the number of the secondary infections caused by the first infectious individual in a primarily susceptible population. The mathematical expected value of R_0 is obtained by multiplying the average daily contacts, the transmission rate, and the infectious period. For the parameters shown in Table 3.1, the expected R_0 value is 2.4.

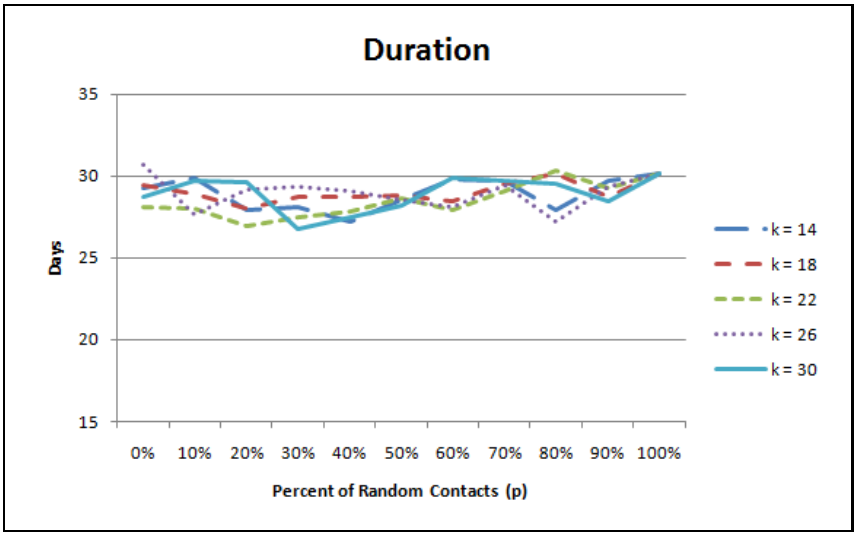
TABLE 3.1. Simulation Parameters for Small Graph Experiments

Parameter	Value	Explanation
Loop	100	Number of times to loop through the simulation
CR	20	Number of contacts per person, per day
TR	0.03	Probability that transfer of infection will occur when a contact is made between an infectious person and a susceptible person
DaysL	3	Number of days in the latent stage
DaysI	4	Number of days in the infectious stage

For experiments performed on a population size of 30, the minimum average value of R_0 is 1.3 which occurs when $k = 2$ and $p = 0$. The maximum value of R_0 is 2.59 and occurs when $k = 26$ and $p = 0.7$. The average value of R_0 is 2.20. Although the average is lower than the expected



(a)

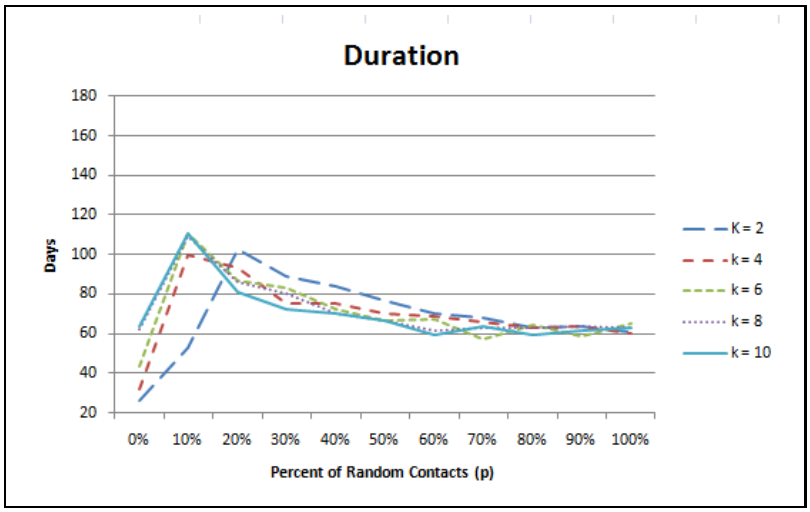


(b)

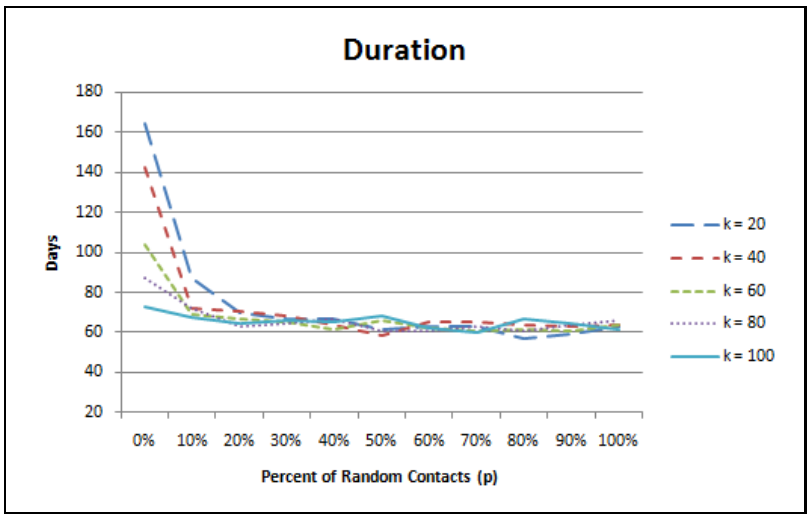
FIGURE 3.3. Outbreak durations on population size 30, $p = 0.0$ to 1.0 for (a) $k = 2$ to 10 and (b) $k = 14$ to 30

value, it is not a surprising number considering the small population size and the low values of k that are included in the average. Figure 3.5 reveals that small neighborhood sizes, accompanied by a low probability of outside contacts, results in a decrease in secondary infections. As the neighborhood size is increased, the value of R_0 becomes more stable.

For experiments performed on a population size of 500, the minimum average value of R_0 is 1.4 which occurs when $k = 2$ and $p = 0$. The maximum value of R_0 is 2.74 and occurs when



(a)



(b)

FIGURE 3.4. Outbreak duration on population size 500, $p = 0.0$ to 1.0 for (a) $k = 2$ to 10 and (b) $k = 20$ to 100

$k = 100$ and $p = 0.8$. The average value for R_0 over all simulations is 2.29. The average value of all simulations for $k = 2$ through $k = 10$, as shown in Figure 3.6 (a), is 2.17 which is below the expected value of 2.4. However, the average of all simulations $k = 20$ through $k = 100$, as shown in Figure 3.6 (b), is exactly the expected value, 2.4. As the neighborhood size or the proportion of outside contacts increase, the value of R_0 is shown to approximate the expected mathematical value. Conversely, a small neighborhood size coupled with limited random contacts reduces the value of R_0 by restricting the number of susceptible individuals that are available for contact.

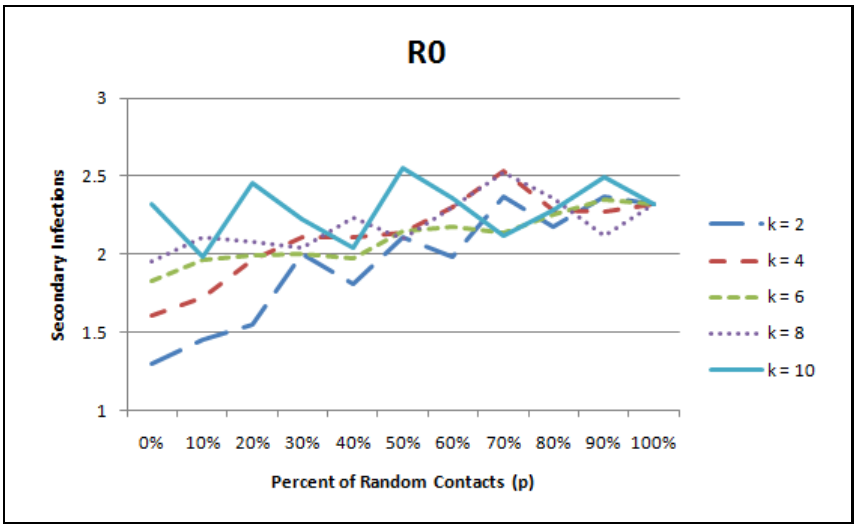
TABLE 3.2. Comparison of R_0 and Percent of Infected Population on Ordered Graph ($p = 0$) with Small Neighborhood Sizes.

Neighborhood Size (k)	R_0	Percent Infected
2	1.4	1
4	1.9	2
6	1.9	4
8	2.0	6
10	2.1	7

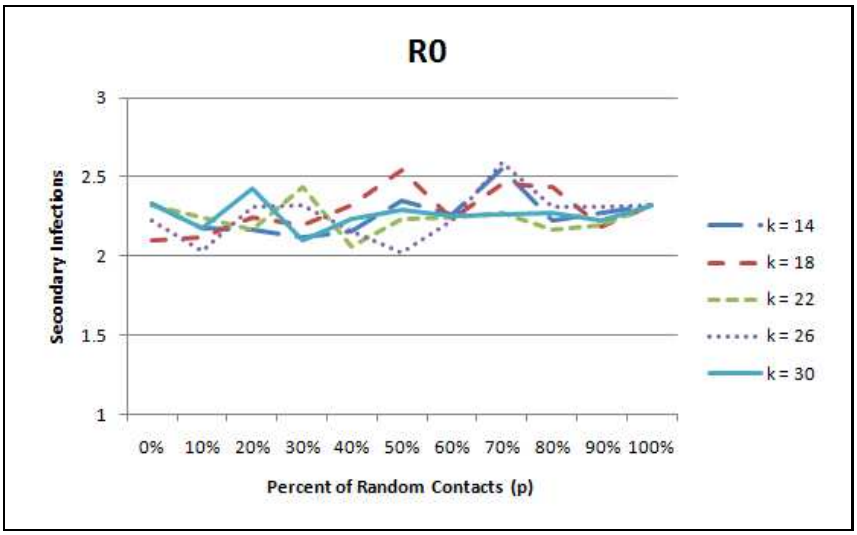
A comparison of the data obtained for R_0 and the proportion of the population infected provides an interesting observation regarding secondary infections. Theoretically, values of $R_0 > 1$ should produce significant outbreaks in a population. However, these experiments do not support this theory when neighborhood sizes are 10 or less and no random contacts are permitted. This is indicated by the very low proportion of the population that is infected as shown in Figure 3.8 (a) when $p = 0$. See Table 3.2 for a comparison of values. Even though secondary infections are above the threshold value of one, the outbreak is not maintained long enough to infect a substantial portion of the population. This is likely caused by a saturation of infected individuals within a neighborhood. The primary case is able to infect more than one individual in their neighborhood leading to an R_0 value greater than unity, but secondary infections by other individuals are limited by competition. For example, suppose Node 3 is the primary case and infects Nodes 2 and 4, resulting in a literal R_0 value of 2. At this point, Node 4 is unable to infect two of its immediate neighbors, Nodes 2 and 3, because they are already infected. Node 2 is limited in the same way. This leads to a *suffocation* of the outbreak. While the actual R_0 is greater than one, the *effective* value of R_0 is quickly reduced. In reality, it is possible to cause a similar affect by reducing the neighborhood size of susceptible individuals through vaccination.

3.4.1. Total Infections

With regard to the population as a whole, the overall proportion of individuals infected during the course of an outbreak is the strongest indicator of severity. Intervention strategies, such as public awareness, vaccination, quarantine measures, etc., are designed to reduce secondary infections with the ultimate goal of lowering the total number of individuals who become infected. The three



(a)

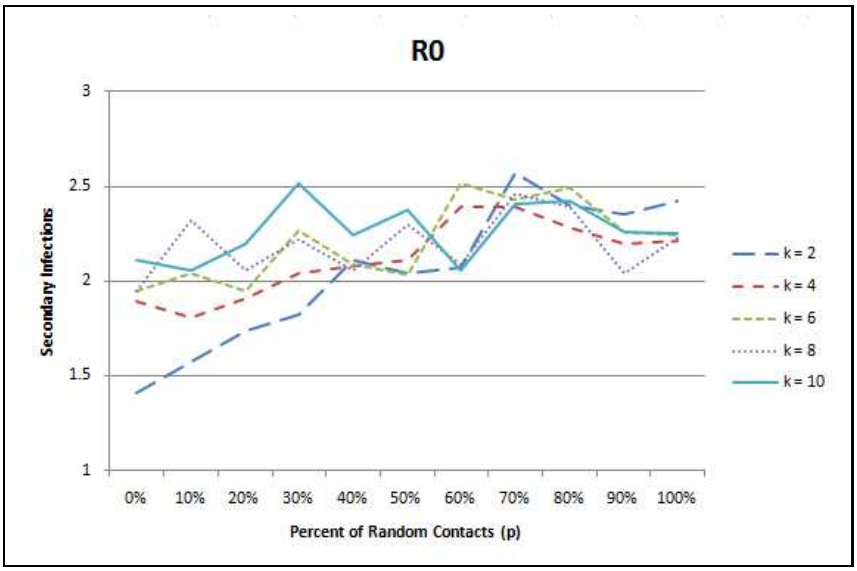


(b)

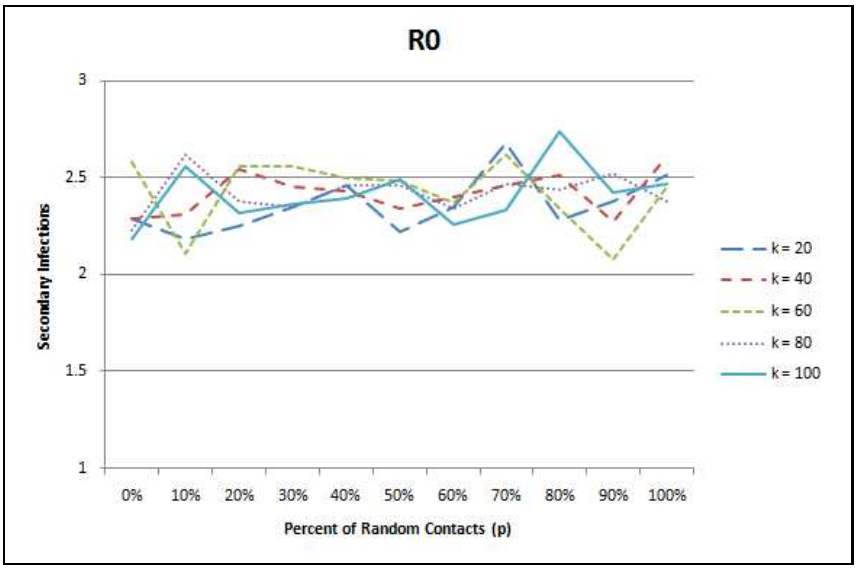
FIGURE 3.5. R_0 values on population size 30, $p = 0.0$ to 1.0 for (a) $k = 2$ to 10 and (b) $k = 14$ to 30

disease-spread indicators examined in this dissertation (duration, R_0, and proportion infected) are unquestionably interrelated. However, it is the proportion of the population infected that provides the most obvious measure of severity.

With a population of size 30, the minimum average value of total infections is 5.43, or 18% of the population which occurs at $k = 2$ and $p = 0$. The maximum average is 24.3, or 81% of the population which occurs when $k = 18$ and $p = 70\%$. Small neighborhoods coupled with few or



(a)



(b)

FIGURE 3.6. R_0 values on population size 500, $p = 0.0$ to 1.0 for (a) $k = 2$ to 10 and (b) $k = 20$ to 100

no random contacts significantly reduce the proportion of the population that becomes infected as demonstrated by Figure 3.7 (a). However, the proportion infected rises rapidly with an increase in random contacts. Additionally, if the neighborhood size is large enough, i.e. $k \geq 14$, the proportion infected is not drastically reduced by a decrease in random contacts (see Figure 3.7 (b)). The average percent of the population infected for all $k \geq 14$ is 74% with a standard deviation of

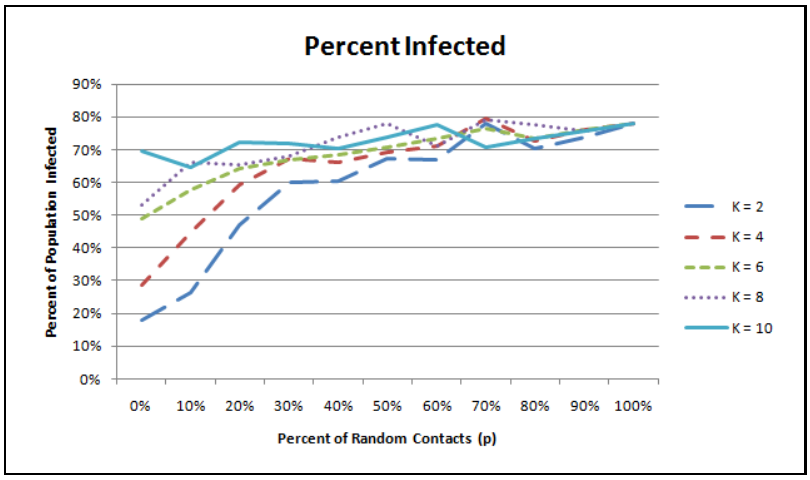
3.1%. For $k \geq 14$ and $p \geq 60\%$, the percent infected rises to 76% with a standard deviation of 2.4%.

With a population of size 500, a mean of 70% of the population becomes infected over all averaged simulations. The lowest average of total infections is 6.58, or 1% of the population which occurs when $k = 2$ and $p = 0$. The highest average is 84% at $k = 100$ and $p = 80\%$. Similar to the results with a population of size 30, Figure 3.8 (a) reveals that simulations with neighborhood sizes of $k = 2$ to $k = 10$ and $p = 0$ result in a very small proportion of the population becoming infected, ranging from 1% to 7%. There is a sharp incline as k and p increase. In fact, for all values of $k \geq 4$ and $p \geq 30\%$ the average remains above 70%, with a mean of 77% and a standard deviation of 3.1%. Figure 3.8 (b) illustrates that, with the exception of $k = 20$ and $p = 0$ at 38%, the proportion of the population infected remains relatively consistent. For these values, exception noted, the mean is 77% with a standard deviation of 2.9%.

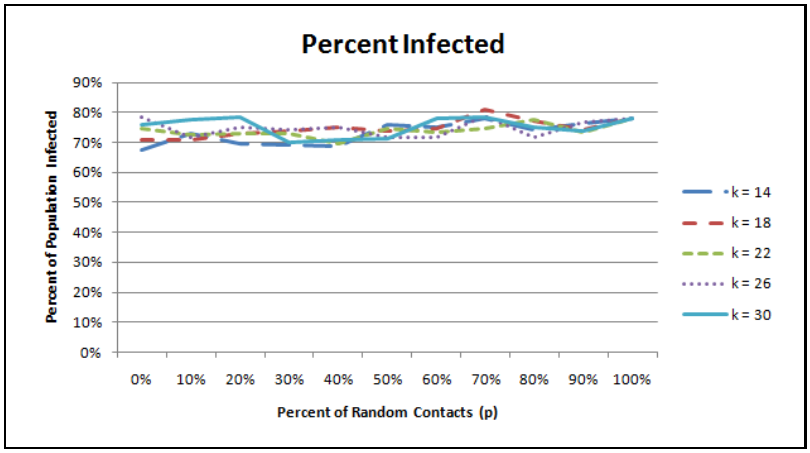
Similar results are observed in both the small and large graph simulations. When the neighborhood size is restricted and the probability of random contacts is low, the proportion of the population that is infected is greatly reduced. Moderate increases in either one or both of these parameters greatly increases this proportion. In the small graph simulations, approximately 65% or more of the population is infected for all values of k when $p \geq 50\%$; for all values of p when $k \geq 10$; and when $k \geq 6$ and $p \geq 30\%$. In the large graph simulations, approximately 65% or more of the population is infected for all values of k when $p \geq 40\%$; for all values of p when $k \geq 40$; and when $k \geq 6$ and $p \geq 20\%$. This implies that small-world graphs are very conducive to the spread of disease, even with relatively small values for k and p. It should be noted, however, that the parameters selected for these experiments generate a high probability of producing an epidemic and no preventative measures are taken during any simulations.

3.5. Summary

Two groups of experiments were presented and discussed in this Chapter. The first involves a series of simulated outbreaks on a population of size 30. This small population size was purposely chosen to allow visual inspection of the resulting contact and outbreak graphs. The second involves



(a)

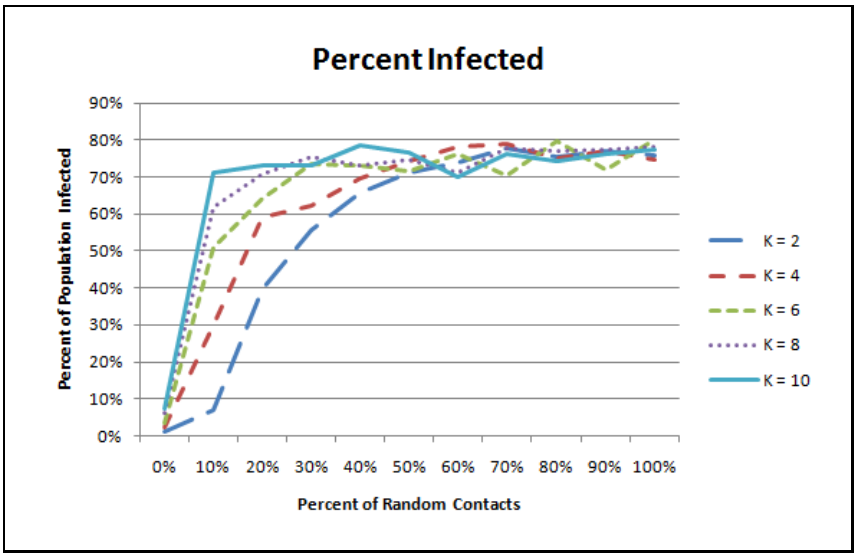


(b)

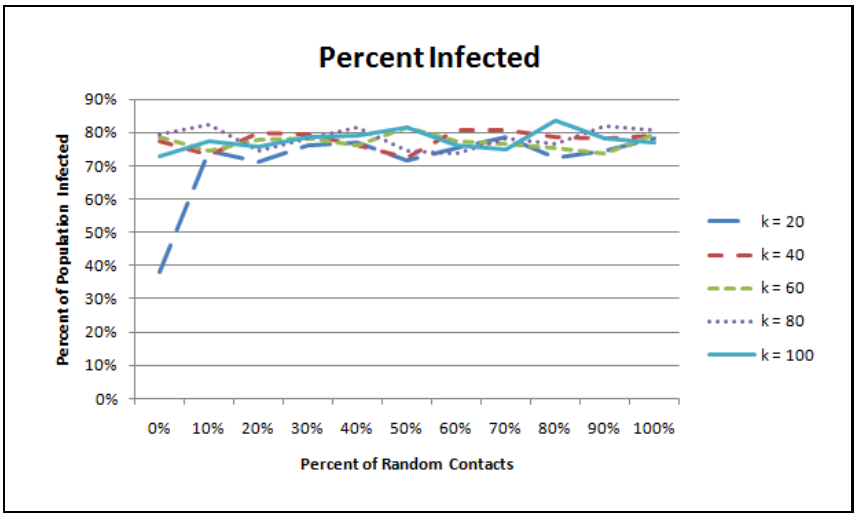
FIGURE 3.7. Total infected on population size 30, $p = 0.0$ to 1.0 for (a) $k = 2$ to 10 and (b) $k = 14$ to 30

identically constructed experiments on graphs of size 500. Even though a size of 500 is smaller than the population of most communities, the parameter variation and number of simulations are prohibitive on larger populations. It is also important to consider that a node in a graph can be representative, not only of an individual, but also of a group of individuals. Therefore, a *population size* of 500 might represent 500 families or 500 cities. The results of the first set of experiments scale nicely to those of the second and it is likely that similar results would be observed on larger data sets.

It is evident that the neighborhood size, k, can have a considerable impact on the severity of an outbreak when both k and p are relatively small. However, there appears to be a threshold value of



(a)



(b)

FIGURE 3.8. Total infected on population size 500, $p = 0.0$ to 1.0 for (a) $k = 2$ to 10 and (b) $k = 20$ to 100

k beyond which there is negligible difference regardless of the value of p. A probable explanation for this threshold value is that there is a level of competition for susceptible individuals when contacts among infected individuals overlap. Once the neighborhood size becomes large enough to eliminate this competition, there is no longer a gain by increasing the size even larger.

In the interest of public health, it is beneficial to employ intervention strategies that effectively reduce the neighborhood size to a level below the threshold. To accomplish this, there are several

preventative measures that can be initiated. For example, the number of daily contacts conducive to disease transfer can be reduced through hand-washing, the use of anti-bacterial products, and social distancing. With proper medical treatment of those who are infectious, it may be possible to reduce the infectious period, and thereby lower the number of secondary infections. Furthermore, the number of susceptible individuals can be reduced through vaccination. Vaccination is a common preventative strategy and the effectiveness of particular vaccination strategies is the topic of discussion in the following chapter.

CHAPTER 4

VACCINATION STRATEGIES BASED ON CENTRALITY MEASURES

Although epidemics are inevitable, it is possible to reduce their impact on society. Ideally, enough individuals could be vaccinated to stop an outbreak from ever reaching epidemic status, a concept referred to as herd immunity. In most cases, however, this is not a practical solution. Herd immunity is achieved if the effective basic reproductive number is brought to a level below unity. Unfortunately, large intrinsic values of R_0, require very high levels of vaccination. In a paper published in 1982 by Anderson and May, it is reported that the proportion, p, of the population that must be vaccinated to achieve herd immunity is given by Equation 19 [8]. Therefore, a disease with an intrinsic $R_0 = 3$ would require that more than $\frac{2}{3}$ of the population be vaccinated. Data from the Centers for Disease Control (See Appendix 5.2.1) indicates that even the yearly influenza vaccine, in anticipation of expected outbreaks, is distributed in much lower quantities. It is highly unlikely that an adequate vaccine supply would be available in the event of an unforeseen disease outbreak.

(19) $$p > 1 - \frac{1}{R_0}$$

The experiments in this chapter explore vaccination methods based on centrality. The results found previously imply that small world graphs effectively facilitate disease spread in a simulated environment even when the neighborhood size and probability of contacts outside the neighborhood are relatively small. Discussing similar results, a research article by Watts and Strogatz states, “Infectious diseases are predicted to spread much more easily and quickly in a small world; the alarming and less obvious point is how few short cuts are needed to make the world small” [80]. In the previously presented experiments, no intervention strategies are implemented and a large

portion of the population became infected. This chapter analyzes the effectiveness of various vaccination strategies based on modifications of the centrality measures discussed in Section 2.4.3. Experimentation follows the steps below which are repeated over various graph structures and includes several vaccination policies for each distinct graph structure.

(i) Create a graph-based social network utilizing parameters given in Table 4.1.

(ii) Vaccinate individuals in the population.

(iii) Simulate multiple outbreaks in the established social network and collect data to assess the severity of the outbreaks.

In contrast to the experiments presented in Chapter 3, the contact graphs are generated prior to each outbreak to allow targeted vaccination of specific nodes based on centrality. The same contact graphs are utilized for each vaccination policy and outbreak simulation. Statistics are recorded for each simulation, including values of R_0, duration, and the proportion of the population infected. Comparisons of each indicator are presented in Section 4.4.

4.1. Creating a Social Network Graph

A population of size N is represented as a graph $G(V, E)$ in which each vertex in the graph, $v \in V$, represents an individual and each edge in the graph, $e(v, w) \in E$ represents a contact between two individuals. Each individual is labeled with a unique identification number between 0 and $n - 1$, inclusive, and Node $n - 1$ is adjacent to Node 0. Each member of the population has an assigned neighborhood of size k, such that the neighborhood extends $k/2$ to the left and $k/2$ to the right of that individual.

The contact graph is established based on the parameters listed in Table 4.1. Specific values for these parameters are discussed in Section 4.3. The total number of contacts for the entire population is calculated as the size of the population, N, times the average number of contacts per person, per day, CR. The procedure of building the contact graph continues until the total number of contacts has been exhausted. The algorithm for creating the contact graph is outlined in Algorithm 1.

Algorithm 1 CONTACT GRAPH

```
ContactCount ← 0
while ContactCount < TotalContacts do
  P1 ← random number from 0 to N − 1
  rn ← random number between 0 and 1
  if rn < p then {Contact should be global}
    P2 ← index from 0 to N − 1, ∉ neighborhood of P1
  else
    P2 ← index ∈ neighborhood of P1
  end if
  if ContactGraph HasEdge (P1, P2) then
    ContactGraph EdgeWeight(P1, P2) ← ContactGraph EdgeWeight(P1, P2) + 1
  else
    ContactGraph AddWeightedEdge(P1, P2, 1)
  end if
  ContactCount ← ContactCount + 2
end while
```

TABLE 4.1. Contact Graph Parameters

Parameter	Explanation
N	Population Size
CR	Average number of contacts per person, per day
k	Neighborhood size
p	Probability of a random contact (contact outside of neighborhood)

4.2. Vaccinating Key Individuals

Vaccination policies are often designed with the primary purpose of protecting individuals. For this reason, vaccines are often recommended for the very young and the very old. Although it appears to be rational thought to safeguard the most vulnerable, this may not be the best strategy for protecting a population. In the event of a limited supply of vaccination, the entire population would likely benefit from a policy that completely restricts or greatly reduces the disease spread. This chapter explores targeted vaccination of ten percent of population sizes 50, 150, and 250 by identifying central nodes in a social contact graph.

The centrality measures of *degree*, *betweenness*, and *closeness* were previously defined in Chapter 2 for unweighted graphs. Recent research suggests that it is beneficial to represent social networks as weighted graphs [12, 62, 63, 74]. This is especially relevant in the domain of

disease spread where repeated contacts increase the likelihood that a disease will transfer from one individual to another. Outlined below are centrality measures designed specifically for the purpose of identifying individuals in a social network who are more prone to facilitate disease spread.

4.2.1. Contact Centrality

Contact Centrality measures the number of contacts an individual makes within a unit of time, including those contacts which are unique and those which are repeated. Represented in a graph, an edge with a weight of 1 between two nodes is initially created upon the first contact between the two nodes. Each additional contact between the same two nodes increases the edge weight by one. The contact centrality for node *i*, $C_N(i)$, is calculated as a sum of the edge weights between *i* and all neighbors of *i*. This is easily calculated through the use of an adjacency matrix, A_{ij}. Each entry in A_{ij} represents the weight of the edge between *i* and *j*. This calculation (see Equation 20) is identical to that of degree centrality presented previously with the exception that the adjacency matrix is weighted rather than binary.

DEFINITION 4.1. Contact centrality is defined as the average number of contacts an individual makes within a specified unit of time.

$$C_N(i) = \sum_{j=1, i \neq j}^{n} A_{ij} \tag{20}$$

$$C_{\acute{N}}(i) = \frac{\sum_{j=1, i \neq j}^{n} A_{ij}}{\sum_{i=1}^{n} \sum_{j=i+1}^{n} A_{ij}} = \frac{C_N(i)}{\sum_{i=1}^{n} \sum_{j=i+1}^{n} A_{ij}} \tag{21}$$

Contact centrality is illustrated in Figure 4.1. In this graph, Node 4 has a contact centrality of 16 which is the highest value in this network. The edge weight of 7 between nodes 1 and 4 implies that 7 contacts are made between these two individuals that are capable of disease transfer if one

individual is infectious and the other is susceptible. Likewise, there are 9 possible opportunities for transfer between nodes 4 and 8. A measure of *non-weighted* degree centrality in this same network would identify Node 3 as the most central, even though Node 3 makes fewer overall contacts than Node 4.

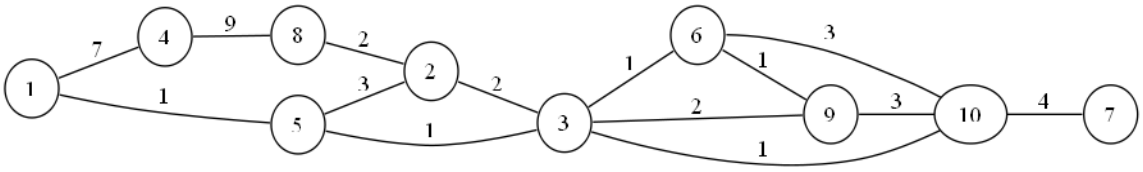


FIGURE 4.1. Contact centrality illustrated by Node 4

Degree centrality is normalized by dividing each centrality measure by the number of possible edges, which is $n - 1$ in a graph of size n. Because the weight of edges in a weighted graph is potentially unlimited, contact centrality is normalized by dividing by the total of all edge weights. Like standard normalization techniques, this will produce a centrality value between 0 and 1, inclusive. A normalized contact centrality of 0 indicates that the node is disconnected, as illustrated by Node 3 in Figure 4.2. A normalized contact centrality of 1 indicates that the graph is structured as a star or wheel, as illustrated by Node 4 in Figure 4.3. The formula for normalized contact centrality of Node i, $C_{\acute{N}}(i)$, is given in Equation 21. This is simply the contact centrality of Node i divided by the sum of half of the undirected weighted adjacency matrix.

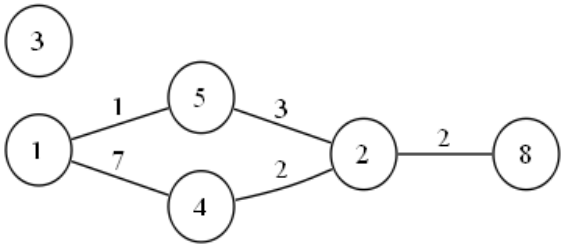


FIGURE 4.2. Disconnected node with a contact centrality of 0 illustrated by Node 3

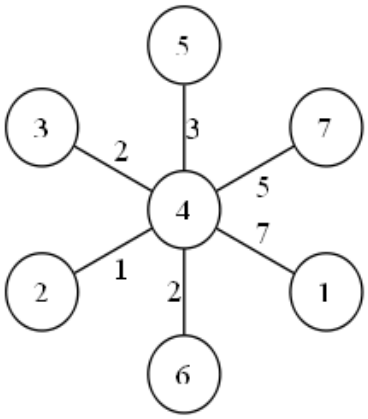


FIGURE 4.3. Node 4 has a normalized contact centrality value of 1

4.2.2. Transmission Centrality

Transmission Centrality measures the degree to which an individual lies on a potential transmission path between other individuals. Transmission centrality is a variation of betweenness centrality such that edge weights are taken into consideration. Ulrik Brandes provides an algorithm for determining betweenness centrality in weighted graphs in the paper "On Variants of Shortest-Path Betweenness Centrality and their Generic Computation" [14]. Transmission centrality, as defined in this paper, applies this algorithm, as well as the suggested use of weight inverses to indicate that stronger weights represent closer ties.

DEFINITION 4.2. Transmission centrality is defined as the likelihood that an individual lies on a randomly selected shortest weighted path between any two other individuals in a network.

To determine overall transmission centrality, it is advantageous to calculate partial transmissibility, $p_{st}(i)$, for each node, as shown in Equation 22. In this formula, g_{st} represents the number of geodesic paths from s to t and $g_{st}(i)$ the number of geodesic paths from s to t that contain i. Partial transmissibility represents the probability that a node i lies on the geodesic path from node s to node t for distinct nodes $s \neq i \neq t$.

The transmission centrality of a node is the sum of all the partial transmissibilities. Equation 23 displays the calculation for transmission centrality of node i. The double summation is required in the formula because for each node s, all other points t must be considered in determining whether

i lies on one or more geodesic paths from s to t. If node i lies on every geodesic path from s to t, the transmission centrality for node i is increased by one. If node i only lies on a portion of the geodesic paths from s to t, the transmission centrality for node i is increased by that proportion.

The largest possible transmission centrality occurs when node i lies on every geodesic path between every two nodes s and t, $s \neq i \neq t$. In terms of public health, if an individual is in a position to have maximum transmission centrality, vaccinating this individual effectively protects an entire segment of the population. For example, a person or group of people who bring supplies to a remote village may have maximum transmission centrality to and from the village. Vaccination prevents transfer of disease from the greater population to the village and likewise, transfer from the village to the greater population.

Since there are $n-1$ nodes not equal to i and $n-2$ nodes not equal to i or s, the maximum possible transmission centrality is one-half $n-1$ times $n-2$ as shown in Equation 24. Division by 2 is necessary in an undirected graph since the path from s to t is equivalent to the path from t to s. This maximum value is used to normalize transmission centralities. The normalized value calculation for node i, $C_T'(i)$, is shown in Equation 25.

The formulas for transmissibility are identical to those of betweenness centrality. However, there is a difference in the definition of the geodesic path between two points. In a non-weighted graph, the path length between point s and point t is measured by the number of edges between the two points. A geodesic path from s to t, therefore, is one that has the fewest edges between s and t. In a weighted graph in which the weight indicates the number of contacts between two individuals, it is reasonable to consider a path to be shorter along more heavily weighted edges. Thus, the sum of the inverse of all edge weights along each path from s to t is calculated to determine the geodesic path.

This calculation of the geodesic path makes a significant difference when considering disease transmission. Consider a situation illustrated by Figure 4.4 in which the individual represented by Node 0 is infectious and all other nodes are susceptible. Considering path length only, Node 3 is most likely to become infected via Node 4. However, if the frequency of contacts is taken

into account, Node 3 is most likely to become infected via Nodes 1 and 2. Node 4 has a higher betweenness centrality, whereas Nodes 1 and 2 have higher Transmission centralities.

(22) $$p_{st}(i) = \left(\frac{1}{g_{st}}\right)(g_{st}(i)) = \frac{g_{st}(i)}{g_{st}}$$

(23) $$C_T(i) = \sum_{s=1}^{n}\sum_{t=s+1}^{n} p_{st}(i)$$

(24) $$maxC_T(i) = \frac{(n-1)(n-2)}{2} = \frac{n^2-3n+2}{2}$$

(25) $$C_T^{'}(i) = \frac{2C_T(i)}{n^2-3n+2}$$

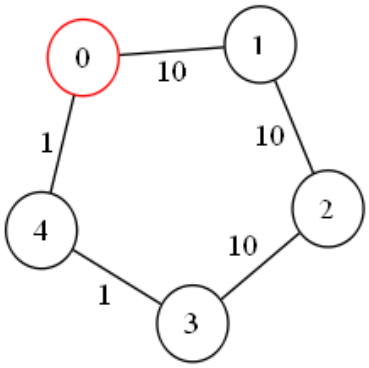


FIGURE 4.4. Weighted geodesic path example: The geodesic path from 0 to 3 in this weighted contact graph is $0-1-2-3$

4.2.3. Spread Centrality

Spread Centrality measures the social distance between an individual and every other individual in the population. Spread centrality is based upon closeness centrality with an additional condition that a larger edge weight value indicates a closer social distance. The formula used to

calculate spread centrality, $C_S(i)$, is the sum of the geodesic distances from i to j, $d(i,j)$, for all points $j \neq i$ as given in Equation 26. The formula for spread centrality is the same as that for closeness centrality except that geodesic distances are calculated as discussed above for transmission centrality.

DEFINITION 4.3. Spread centrality is defined as the weighted social distance between an individual and every other individual in the social network.

Normalization for spread centrality must consider edge weights in addition to size of the graph. Normalization for closeness centrality is achieved by multiplying the raw value by $n-1$. This standardizes across any network size so that the maximum closeness of any node is obtained when that node is connected to every other node and the normalized value is unity. In a weighted graph, a maximum normalized value of unity for spread centrality is obtained when the unweighted normalization is divided by the maximum weight evenly distributed over all other nodes as shown in Equation 27.

$$C_S(i) = \sum_{j=1}^{n} \frac{1}{d(i,j)} \tag{26}$$

$$C_{\acute{S}}(i) = \frac{(n-1)\frac{1}{C_S(i)}}{\frac{\sum_{i=1}^{n}\sum_{j=i+1}^{n} A_{ij}}{(n-1)}} = \frac{(n-1)^2\frac{1}{C_S(i)}}{\sum_{i=1}^{n}\sum_{j=i+1}^{n} A_{ij}} \tag{27}$$

4.3. Simulating an Outbreak on an Established Contact Graph

The third and final stage of the experiments presented herein is the simulation of outbreaks across an established contact network once central nodes have been vaccinated. Outbreaks are based on the susceptible-infectious-removed (SIR) model discussed in Chapter 2. Disease parameters are defined in Table 4.2. The total number of contacts for the entire population is calculated as the size of the population, N, times the average number of contacts per person, per day, CR.

There are two primary components to the outbreak simulation: Make Contacts, Algorithm 2, and Update Population, Algorithm 3. The simulation loops through these two modules until there are no individuals in either the Latent or Infectious state.

Algorithm 2 MAKE CONTACTS

```
TotalContactCount ← 0
while TotalContactCount < TotalContacts do
  P1 ← random number from 0 to N − 1
  P2 ← index of neighbor of Person[P1] by weighted selection
  Person[P1].ContactCount ← Person[P1].ContactCount + 1
  Person[P2].ContactCount ← Person[P2].ContactCount + 1
  TotalContactCount ← TotalContactCount + 2
  rn ← random number between 0 and 1
  if Person[P1].State == Infectious and Person[P2].State == Susceptible then
    if rn ≤ TR then
      Person[P2].State ← Latent
      Person[P2].StateCount ← 0
      Person[P1].SecondaryInfections ← Person[P1].SecondaryInfections + 1
      InfectiousGraph AddDirectedEdge(P1, P2)
    end if
  else if Person[P2].State == Infectious and Person[P1].State == Susceptible then
    if rn ≤ TR then
      Person[P1].State ← Latent
      Person[P1].StateCount ← 0
      Person[P2].SecondaryInfections ← Person[P2].SecondaryInfections + 1
      InfectiousGraph AddDirectedEdge(P2, P1)
    end if
  end if
end while
```

TABLE 4.2. Disease Parameters

Parameter	Explanation
N	Population Size
CR	Average number of contacts per person, per day
TR	Transmission rate of the disease
DaysL	Number of days in the latent state
DaysI	Number of days in the infectious state

The experiments for this dissertation are conducted using programs written in Perl. Perl is selected as the language of choice primarily because it offers a powerful graph library and has

Algorithm 3 UPDATE POPULATION

```
LatentCount ← 0
InfectiousCount ← 0
RemovedCount ← 0
for i = 0 to N − 1 do
  if Person[i].State == Latent then
    if Person[i].StateCount = DaysL then
      Person[i].State ← Infectious
      InfectiousCount ← InfectiousCount + 1
      Person[i].StateCount ← 0
    else
      LatentCount ← LatentCount + 1
    end if
  else if Person[i].State == Infectious then
    if Person[i].StateCount = DaysI then
      Person[i].State ← Removed
      RemovedCount ← RemovedCount + 1
      Person[i].StateCount ← 0
    else
      InfectiousCount ← InfectiousCount + 1
    end if
  else if Person[i].State == Removed then
    RemovedCount ← RemovedCount + 1
  end if
  Person[i].StateCount ← Person[i].StateCount + 1
end for
```

extensive community support. As open-source software, the code is modifiable which is advantageous in meeting the needs of this study. For example, the Graph module in Perl includes not only the standard functions such as add_vertex, add_edge, etc., but also more complex functions, such as APSP (All-Pairs Shortest Path) and Betweenness [40]. The Betweenness function offered by the Graph package returns a betweenness value as described by Freeman [32]. This open-source code is modified to create a module for calculating contact, transmission, and spread centralities as previously described. These three measures are efficiently combined into a single module because of the overlap in the required calculations.

Before simulating any outbreaks, statistical analyses are performed on social network graphs to determine if graphs created with the same parameters produce statistically equivalent network structures with regard to the distribution of node centralities. The Contact Rate (CR) is assigned

a value of 20 and the Neighborhood Size (k) is assigned a value of 6. Three population sizes are tested, $N = 50$, $N = 150$, and $N = 250$. Four values of p are tested for each population size, $p = 0$, $p = 0.01$, $p = 0.25$, and $p = 0.5$. The value $p = 0$ represents an ordered graph. Larger values of p represent small-world graphs that approach random graphs as p is increased. These four values are tested to provide a range of graph structures. For each value of p, thirty distinct graphs are created and for every graph, and the contact, spread, and transmission centrality measures are recorded for every node. From this information, two groups of data for each centrality measure and each distinct p value are compared. Data Set 1 is comprised of the first ten graphs and Data Set 2 is comprised of the last twenty graphs. This division of the data allows n trials to be compared with $2n$ trials to ascertain if more than n simulations are necessary to generate representative graph structures. The mean and standard deviation of each centrality measure are calculated for every graph as well as the average of the means and standard deviations for each data set. The results of the graph structure analyses are summarized in Tables 4.3, 4.4, and 4.5 and discussed in Section 4.4.

Disease outbreaks are simulated over ten distinct graph structures for each set of graph parameters. Additionally, ten outbreaks for each vaccination strategy are simulated within each graph structure. Experiments are performed on population sizes of 50, 150, and 250. The value of p is set to 0, 0.01, 0.25, and 0.5 for each set of experiments. A structural layout of the experiments for this chapter is depicted in Figure 4.5.

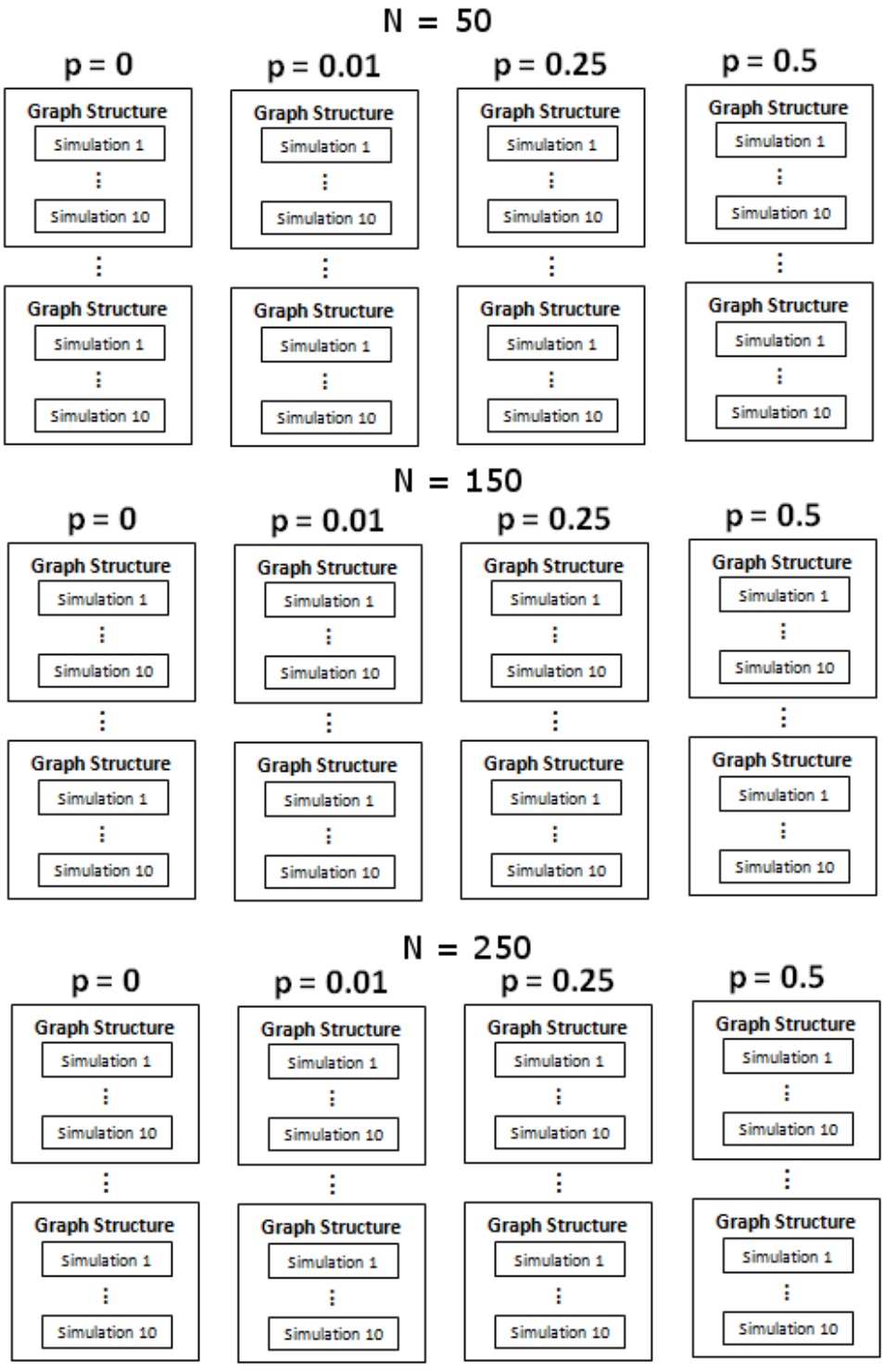


FIGURE 4.5. Vaccination simulation structure

Figure 4.5 describes the organization of the experiments, however, it should be noted that each simulation in the diagram represents distinct outbreaks after each of the following vaccination strategies: low contact centrality, high contact centrality, transmission centrality, spread centrality, and randomness. For comparison purposes, outbreaks are simulated without vaccination as well. This results in a total of $7,200$ simulated outbreaks (6 strategies, 10 outbreaks on each graph, 10 graphs, 4 values for p, and 3 values for n).

4.4. Results

From the experiments discussed in this chapter, results are presented based on observations regarding graph structure and centrality distribution, graph structure and outbreak analyses, and graph structure and vaccination methods. Observations in each of these areas provide interesting findings which will hopefully generate continued research in this area.

4.4.1. Graph Structure and Centrality Distribution

Prior to outbreak simulations, graph analyses are performed to provide a guideline regarding the number of simulations necessary for statistically significant results. In addition to increasing the level of confidence regarding further simulations, the results from this preliminary study offer insight regarding the relationship between graph structure and centrality distribution. The results are summarized in Tables 4.3, 4.4, and 4.5, in which Data Set 1 is comprised of averages from a set of 10 graphs and Data Set 2 is comprised of averages from 20 graphs, such that all 30 graphs are created using the same set of parameters. The consistency between Data Sets 1 and 2 for each value of N and p over all distributions implies that experimentation over ten distinct graph structures should produce reliable results.

The average means and standard deviations for contact centrality over population sizes 50, 150, and 250 are presented in Table 4.3. The average contact centrality has minimal variation across all values of p for each specific N value as expected. The graphs are created with a specific number of edges, (N * CR)/2, therefore, the average contact centrality should be equivalent for a given population size. Due to normalization, which is a division by the total of all edge weights, the average contact centrality decreases as the population size increases.

The average means and standard deviations for transmission centrality over population sizes 50, 150, and 250 are given in Table 4.4. Unlike the average contact centrality, the average transmission centrality does not remain consistent within a given population size. As the probability of non-local contacts increases, the average transmission centrality decreases. Comparing the transmission centrality among the three populations sizes, it is noted that when $p = 0$ (all contacts are local), the transmission centrality is very similar regardless of the population size. However, for $p > 0$, these values do not remain consistent over the various population sizes.

The average means and standard deviations for spread centrality over population sizes 50, 150, and 250 are given in Table 4.5. The average mean spread centrality tends to increase as the probability of non-local contacts increases from the lower values, $p = 0$ and $p = 0.01$ to the larger values, $p = 0.25$ and $p = 0.5$, although in every case there is a slight drop in spread centrality from $p = 0.25$ to $p = 0.5$. This is an interesting point, suggesting that spread centrality may reach a peak in a small-world graph in which p has a value somewhere between $p = 1\%$ and $p = 50\%$. Although outside the realm of this study, this result indicates that further investigation is warranted in this area.

TABLE 4.3. Contact Centrality Distribution Statistics

Contact Centrality		Average Mean	Standard Deviation
N = 50			
p = 0	Data Set 1	0.04000	0.00880
	Data Set 2	0.04000	0.00890
p = 0.01	Data Set 1	0.04000	0.00941
	Data Set 2	0.04000	0.00912
p = 0.25	Data Set 1	0.04000	0.00934
	Data Set 2	0.04000	0.00881
p = 0.5	Data Set 1	0.04000	0.00884
	Data Set 2	0.04000	0.00878
N = 150			
p = 0	Data Set 1	0.01333	0.00303
	Data Set 2	0.01333	0.00300
p = 0.01	Data Set 1	0.01333	0.00301
	Data Set 2	0.01333	0.00302
p = 0.25	Data Set 1	0.01333	0.00299
	Data Set 2	0.01333	0.00300
p = 0.5	Data Set 1	0.01333	0.00291
	Data Set 2	0.01333	0.00299
N = 250			
p = 0	Data Set 1	0.00799	0.00181
	Data Set 2	0.00800	0.00183
p = 0.01	Data Set 1	0.00800	0.00178
	Data Set 2	0.00800	0.00178
p = 0.25	Data Set 1	0.00801	0.00211
	Data Set 2	0.00800	0.00210
p = 0.5	Data Set 1	0.00799	0.00247
	Data Set 2	0.00800	0.00250

TABLE 4.4. Transmission Centrality Distribution Statistics

Transmission Centrality		Average Mean	Standard Deviation
N = 50			
p = 0	Data Set 1	0.09227	0.08425
	Data Set 2	0.09151	0.08379
p = 0.01	Data Set 1	0.07304	0.07063
	Data Set 2	0.07490	0.06897
p = 0.25	Data Set 1	0.03146	0.02338
	Data Set 2	0.03222	0.02439
p = 0.5	Data Set 1	0.02400	0.02113
	Data Set 2	0.02399	0.01936
N = 150			
p = 0	Data Set 1	0.09568	0.10560
	Data Set 2	0.09516	0.10415
p = 0.01	Data Set 1	0.03838	0.03887
	Data Set 2	0.03953	0.03950
p = 0.25	Data Set 1	0.01418	0.00895
	Data Set 2	0.01421	0.00897
p = 0.5	Data Set 1	0.01086	0.00720
	Data Set 2	0.01085	0.00734
N = 250			
p = 0	Data Set 1	0.09625	0.11009
	Data Set 2	0.09560	0.10919
p = 0.01	Data Set 1	0.02791	0.02749
	Data Set 2	0.02805	0.02743
p = 0.25	Data Set 1	0.00968	0.00656
	Data Set 2	0.00980	0.00655
p = 0.5	Data Set 1	0.00736	0.00543
	Data Set 2	0.00739	0.00568

TABLE 4.5. Spread Centrality Distribution Statistics

Spread Centrality		Average Mean	Standard Deviation
N = 50			
p = 0	Data Set 1	0.07994	0.00471
	Data Set 2	0.07970	0.00473
p = 0.01	Data Set 1	0.08362	0.00635
	Data Set 2	0.08195	0.00584
p = 0.25	Data Set 1	0.09090	0.00636
	Data Set 2	0.09121	0.00634
p = 0.5	Data Set 1	0.08811	0.00739
	Data Set 2	0.08817	0.00693
N = 150			
p = 0	Data Set 1	0.02879	0.00082
	Data Set 2	0.02870	0.00086
p = 0.01	Data Set 1	0.04746	0.00421
	Data Set 2	0.04674	0.00418
p = 0.25	Data Set 1	0.06735	0.00353
	Data Set 2	0.06740	0.00359
p = 0.5	Data Set 1	0.06524	0.00367
	Data Set 2	0.06528	0.00387
N = 250			
p = 0	Data Set 1	0.01763	0.00050
	Data Set 2	0.01741	0.00047
p = 0.01	Data Set 1	0.03918	0.00321
	Data Set 2	0.03893	0.00341
p = 0.25	Data Set 1	0.06065	0.00348
	Data Set 2	0.06055	0.00354
p = 0.5	Data Set 1	0.05872	0.00398
	Data Set 2	0.05893	0.00402

4.4.2. Graph Structure and Outbreak Analysis

The severity of every simulated outbreak is measured based upon the proportion of the population that becomes infected, the value of R_0, and the duration. These findings support similar results presented in Chapter 3. The results are summarized in Tables 4.6, 4.7, and 4.8. Although there are exceptions as noted in the tables, general trends are observed regarding each of the severity measures as discussed below.

The proportion of the population that becomes infected during the simulations after implementation of a vaccination policy (see Table 4.6) ranges from 15.4% to 88.7%. The lowest average occurs at $N = 250$ and $p = 0$ under the high contact vaccination policy. The highest average occurs at $N = 150$ and $p = 0.25$ under the low contact vaccination policy. Consistent with earlier experiments, the proportion of infected individuals increases with the number of non-local contacts regardless of the vaccination policy. This is an indication that restricted contacts have a tendency to confine the spread of an outbreak. Additionally, the proportion of infected individuals is found to be considerably higher in smaller populations in simulations in which there are no, or very few ($p = 0$ or $p = 0.01$) outside contacts. This disparity is not observed in simulations with a larger probability of non-local contacts ($p = 0.25$ and $p = 0.5$). The neighborhood size, $k = 6$, is held constant for these experiments regardless of the population size which may account for this discrepancy. In smaller populations, the neighborhood size is proportionally larger, thereby increasing the probability that the disease will transfer to a higher proportion of individuals in the population.

Average values of R_0, as shown in Table 4.7 range from 2.65 to 4.07 in simulations with vaccination implementation. The low value of 2.65 occurs at $N = 50$ and $p = 0$ under the random vaccination policy. The highest average of 4.07 occurs at $N = 50$ and $p = 0.5$ under the low contact vaccination policy. All vaccination strategies are shown to lower the value of R_0. Regardless of the population size or the type of vaccination, the value of R_0 tends to increase with the probability of outside contacts. The population size, N, does not appear to have as much influence over R_0 as the probability of non-local contacts and the vaccination method. Perhaps

TABLE 4.6. Comparison of Percent of Population Infected based on Specific Vaccination Strategies

Percent of Population Infected									
High Contact Vaccination					**Transmission Vaccination**				
N	$p=0$	$p=0.01$	$p=0.25$	$p=0.5$	N	$p=0$	$p=0.01$	$p=0.25$	$p=0.5$
50	62.1	66.3	83.3	86.9	50	63.3	65.4	85.1	84.3
150	26.4	47.2	85.7	84.2	150	32.4	41.0	83.2	84.2
250	15.4	36.7	85.7	82.2	250	23.2	28.8	84.7	84.1
Random Vaccination					**Spread Vaccination**				
N	$p=0$	$p=0.01$	$p=0.25$	$p=0.5$	N	$p=0$	$p=0.01$	$p=0.25$	$p=0.5$
50	65.6	73.9	86.3	86.3	50	72.6	69.1	83.8	85.9
150	34.4	52.7	86.7	85.4	150	48.6	47.6	83.6	84.4
250	21.3	43.3	85.9	87.5	250	34.2	41.8	84.5	84.1
Low Contact Vaccination					**No Vaccination**				
N	$p=0$	$p=0.01$	$p=0.25$	$p=0.5$	N	$p=0$	$p=0.01$	$p=0.25$	$p=0.5$
50	73.9	77.8	87.7	87.7	50	93.7	94.1	96.8	98.7
150	34.6	60.4	88.7	86.1	150	68.3	87.0	96.9	98.9
250	22.2	43.7	88.4	87.9	250	53.1	85.6	97.1	97.9

the most notable finding is that the value of R_0 is *not* a good predictor of the proportion of the population that will become infected without additional consideration of the graph structure. This implies that a given disease with an estimated R_0 value is not likely to manifest itself in the same manner under different population dynamics.

TABLE 4.7. Comparison of R_0 Values based on Specific Vaccination Strategies

R_0 Values									
High Contact Vaccination					**Transmission Vaccination**				
N	$p=0$	$p=0.01$	$p=0.25$	$p=0.5$	N	$p=0$	$p=0.01$	$p=0.25$	$p=0.5$
50	2.75	2.69	3.04	3.20	50	2.86	2.87	3.17	3.29
150	2.78	2.70	3.29	3.22	150	2.70	2.75	3.92	3.38
250	2.80	2.63	3.21	3.20	250	2.92	2.80	2.85	3.39
Random Vaccination					**Spread Vaccination**				
N	$p=0$	$p=0.01$	$p=0.25$	$p=0.5$	N	$p=0$	$p=0.01$	$p=0.25$	$p=0.5$
50	2.65	2.77	3.01	3.31	50	2.97	2.89	3.23	3.11
150	2.84	2.70	3.20	3.43	150	2.96	3.10	3.09	3.22
250	2.85	2.85	3.16	3.53	250	3.02	3.00	3.14	3.19
Low Contact Vaccination					**No Vaccination**				
N	$p=0$	$p=0.01$	$p=0.25$	$p=0.5$	N	$p=0$	$p=0.01$	$p=0.25$	$p=0.5$
50	2.83	2.94	3.39	4.07	50	3.18	3.24	3.66	4.05
150	2.94	3.11	3.46	3.58	150	3.02	2.94	3.56	4.15
250	2.73	3.03	3.45	3.67	250	3.09	2.99	3.54	4.05

The average duration, as shown in Table 4.8 ranges from a low of 30 days which occurs at $N = 50$ and $p = 0.5$ after low contact vaccination, as well as no vaccination, to a high average value, with vaccination, of 145 which occurs at $N = 250$ and $p = 0$ after spread vaccination and a high average value of 211 days without vaccination. Under the same vaccination strategy and the same graph structure, it is observed that larger populations sustain the disease for a longer period of time, but this is not an indication of the severity of the outbreak. As with R_0, there is not a direct correlation between the duration of the outbreak and the proportion of the population infected.

TABLE 4.8. Comparison of Outbreak Duration (in days) based on Specific Vaccination Strategies

Outbreak Duration (Days)									
High Contact Vaccination					**Transmission Vaccination**				
N	$p=0$	$p=0.01$	$p=0.25$	$p=0.5$	N	$p=0$	$p=0.01$	$p=0.25$	$p=0.5$
50	61	56	36	34	50	63	57	35	32
150	78	95	50	41	150	89	91	49	41
250	74	116	53	46	250	103	109	53	45
Random Vaccination					**Spread Vaccination**				
N	$p=0$	$p=0.01$	$p=0.25$	$p=0.5$	N	$p=0$	$p=0.01$	$p=0.25$	$p=0.5$
50	64	58	37	32	50	68	62	35	33
150	97	102	47	40	150	123	99	47	41
250	101	130	51	44	250	145	122	52	45
Low Contact Vaccination					**No Vaccination**				
N	$p=0$	$p=0.01$	$p=0.25$	$p=0.5$	N	$p=0$	$p=0.01$	$p=0.25$	$p=0.5$
50	65	60	33	30	50	68	61	32	30
150	99	104	46	38	150	167	123	43	37
250	110	122	51	43	250	211	178	47	41

4.4.3. Graph Structure and Vaccination Methods

All of the investigations presented in this dissertation indicate that the underlying social structure has an effect on the severity of a disease outbreak. Graphs that are ordered, i.e. contacts are limited to a specific neighborhood, do not facilitate disease spread as well as small-world or random networks. However, the indication that a specific vaccine strategy is always more effective on a particular graph structure is not supported by experiments presented herein. Nevertheless, based on the results shown in Tables 4.9, 4.10, and 4.11, vaccination of individuals with low contact centrality is generally not as effective as the other vaccination strategies. This is an important finding

because this is a strategy that is represented by policies that are directed at vaccinating the elderly and infants. Among the vaccination strategies of high contact, transmission, spread, and random, there is no confirmation that any one strategy is markedly more effective than the others, although a general pattern does appear. High contact and transmission vaccination demonstrate a tendency to produce a slightly lower proportion of infected population than the other policies, and random vaccination has a propensity to be more effective than low contact vaccination, but less effective than the other strategies.

TABLE 4.9. Vaccination Efficacy, Population Size N = 50: Vaccination Method (Percent of Population Infected, Percent of Outbreaks Infecting 20% of Population)

$p = 0$	$p = 0.01$
Contact $(62.1, 93)$	Transmission $(65.4, 87)$
Transmission $(63.6, 88)$	Contact $(66.3, 93)$
Random $(65.6, 91)$	Spread $(69.1, 95)$
Spread $(72.6, 95)$	Random $(73.9, 95)$
Low Contact $(73.9, 93)$	Low Contact $(77.8, 96)$
No Vaccination $(93.7, 99)$	No Vaccination $(94.1, 97)$
$p = 0.25$	$p = 0.5$
Contact $(83.3, 95)$	Transmission $(84.3, 96)$
Spread $(83.8, 96)$	Spread $(85.9, 98)$
Transmission $(85.1, 97)$	Random $(86.3, 98)$
Random $(86.3, 98)$	Contact $(86.9, 99)$
Low Contact $(87.7, 99)$	Low Contact $(87.7, 99)$
No Vaccination $(96.8, 98)$	No Vaccination $(98.7, 100)$

4.5. Summary

Emerging diseases pose a threat to society. One of the most practical defenses against such a threat is through vaccination. However, vaccine availability is often a concern that public health officials must address. Limited supplies create a dilemma regarding who should receive the existing doses. The purpose of the experiments presented in this chapter is to better understand the effects of vaccination within various social structures and to determine if any of the particular vaccination strategies under examination are more effective than the others.

TABLE 4.10. Vaccination Efficacy, Population Size N = 150: Vaccination Method (Percent of Population Infected, Number of Outbreaks 20%)

$p = 0$	$p = 0.01$
Contact $(26.4, 57)$	Transmission $(41.0, 73)$
Transmission $(32.4, 67)$	Contact $(47.2, 75)$
Random $(34.4, 71)$	Spread $(47.6, 81)$
Low Contact $(34.6, 74)$	Random $(52.7, 78)$
Spread $(48.6, 78)$	Low Contact $(60.4, 88)$
No Vaccination $(68.3, 91)$	No Vaccination $(87.0, 95)$
$p = 0.25$	$p = 0.5$
Transmission $(83.2, 95)$	Contact $(84.2, 96)$
Spread $(83.6, 98)$	Transmission $(84.2, 96)$
Contact $(85.7, 98)$	Spread $(84.4, 96)$
Random $(86.7, 98)$	Random $(85.4, 96)$
Low Contact $(88.7, 100)$	Low Contact $(86.1, 97)$
No Vaccination $(96.9, 98)$	No Vaccination $(98.9, 100)$

TABLE 4.11. Vaccination Efficacy, Population Size N = 250: Vaccination Method (Percent of Population Infected, Number of Outbreaks 20%)

$p = 0$	$p = 0.01$
Contact $(15.4, 33)$	Transmission $(28.8, 56)$
Random $(21.3, 44)$	Contact $(36.7, 65)$
Low Contact $(22.2, 46)$	Spread $(41.8, 73)$
Transmission $(23.2, 52)$	Random $(43.3, 71)$
Spread $(34.2, 66)$	Low Contact $(43.7, 72)$
No Vaccination $(53.1, 85)$	No Vaccination $(85.6, 94)$
$p = 0.25$	$p = 0.5$
Spread $(84.5, 96)$	Contact $(82.2, 94)$
Transmission $(84.7, 97)$	Transmission $(84.1, 96)$
Contact $(85.7, 98)$	Spread $(84.1, 96)$
Random $(85.9, 97)$	Random $(87.5, 99)$
Low Contact $(88.4, 100)$	Low Contact $(87.9, 99)$
No Vaccination $(97.1, 98)$	No Vaccination $(97.9, 99)$

In corroboration with the results of Chapter 3, it is observed that outbreak severity is diminished when social contacts are confined to a specific neighborhood. The severity increases as the probability of contacts outside of the neighborhood rises. This finding is consistent for all population sizes and all vaccination methods. It is additionally observed that the proportion of infected

individuals in smaller populations is found to be considerably higher than that of larger populations in simulations in which there are no, or very few, non-local contacts.

All vaccination methods are shown to lower the value of R_0 and the proportion of the infected population, but no single policy is determined to be significantly more effective than the others. Alternatively, a vaccination policy based on low contact is consistently found to be less effective than the other policies. Implementation in real life is often related to random vaccination, low contact vaccination, or a combination thereof. Under the specific circumstances of this study, there is no substantial gain in changing this policy.

CHAPTER 5

CONCLUSION

Epidemiology has evolved and continues to evolve with the advancement of technology. Contributions to this field by Hippocrates, John Snow, and others created a foundation for current and future research. It is not only possible, but essential, that the discipline of computer science integrate with epidemiology and public health to combat disease spread. This dissertation is one effort of many that is designed to bridge the gap among these fields.

As computational models become more prevalent, it is important to recognize that the structure used to model a social network has an influence on the results of the simulated outbreak. If the social network is not accurately modeled, the results obtained may be unreliable. A basic assumption for the work presented in this dissertation is that there is a strong connection between the underlying social network and disease spread. Social networks, made up of individuals or groups who are connected through family, friendship, work relations, or another type of interdependent bond, can be modeled as a graph in which each individual or group is represented by a node and each relationship is denoted as an edge between two nodes. The structure of a graph that accurately represents a social network is a subject of debate. An ordered graph structure implies that individuals are only allowed to make contacts within their neighborhood, while a random structure indicates that contacts can be initiated with anyone in the population. Small-world graphs are those that fall between ordered and random and are commonly used to represent social networks. Many experts agree that a small-world graph, as discussed in Section 2.4.2 characterizes two essential properties of social networks, clustering and the small-world effect. Clustering refers to the tendency of people to form groups and the small-world effect theorizes that, on average, there is a relatively short distance between any two individuals in a population. This study explores ordered, random, and small-world graphs as underlying social networks.

Investigations presented herein are the result of multiple analyses of disease spread in simulated environments. The creation of social networks and subsequent disease outbreaks are based on graph theoretical concepts. This design allows the established field of graph theory to be applied to the area of epidemiology. A well-recognized paradigm, the SIR model, is superimposed onto the social network graph structure. Initial investigations in this dissertation measure changes in outbreak severity as a result of modifications to the social structure. Subsequent experiments explore the efficacy of several vaccination strategies.

Several conclusions are drawn from these experiments. As a social network progresses from ordered to random, the neighborhood size becomes less important. A small neighborhood size with a low probability for contact outside of the neighborhood has a significant effect on the severity of a disease outbreak, however, as the neighborhood size or the probability of random contacts increase, the variation in severity is very minor. In fact, as the neighborhood size approaches the size of the population, the structure of the graph inherently moves from ordered to random regardless of the probability of random contacts.

It is also observed that the duration of an outbreak and the initial number of secondary infections, R_0, are not reliable indicators of the severity of an outbreak. A short duration may result due to the lack of progression of the disease throughout the population, infecting very few individuals, or it may result because the disease spreads very quickly, infecting many. A value of $R_0 > 1$ generally indicates that an epidemic is likely to occur, however, this is not always the case. In circumstances when the neighborhood size is limited and the probability for random contacts is low, it is observed that $R_0 > 1$ is not an accurate indicator. Although duration and R_0 are useful in conjunction with the proportion of population infected, they do not provide enough information to stand alone.

Chapter 4 explores the efficacy of targeted vaccination policies under the assumption of a limited supply of vaccine. Unlike the earlier experiments in which the contact graph is created dynamically as the disease spreads, these experiments first create a social contact graph so that

key individuals can be identified for vaccination. Vaccination methods include high contact, transmission, spread, random, and low contact. After vaccination of ten percent of the population, an outbreak is simulated and measurements of R_0, duration, and the proportion of the population infected are recorded. All vaccination methods are found to lower the value of R_0 and decrease the proportion of the population infected. Vaccination of individuals who make fewer contacts is found to be the least effective strategy, but none of the vaccination methods are consistently more effective than the others. Random vaccination generally attains better results than low contact vaccination, but is found to be slightly less successful than the other strategies.

5.1. Implications to Public Health and Policy Development

The experiments presented in this dissertation suggest that both reducing an individual's effective neighborhood size and limiting the number random contacts have the potential to decrease the severity of a disease outbreak. This does not necessarily require an alteration of the actual personal connections, rather a reduction in the ability for the connections to transfer disease. Public awareness, prophylactic use, quarantine, and vaccination are all methods that can effectively reduce disease transfer. Early intervention may prevent the occurrence of an epidemic/pandemic or limit the severity of an outbreak. When vaccination methods are employed, the findings herein suggest that random vaccination is nearly as effective as targeted policies if the proportion of the population to be vaccinated is low (10% for the studies conducted in this research). Future studies may reveal that targeted strategies are more effective if the proportion of individuals that are vaccinated is increased.

5.2. Limitations

It is not possible to simulate a disease outbreak with complete accuracy. The variance in disease and population parameters along with the random nature of disease spread make it a tremendous challenge to portray an epidemic/pandemic in a simulated environment. Nevertheless, this is a challenge that must be addressed in order to advance our knowledge and understanding of disease dynamics.

The computational complexity of the graph algorithms implemented in this research restrict the breadth and depth of the experiments presented in this dissertation. Many of the experiments presented in Chapter 4 have execution times in excess of 30 hours. Therefore, it is impractical to increase the graph size to a level that simulates a larger population of individuals, such as a metropolitan area. It should be emphasized, however, that nodes in a graph can represent groups as well as individuals.

In the experiments conducted for this research, disease attributes are held constant and parameter changes are limited to graph structure. Altering the transmission rate, latent period, and/or infectious period will almost certainly change the rate at which a disease progresses through a population, but the relationship between graph structure and outbreak severity is likely to be the same. A more significant limitation of this study is that vaccination policies are only implemented over ten percent of the population. Vaccination of a larger portion of the population may reveal a more distinctive pattern among vaccination strategies. Although a more comprehensive study may provide additional insights regarding the relationship between graph structure and disease dynamics, the findings in this dissertation are substantial and provide direction for future studies.

5.2.1. Future Work

Computation epidemiology is a growing field with unlimited open questions. From the experiments presented in this dissertation, there is much room for expansion. While population, neighborhood size, and probability of non-local contacts are altered, many other parameters are held constant. Changes in these parameters, such as the contact rate, days latent, and days infectious, may produce additional results. Additionally, an increase in population size could be performed to test scalability. In regard to neighborhood size, it can be argued that an individual makes approximately the same number of contacts in a day regardless of the population size or, alternatively, it might be assumed that an increase in population size increases the contact rate. The experiments involving vaccination (Chapter 4) maintain a constant neighborhood size, leaving a room for future research involving variable neighborhood sizes. The vaccination simulations presented in Chapter 4 are designed to model a situation in which the vaccine supply is very limited.

Future studies might include a gradual increase in the proportion of the population vaccinated to compare the efficacy as more individuals become vaccinated.

Computational epidemiology is becoming increasingly important in our global society. The vast nature of this field of study requires a concerted effort from many agencies and across several disciplines. Successful development of reliable models depends on a collaborative effort and ongoing research such as that presented in this paper. No single endeavor is sufficient, but each contribution is valuable.

APPENDIX

CDC VACCINATION TABLE

TABLE: Self-reported influenza vaccination coverage trends 1989 - 2008 among adults by age group, risk group, race/ethnicity, health-care worker status, and pregnancy status, United States, National Health Interview Survey (NHIS)

Characteristics	Survey Year																
	1989	1991	1993	1994	1995	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Age Group																	
18-49	3.4±0.2 (54664)	6.1±0.4 (26900)	10.2±0.6 (12954)	12.1±0.7 (11993)	13.1±0.7 (10757)	14.3±0.5 (23328)	15.5±0.6 (19546)	16.4±0.7 (18232)	17.1±0.7 (19409)	15.1±0.6 (20031)	16.3±0.7 (18427)	16.8±0.7 (18033)	17.9±0.7 (18039)	10.4±0.5 (17653)	15.5±0.8 (13861)	17.7±0.8 (13030)	20.0±0.9 (11856)
50-64	10.6±0.6 (15655)	15.0±1.0 (7618)	23.0±1.6 (3620)	25.6±1.6 (3443)	27.0±1.8 (3000)	31.9±1.4 (6513)	33.1±1.4 (6194)	34.1±1.3 (6110)	34.6±1.5 (6403)	32.1±1.3 (6804)	34.0±1.3 (6424)	36.8±1.4 (6666)	35.9±1.4 (6933)	22.9±1.1 (7251)	33.1±1.5 (5471)	36.2±1.7 (5408)	39.5±1.6 (5185)
≥65	30.4±1.1 (14244)	41.7±1.6 (8453)	52.0±1.6 (4198)	55.3±1.8 (3971)	58.2±1.8 (3442)	63.2±1.4 (6878)	63.3±1.4 (6257)	65.7±1.5 (5887)	64.3±1.4 (6091)	63.0±1.4 (6046)	65.6±1.5 (5757)	65.5±1.4 (5662)	64.6±1.4 (5922)	59.6±1.5 (5940)	64.1±1.8 (4573)	66.4±1.6 (4474)	67.0±1.8 (4358)
Age by Risk Status																	
18-49																	
High Risk	N/A	N/A	N/A	N/A	N/A	20.7±1.6 (3263)	22.7±1.8 (2758)	22.6±1.9 (2505)	24.7±2.0 (2676)	20.9±1.7 (3099)	23.1±2.0 (2482)	24.2±2.1 (2341)	26.0±2.1 (2555)	18.1±1.7 (2589)	24.5±2.5 (1872)	27.2±2.7 (1670)	29.8±2.6 (1713)
Not High Risk	N/A	N/A	N/A	N/A	N/A	16.1±0.6 (23292)	14.4±0.6 (16673)	15.4±0.8 (15649)	16.0±0.7 (16627)	14.0±0.6 (16814)	15.3±0.7 (15891)	15.8±0.7 (15654)	16.6±0.7 (15442)	9.1±0.5 (15031)	14.1±0.8 (11954)	16.3±0.8 (11338)	18.4±1.0 (10125)
50-64																	
High Risk	N/A	N/A	N/A	N/A	N/A	40.5±2.6 (3003)	43.4±2.6 (1826)	45.0±2.7 (1775)	43.9±2.5 (1920)	40.9±2.3 (2147)	43.6±2.4 (1969)	46.3±2.6 (2006)	45.5±2.5 (2104)	33.9±2.2 (2352)	44.4±3.1 (1705)	46.0±3.1 (1641)	49.6±3.1 (1629)
Not High Risk	N/A	N/A	N/A	N/A	N/A	28.2±1.6 (4416)	28.9±1.6 (4289)	29.7±1.5 (4274)	30.6±1.7 (4421)	28.2±1.6 (4568)	29.8±1.5 (4431)	32.7±1.6 (4637)	32.1±1.6 (4807)	17.8±1.2 (4888)	28.2±1.6 (3745)	32.3±1.9 (3749)	35.2±1.9 (3538)
Age by Race/Ethnicity																	
18-49																	
White Not Hispanic	3.3±0.2 (40196)	5.6±0.5 (19655)	10.0±0.7 (9525)	12.4±0.8 (8715)	12.9±0.9 (6962)	15.1±0.7 (13831)	16.1±0.8 (12162)	17.1±0.9 (11249)	18.1±0.9 (11739)	15.5±0.8 (12100)	16.9±0.8 (11032)	18.0±0.9 (10725)	19.8±0.9 (10533)	11.1±0.6 (10306)	16.6±1.0 (7312)	19.0±1.1 (6905)	21.6±1.2 (6455)
Black Not Hispanic	3.9±0.5 (7523)	8.1±1.2 (3736)	10.8±1.7 (1720)	10.0±1.9 (1551)	15.6±2.4 (1394)	13.2±1.3 (3313)	13.1±1.6 (2735)	15.5±1.8 (2688)	14.0±1.3 (3902)	15.0±1.8 (2882)	15.7±1.7 (2691)	16.9±1.6 (2582)	14.3±1.8 (2730)	9.8±1.4 (2594)	14.8±1.9 (2468)	14.6±1.9 (2182)	17.4±2.2 (1923)
Hispanic	3.9±0.6 (4653)	6.8±1.2 (3472)	9.1±2.2 (1068)	10.8±2.2 (1123)	11.1±1.7 (1900)	10.3±1.1 (4294)	13.1±1.3 (3884)	12.9±1.6 (3583)	13.4±1.4 (3970)	11.9±1.2 (4161)	12.4±1.2 (3898)	11.8±1.3 (3960)	11.6±1.2 (4015)	7.8±1.0 (4005)	11.0±1.3 (3108)	13.9±1.6 (3004)	14.9±1.7 (2559)
API	2.8±1.1 (1392)	7.5±2.2 (695)	11.3±2.8 (417)	14.9±3.0 (369)	14.8±4.7 (312)	14.4±3.5 (713)	17.5±3.4 (594)	17.7±4.1 (517)	21.1±3.5 (607)	17.4±2.9 (675)	19.7±3.4 (633)	18.1±3.9 (370)	21.5±4.9 (370)	9.0±3.1 (380)	20.2±4.1 (574)	20.5±4.3 (505)	23.8±5.1 (534)
50-64																	
White Not Hispanic	11.1±0.7 (12252)	15.4±1.1 (6031)	23.8±1.8 (2892)	26.8±1.8 (2696)	28.4±2.1 (2208)	33.8±1.7 (4612)	35.0±1.6 (4511)	35.8±1.5 (4412)	37.0±1.7 (4539)	34.6±1.6 (4825)	35.8±1.5 (4654)	38.8±1.6 (4758)	38.3±1.6 (4848)	24.5±1.3 (5105)	34.8±1.7 (3754)	38.2±2.1 (3661)	41.5±1.9 (3490)
Black Not Hispanic	8.5±1.5 (2060)	11.4±2.6 (1012)	15.0±3.2 (458)	17.5±4.0 (445)	19.8±4.4 (349)	22.5±3.1 (899)	24.7±3.6 (811)	27.3±3.8 (766)	23.8±3.0 (878)	23.2±3.9 (966)	28.0±3.5 (785)	28.4±3.3 (876)	26.0±3.4 (931)	19.9±3.0 (997)	28.2±3.9 (818)	29.1±3.6 (827)	35.8±4.1 (808)
Hispanic	7.7±2.6 (856)	14.3±3.4 (362)	19.1±7.3 (178)	22.7±7.6 (194)	23.8±5.2 (348)	22.8±3.4 (809)	24.1±4.2 (693)	26.0±3.4 (774)	22.3±3.5 (783)	22.1±3.4 (821)	25.7±3.8 (790)	27.3±3.7 (847)	27.7±3.6 (916)	15.4±2.7 (891)	25.0±3.8 (647)	27.0±4.2 (659)	29.4±4.2 (617)
API	*	16.1±8.8 (142)	33.6±10.2 (54)	22.3±11.6 (58)	27.0±15.9 (54)	33.6±8.5 (140)	31.2±8.3 (132)	28.2±8.3 (115)	35.5±9.6 (148)	20.6±7.8 (134)	31.0±8.5 (151)	31.7±10.7 (78)	33.7±9.7 (113)	12.7±6.6 (121)	27.7±9.4 (130)	31.4±10.5 (127)	33.0±9.5 (128)
≥65																	
White Not Hispanic	32.1±1.2 (11871)	43.4±1.7 (7144)	54.0±1.7 (3585)	58.1±1.9 (3333)	60.7±2.0 (2761)	65.8±1.5 (5481)	65.6±1.5 (4934)	67.9±1.6 (4581)	66.6±1.6 (4744)	65.4±1.5 (4691)	68.6±1.6 (4485)	68.7±1.4 (4401)	67.3±1.5 (4562)	63.2±1.6 (4609)	67.2±2.1 (3267)	69.0±1.9 (3227)	69.8±2.0 (3113)
Black Not Hispanic	17.8±2.3 (1606)	27.5±4.1 (904)	32.6±4.9 (423)	39.0±5.9 (414)	39.9±5.6 (361)	44.8±4.4 (774)	45.8±4.4 (667)	49.7±4.4 (642)	48.0±4.5 (678)	48.1±4.5 (662)	49.6±4.5 (626)	48.0±4.6 (609)	45.4±4.4 (647)	39.7±4.2 (669)	46.5±4.5 (661)	55.4±4.6 (594)	50.6±4.7 (583)
Hispanic	23.8±4.8 (464)	34.0±8.6 (259)	47.3±12.8 (121)	37.9±8.1 (145)	49.9±7.8 (245)	52.7±5.9 (520)	50.3±5.1 (532)	55.1±5.1 (543)	55.7±5.0 (566)	51.9±5.2 (567)	48.5±5.0 (527)	45.4±5.2 (528)	54.6±5.2 (584)	41.7±5.0 (511)	44.8±5.9 (410)	52.1±5.9 (449)	54.5±5.6 (438)
API	19.4±8.5 (123)	30.1±9.2 (74)	54.7±18.1 (30)	43.1±22.1 (35)	50.9±15.8 (50)	51.2±11.2 (80)	67.1±10.6 (93)	71.7±9.4 (87)	60.1±11.0 (82)	57.5±11.6 (89)	57.8±11.1 (96)	63.6±15.5 (58)	52.7±16.8 (54)	56.1±14.1 (68)	61.4±11.2 (106)	62.7±11.3 (103)	58.8±12.1 (101)

FIGURE A.1. CDC vaccination table

18-49 High Risk by Race/Ethnicity																
White Not Hispanic	N/A	N/A	N/A	N/A	N/A	22.3±1.9 (2153)	22.8±2.2 (1812)	22.4±2.3 (1640)	25.2±2.5 (1759)	21.0±2.0 (2013)	23.0±2.4 (1591)	24.9±2.5 (1466)	26.6±2.6 (1584)	18.4±2.2 (1633)	24.8±3.1 (1078)	29.0±3.5 (956)	30.4±3.2 (1058)
Black Not Hispanic	N/A	N/A	N/A	N/A	N/A	17.3±3.5 (519)	22.1±4.8 (452)	20.4±5.2 (401)	22.6±4.3 (426)	20.0±4.0 (496)	23.9±5.2 (410)	24.3±5.4 (378)	22.8±4.8 (424)	15.9±3.8 (427)	23.6±5.5 (370)	19.6±5.5 (302)	29.8±6.7 (304)
Hispanic	N/A	N/A	N/A	N/A	N/A	12.6±3.2 (498)	22.1±5.0 (434)	24.2±5.2 (396)	22.7±4.3 (410)	19.2±3.8 (493)	21.2±4.8 (394)	20.0±4.8 (433)	21.2±4.9 (468)	19.4±4.4 (452)	20.6±5.0 (332)	24.6±5.7 (327)	23.7±5.9 (281)
API	N/A	N/A	N/A	N/A	N/A	*	*	*	*	25.5±14.2 (56)	29.5±15.5 (48)	*	44.7±20.2 (30)	*	39.5±17.3 (47)	*	32.4±18.6 (34)
50-64 High Risk by Race/Ethnicity																
White Not Hispanic	N/A	N/A	N/A	N/A	N/A	41.8±3.1 (1359)	44.8±3.0 (1294)	47.1±3.1 (1234)	46.1±3.0 (1282)	43.8±2.8 (1478)	45.2±2.7 (1382)	47.9±3.0 (1406)	47.9±3.0 (1421)	35.4±2.6 (1649)	45.2±3.7 (1137)	46.7±3.7 (1044)	49.1±3.8 (1048)
Black Not Hispanic	N/A	N/A	N/A	N/A	N/A	34.4±5.8 (332)	35.3±7.4 (271)	34.1±7.2 (243)	35.5±6.0 (344)	30.4±5.4 (359)	38.1±7.5 (281)	39.7±6.6 (285)	33.9±6.4 (329)	32.2±6.1 (350)	43.5±6.3 (298)	44.6±6.6 (310)	48.7±7.4 (288)
Hispanic	N/A	N/A	N/A	N/A	N/A	33.1±6.5 (253)	38.3±8.0 (211)	38.1±7.0 (248)	34.5±8.0 (242)	28.0±7.1 (258)	34.6±8.2 (244)	39.7±7.2 (270)	38.5±6.6 (274)	24.5±5.6 (290)	42.3±8.6 (207)	38.6±7.4 (216)	47.7±9.3 (214)
API	N/A	N/A	N/A	N/A	N/A	45.1±21.1 (30)	46.1±18.5 (33)	*	*	*	46.6±18.7 (41)	*	38.8±17.0 (34)	*	*	*	57.0±21.2 (31)
Age by Diabetes Status																
18-49																
With Diabetes	N/A	N/A	N/A	N/A	N/A	25.1±4.7 (472)	31.2±5.2 (434)	32.7±5.5 (419)	32.5±5.3 (483)	28.6±4.3 (542)	31.9±5.3 (486)	36.4±5.6 (461)	34.0±4.6 (553)	29.5±4.7 (571)	35.0±5.4 (470)	30.0±5.4 (406)	37.0±6.0 (388)
Without Diabetes	N/A	N/A	N/A	N/A	N/A	14.1±0.5 (21627)	15.1±0.6 (18981)	16.0±0.7 (17724)	16.8±0.7 (18803)	14.7±0.6 (19374)	15.9±0.7 (17925)	16.4±0.7 (17567)	17.4±0.7 (17477)	9.8±0.5 (17073)	14.9±0.8 (13381)	17.4±0.8 (12619)	19.5±0.9 (11465)
50-64																
With Diabetes	N/A	N/A	N/A	N/A	N/A	43.2±4.6 (671)	47.4±4.8 (604)	50.9±4.6 (621)	46.6±4.2 (678)	47.7±4.2 (765)	46.8±4.4 (744)	51.6±4.0 (770)	48.9±3.7 (835)	40.3±3.5 (936)	51.0±5.0 (716)	45.4±4.7 (717)	53.7±4.6 (737)
Without Diabetes	N/A	N/A	N/A	N/A	N/A	30.8±1.4 (5731)	31.5±1.4 (5494)	32.1±1.3 (5410)	33.1±1.6 (5647)	30.1±1.4 (5938)	32.4±1.3 (5670)	35.0±1.5 (5891)	34.3±1.5 (6090)	20.5±1.1 (6313)	30.6±1.5 (4746)	34.9±1.8 (4687)	37.3±1.7 (4443)
Age by Asthma Status																
18-49																
With Asthma	N/A	N/A	N/A	N/A	N/A	23.2±3.1 (901)	21.5±3.5 (762)	23.3±3.7 (703)	28.1±3.9 (734)	26.6±3.3 (880)	23.9±3.2 (763)	28.9±4.3 (675)	28.4±4.0 (667)	21.5±3.4 (728)	24.6±4.4 (534)	30.4±5.3 (528)	30.7±4.9 (468)
Without Asthma	N/A	N/A	N/A	N/A	N/A	13.9±0.5 (21414)	15.2±0.6 (18759)	16.1±0.7 (17516)	16.7±0.7 (18661)	14.6±0.6 (19131)	14.6±0.6 (19131)	16.4±0.7 (17345)	17.5±0.7 (17357)	9.9±0.5 (16911)	15.2±0.8 (13314)	17.2±0.8 (12504)	19.6±0.9 (11378)
50-64																
With Asthma	N/A	N/A	N/A	N/A	N/A	46.0±6.7 (250)	55.2±6.7 (257)	50.3±7.8 (226)	55.1±7.3 (247)	41.9±6.6 (279)	51.0±6.7 (263)	47.9±7.0 (275)	52.4±6.4 (292)	40.2±6.3 (310)	46.8±7.8 (258)	57.9±7.6 (228)	50.8±8.1 (231)
Without Asthma	N/A	N/A	N/A	N/A	N/A	31.3±1.5 (6248)	32.2±1.4 (5919)	33.5±1.3 (5875)	33.8±1.6 (6145)	31.8±1.3 (6514)	31.8±1.3 (6514)	36.3±1.4 (6376)	35.3±1.4 (6624)	22.2±1.1 (6931)	32.5±1.5 (5205)	35.4±1.8 (5173)	39.1±1.6 (4950)
HCW Status																
Health-Care Workers	10.0±0.9 (4849)	16.7±1.7 (2646)	25.7±2.5 (1387)	31.4±2.0 (1292)	29.5±2.0 (1153)	34.0±2.1 (2387)	37.0±2.5 (2129)	36.4±2.6 (2013)	37.6±2.4 (2165)	36.1±2.5 (2270)	38.4±2.5 (2066)	40.1±3.5 (2146)	41.9±2.5 (2031)	32.9±2.1 (2143)	41.7±3.2 (1586)	45.3±3.0 (1643)	48.0±2.9 (1608)
Other Workers	4.6±0.2 (49545)	7.7±0.4 (23526)	12.3±0.7 (11319)	13.8±0.8 (10535)	14.9±0.8 (9571)	17.3±0.6 (19715)	18.0±0.7 (17769)	19.1±0.7 (16821)	20.3±0.7 (17879)	18.1±0.6 (18836)	27.2±0.6 (18826)	20.9±0.8 (16500)	19.4±0.8 (16216)	10.2±0.5 (16181)	17.2±0.8 (12636)	19.8±0.9 (11910)	22.2±1.0 (10798)
Pregnancy Status																
Pregnant	N/A	N/A	N/A	N/A	N/A	11.1±4.6 (372)	8.4±3.2 (335)	7.6±3.3 (332)	10.0±3.4 (337)	10.4±3.6 (294)	12.4±3.9 (319)	12.8±4.4 (315)	12.9±5.0 (263)	15.6±5.0 (303)	13.8±4.9 (240)	15.3±4.9 (254)	19.1±7.1 (187)
Not Pregnant	N/A	N/A	N/A	N/A	N/A	11.9±0.8 (8515)	14.3±0.9 (7239)	14.4±1.0 (6992)	15.3±1.0 (7406)	13.5±0.9 (7391)	15.2±1.1 (6956)	15.8±1.1 (6775)	17.6±1.2 (6657)	10.5±0.9 (6521)	15.5±1.2 (5189)	17.1±1.3 (4874)	20.0±1.4 (4379)

Table cell entries are % reporting influenza vaccination in the past 12 months, ±the 95% confidence interval half-width (SE), with the unweighted sample size (n).
An asterix (*) indicates the estimate was unreliable due to the n being less than 30 or the SE relative to the estimate was greater than 0.3 (n<30 or RSE>0.3).
N/A (not available) indicates the years when the characteristic was not included in the survey. Nasal spray was included starting with 2006.

FIGURE A.2. CDC vaccination table

BIBLIOGRAPHY

[1] "Seasonal influenza," Nov 2009. [Online]. Available: http://www.cdc.gov/flu/about/qa/disease.htm

[2] "World health organization," Oct 2009. [Online]. Available: http://www.who.int

[3] "Maptitude geographic information system," Feb 2010. [Online]. Available: http://www.caliper.com/Maptitude/publichealth/default.htm

[4] A. Akella, S. Chawla, A. Kannan, and S. Seshan, "On the scaling of congestion in the internet graph," *SIGCOMM Comput. Commun. Rev.*, vol. 34, no. 3, pp. 43–56, 2004.

[5] L. Allen and A. Burgin, "Comparison of deterministic and stochastic sis and sir models in discrete time," *Mathematical Biosciences*, vol. 163, no. 1, pp. 1–33, Jan 2000.

[6] E. Allman and J. Rhodes, *Mathematical models in biology: an introduction*. pub-CAMBRIDGE:adr: Cambridge University Press, 2003.

[7] R. Anderson and R. May, *Infectious Diseases of Humans Dynamics and Control*. Oxford University Press, 1992.

[8] R. M. Anderson and R. M. May, "Directly transmitted infectious diseases: Control by vaccination," *Science*, vol. 215, no. 4536, pp. 1053–1060, 1982. [Online]. Available: http://www.jstor.org/stable/1688362

[9] A. Barabsi, R. Albert, and H. Jeong, "Scale-free characteristics of random networks: The topology of the world wide web," *Diameter of the World Wide Web. Nature*, vol. 401, p. 130, 1999.

[10] C. Barrett, K. Bisset, S. Eubank, X. Feng, and M. Marathe, "Episimdemics: an efficient algorithm for simulating the spread of infectious disease over large realistic social networks," *SC '08: Proceedings of the 2008 ACM/IEEE conference on Supercomputing*.

[11] D. BarthJones, A. Adams, and J. Koopman, "Monte carlo simulation experiments for analysis of HIV vaccine effects and vaccine trial design," in *WSC '00: Proceedings of the 32nd conference on Winter simulation*. San Diego, CA, USA: Society for Computer Simulation International, 2000, pp. 1985–1994.

[12] K. Bauman, RF, S. Ennett, AH, and V. Foshee, "Adding valued data to social network measures: Does it add to associations with adolescent substance use?" *Social Networks*, vol. 29, no. 1, pp. 1 – 10, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/B6VD1-4J2M6X9-1/2/82c3b3333c4354d79afaacad086d20fb

[13] K. Bisset, J. Chen, X. Feng, V. Kumar, and M. Marathe, "Epifast: a fast algorithm for large scale realistic epidemic simulations on distributed memory systems," in *ICS '09: Proceedings of the 23rd international conference on Supercomputing*. New York, NY, USA: ACM, 2009, pp. 430–439.

[14] U. Brandes, "On variants of shortest-path betweenness centrality and their generic computation," *Social Networks*, vol. 30, no. 2, pp. 136 – 145, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/B6VD1-4RFJ4K1-1/2/3a32e9305e56b9844dfb72cd93c73aa6

[15] ——, "Social network analysis and visualization," *IEEE Signal Processing Magazine*, vol. 25, no. 6, pp. 147–151, 2008.

[16] E. Britannica, "John graunt," http://www.britannica.com/EBchecked/topic/242312/John-Graunt, Oct 2009.

[17] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," *KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*.

[18] I. Chlamtac and A. Faragó, "A new approach to the design and analysis of peer-to-peer mobile networks," *Wireless Networks*, vol. 5, no. 3, pp. 149–156, 1999. [Online]. Available: http://dx.doi.org/10.1023/A:1019186624837

[19] G. Chowell, C. Ammon, N. Hengartner, and J. Hyman, "Estimation of the reproductive number of the spanish flu epidemic in geneva, switzerland," *Vaccine*, vol. 24, no. 44-46, pp. 6747–50, Nov 2006.

[20] G. Chowell, C. Castillo-Chavez, P. Fenimore, C. Kribs-Zaleta, L. Arriola, and J. Hyman, "Model parameters and outbreak control for sars," *Emerging Infectious Diseases*, vol. 10, no. 7, pp. 1258–63, July 2004. [Online]. Available: http://www.cdc.gov/ncidod/EID/vol10no7/03-0647.htm

[21] S. Crosier, "Center for spatially integrated social science," http://www.csiss.org/classics/content/8, Feb 2010.

[22] S. Dorogovtsev and J. Mendes, "Evolution of networks." *Advances in Physics*, vol. 51, no. 4, pp. 1079 – 1187.

[23] N. Du, C. Faloutsos, B. Wang, and L. Akoglu, "Large human communication networks: patterns and a utility-driven generator," *KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*.

[24] J. DW Bauer and M. Mohtashemi, "An application of parallel monte carlo modeling for real-time disease surveillance," in *WSC '08: Proceedings of the 40th Conference on Winter Simulation*. Winter Simulation Conference, 2008, pp. 1029–1037.

[25] D. Eppstein and M. Goodrich, "Studying (non-planar) road networks through an algorithmic lens," *GIS '08: Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*.

[26] P. Erdős and A. Rényi, "On the evolution of random graphs," *Publ. Math. Inst. Hungar. Acad. Sci.*, vol. 7, pp. 17–61, 1960.

[27] S. Eubank, H. Guclu, V. Kumar, M. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, "Modelling disease outbreaks in realistic urban social networks," *Nature*, vol. 429, no. 6988, pp. 180–184, May 2004.

[28] G. Evans, "With h1n1 vaccine shortage expected, highest-risk groups go to front of line," *Occupational Health Management*.

[29] C. Farrington and H. Whitaker, “Estimation of effective reproduction numbers for infectious diseases using serological survey data,” *Biostat*, vol. 4, no. 4, pp. 621–632, 2003. [Online]. Available: http://biostatistics.oxfordjournals.org/cgi/content/abstract/4/4/621

[30] D. Fell and A. Wagner, “The small world of metabolism,” *Nature Biotech.*, vol. 18, pp. 1121–1122, 2000.

[31] W. Fischer, “Thomas sydenham, the english hippocrates,” *The Canadian Medical Association Journal*, vol. III, no. 11, pp. 931–946, 1913.

[32] L. Freeman, “A set of measures of centrality based on betweenness,” *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977. [Online]. Available: http://www.jstor.org/stable/3033543

[33] ——, “Centrality in social networks: Conceptual clarification,” *Social Networks*, vol. 1, no. 3, pp. 215–239, 1979. [Online]. Available: http://dx.doi.org/10.1016/0378-8733(78)90021-7

[34] A. Galvani, T. Reluga, and G. Chapman, “Long-standing influenza vaccination policy is in accord with individual self-interest but not with the utilitarian optimum,” *Proc Natl Acad Sci USA*, vol. 104, no. 13, pp. 5692–5697, March 2007.

[35] A. Gardner, “Production problems plague delivery of swine flu vaccine,” *HealthDay Consumer News Service*, March 2009.

[36] G. Ghoshal, L. Sander, and I. Sokolov, “Sis epidemics with household structure: the self-consistent field method,” *Mathematical Biosciences*, vol. 190, no. 1, pp. 71–85, July 2004.

[37] P. Grammaticos and A. Diamantis, “Useful and unknown views of the father of modern medicine, hippocrates and his teacher democritus,” *Hellenic Journal of Nuclear Medicine*, vol. 11, no. 1, pp. 2–4, 2008. [Online]. Available: http://web.auth.gr/nuclmed/magazine/eng/jan08/editorial.htm

[38] M. Green, D. Freedman, and L. Gordis, “Reference manual on scientific evidence, second ed,” Public Health, 2000. [Online]. Available: http://www.fjc.gov/library/fjc_catalog.nsf

[39] J. Heffernan, R. Smith, and L. Wahl, “Perspectives on the basic reproductive ratio,” *Journal of the Royal Society Interface*, vol. 2, pp. 281–293, June 2005.

[40] J. Hietaniemi, “Cpan graph-0.91,” Jan 2009. [Online]. Available: http://search.cpan.org/~jhi/Graph-0.91/

[41] A. Hinman, W. Orenstein, J. Santoli, L. Rodewald, and S. Cochi, “Vaccine shortages: History, impact, and prospects for the future,” *Annual Review of Public Health*, vol. 27, no. 1, pp. 235–259, 2006.

[42] S. Huang, “A new seir epidemic model with applications to the theory of eradication and control of diseases, and to the calculation of R_0,” *Mathematical Biosciences*, vol. 215, no. 1, pp. 84 – 104, 2008.

[43] T. Jun, J. Kim, B. Kim, and M. Choi, “Consumer referral in a small world network,” *Social Networks*, vol. 28, no. 3, pp. 232 – 246, 2006.

[44] G. Karypis, R. Aggarwal, V. Kumar, and S. Shekhar, “Multilevel hypergraph partitioning: applications in VLSI domain,” *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*.

[45] L. Kazemi, C. Shahabi, M. Sharifzadeh, and L. Vincent, “Optimal traversal planning in road networks with navigational constraints,” *GIS ’07: Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*.

[46] W. Kermack and A. McKendrick, “A contribution to the mathematical theory of epidemics,” *Proceedings of the Royal Society of London*, vol. 115, no. 772, pp. 700–721, Aug 1927.

[47] P. Killworth, C. McCarty, H. Bernard, and M. House, “The accuracy of small world chains in social networks,” *Social Networks*, vol. 28, no. 1, pp. 85 – 96, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/B6VD1-4GSTS3G-1/2/93da4a3d76c5b11600e08c324cdd1225

[48] M. Li, J. Graef, L. Wang, and J. Karsai, “Global dynamics of a seir model with varying total population size,” *Mathematical Biosciences*, vol. 160, no. 2, pp. 191 – 213, 1999. [Online]. Available: http://www.sciencedirect.com/science/article/B6VHX-3X3T3JP-4/2/85338a74b4dca96fe0d2ff715975d3d7

[49] L. Loslier, “Geographical information systems GIS) from a health perspective,” in *GIS for health and the environment*. Proceedings of an International Workshop, 1995, pp. 12–20.

[50] C. Macal and M. North, “Tutorial on agent-based modeling and simulation,” in *WSC ’05: Proceedings of the 37th conference on Winter simulation*. Winter Simulation Conference, 2005, pp. 2–15.

[51] P. Martin, “Epidemics: Lessons from the past and current patterns of response,” *Comptes Rendus Geosciences*, vol. 340, no. 9-10, pp. 670 – 678, 2008, ecosystmes et vnements climatiques extrmes - Ecosystems and extreme climatic events. [Online]. Available: http://www.sciencedirect.com/science/article/B6X1D-4RV7GX9-1/2/ed1dc322a4a7cea132d8b9f23359c95f

[52] B. McCue, “Another view of the “small world”,” *Social Networks*, vol. 24, no. 2, pp. 121 – 133, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/B6VD1-44TVD99-3/2/d4ea5e61877fdf9ce5ac43971fa005ad

[53] J. Medlock and A. Galvani, “Optimizing influenza vaccine distribution,” *Science*, vol. 325, no. 5948, pp. 1705–1708, 2009.

[54] R. Merrill and T. Timmreck, *Introduction to Epidemiology, 4th Ed.* Jones and Bartlett Publishers, 2006.

[55] N. Metropolis, “The beginning of the monte carlo method,” *Los Alamos Science*, vol. 15, pp. 122–143, 1987.

[56] N. Metropolis and S. Ulam, “The monte carlo method,” *Journal of the American Statistical Association*, vol. 44, no. 247, pp. 335–341, 1949. [Online]. Available: http://www.jstor.org/stable/2280232

[57] L. Meyers, “Contact network epidemiology: Bond percolation applied to infectious disease prediction and control,” *Bulletin of The American Mathematical Society*, vol. 44, no. 1, pp. 63–86, 2007.

[58] A. Mikler, A. Bravo-Salgado, and C. Corley, "Global stochastic contact modeling of infectious diseases," *Proceedings of the 2009 International Conference on Bioinformatics and Computational Biology*, Aug 2009.

[59] A. Mikler, S. Venkatachalam, and K. Abbas, "Modeling infectious diseases using global stochastic cellular automata," *Journal of Biological Systems*, vol. 21, no. 4, pp. 421–439, 2005. [Online]. Available: http://cerl.unt.edu/publications/2005/pdf/GSCA_journal_paper.pdf

[60] M. Newman, "Models of the small world," *Journal of Statistical Physics*, vol. 101, pp. 819–841, June 2000.

[61] F. Nobre and M. Carvalho, "Spacial and temporal analysis of epidemiological data," in *GIS for health and the environment*. Proceedings of an International Workshop, 1995, pp. 21–31.

[62] T. Opsahl, F. Agneessens, and J. Skvoretz, "Node centrality in weighted networks: Generalizing degree and shortest paths," *Social Networks*, vol. 32, no. 3, pp. 245 – 251, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/B6VD1-4YYWRR4-1/2/15339c1c6a6ee5b10eebf01dfd2ff750

[63] T. Opsahl and P. Panzarasa, "Clustering in weighted networks," *Social Networks*, vol. 31, no. 2, pp. 155 – 163, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/B6VD1-4VTVJRG-1/2/482c38f908493aa9b9bd8ba34dd3365b

[64] M. Osterholm, "Preparing for the next pandemic," *N Engl J Med*, vol. 352, no. 18, pp. 1839–42, May 2005. [Online]. Available: http://www.nejm.org

[65] G. Palla, A. Barabsi, and T. Vicsek, "Quantifying social group evolution," *Nature*, vol. 446, pp. 664–667, 2007.

[66] G. Robins, P. Pattison, Y. Kalish, and D. Lusher, "An introduction to exponential random graph (p^*) models for social networks," *ScienceDirect*, vol. 29, pp. 173–191, 2007.

[67] M. Sanchez and S. Blower, "Uncertainty and sensitivity analysis of the basic reproductive rate," *American Journal of Epidemiology*, vol. 145, no. 12, pp. 1127–37, June 1997. [Online]. Available: http://aje.oxfordjournals.org/cgi/reprint/145/12/1127

[68] S. Sanchez and T. Lucas, “Exploring the world of agent-based simulations: simple models, complex analyses: exploring the world of agent-based simulations: simple models, complex analyses,” in *WSC ’02: Proceedings of the 34th conference on Winter simulation*. Winter Simulation Conference, 2002, pp. 116–126.

[69] P. Sarkar, “A brief history of cellular automata,” *ACM Comput. Surv.*, vol. 32, no. 1, pp. 80–107, 2000.

[70] S. Schnettler, “A structured overview of 50 years of small-world research,” *Social Networks*, vol. 31, no. 3, pp. 165 – 178, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/B6VD1-4W09GJ2-1/2/52260d49171cb9f950341d94a6c3465b

[71] D. Smith, F. McKenzie, R. Snow, and S. Hay, “Revisiting the basic reproductive number for malaria and its implications for malaria control,” *PLoS Biology*, vol. 5, no. 3, p. e42, March 2007. [Online]. Available: http://biology.plosjournals.org/perlserv/?request=get-document&doi=10.1371/journal.pbio.0050042

[72] R. Solomonoff and A. Rapoport, “Connectivity of random nets,” *Bulletin of Mathematical Biophysics*, vol. 13, pp. 107–117, 1951.

[73] P. Tan and F. Pio, “Predicting protein complex core components through data integration,” *Proceedings of the 2009 International Conference on Bioinformatics and Computational Biology*, vol. I, pp. 17–23, July 2009.

[74] R. Toivonen, J. Kumpula, J. Saramäki, J. Onnela, J. Kertész, and K. Kaski, “The role of edge weights in social networks: modelling structure and dynamics,” *Noise and Stochastics in Complex Systems and Finance*, vol. 6601, no. 1, pp. B1–B8, 2007.

[75] G. Treu, A. Kupper, O. Neukumand, and C. Linnhoff-Popien, “Efficient clique detection among mobile targets,” *Mobility ’08: Proceedings of the International Conference on Mobile Technology, Applications, and Systems*.

[76] A. Vazquez, A. Flammini, A. Maritan, and A. Vespignani, “Modeling of protein interaction networks,” *ComPlexUs*, vol. 1, pp. 38–44, Oct 2003.

[77] E. Vynnycky, A. Trindall, and P. Mangtani, “Estimates of the reproduction numbers of spanish influenza using morbidity data,” *International Journal of Epidemiology*, vol. 36, no. 4, pp. 881–889, March 2007. [Online]. Available: http://ije.oxfordjournals.org/cgi/content/full/36/4/881?ck=nck

[78] J. Wallinga and P. Teunis, “Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures,” *American Journal of Epidemiology*, vol. 160, no. 6, pp. 509–516, Sept 2004.

[79] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.

[80] D. Watts and S. Strogatz, “Collective dynamics of ’small-world’ networks,” *Nature*, vol. 393, pp. 440–442, June 1998.

[81] H. Wearing, P. Rohani, and M. Keeling, “Appropriate models for the management of infectious diseases,” *PLoS Medicine*, vol. 2, no. 7, p. e174, July 2005. [Online]. Available: http://dx.doi.org/10.1371%2Fjournal.pmed.0020174

[82] D. West, *Introduction to Graph Theory, 2nd Ed*. Prentice Hall, 2001.

[83] L. White and M. Pagano, “Transmissibility of the influenza virus in the 1918 pandemic,” *PLoS One*, vol. 3, no. 1, p. e1498, Jan 2008.

[84] E. Zegura, K. Calvert, and M. Donahoo, “A quantitative comparison of graph-based models for internet topology,” *IEEE/ACM Trans. Netw.*, vol. 5, no. 6, pp. 770–783, 1997.