

**RESPIRATORY INFECTION IN THE HOSPITAL SETTING: MODELING
TRANSMISSION PATTERNS AND INTERVENTION STRATEGIES USING
MULTITYPE CONTACT NETWORKS**

by

Jillian Gauld

BScH, Queen's University, 2012

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

in

The Faculty of Graduate and Postdoctoral Studies

(Population and Public Health)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

August 2014

© Jillian Gauld, 2014

Abstract

Respiratory infections in the hospital setting are a burden to patients, healthcare workers and the surrounding community. The patterns of disease spread in hospital settings are difficult to capture, due to the heterogeneous rates of contact among healthcare workers (HCWs). Patterns of healthcare worker contact were analyzed for a large Canadian hospital. A novel tool was developed to capture the heterogeneous patterns in contact in hospital settings, and data from the Canadian hospital was used to inform a realistic contact network. The spread of respiratory infections with reproductive numbers approximating SARS, influenza, and measles were simulated on the network, and control strategies including vaccination, transmission reduction, and social distancing were tested. The efficacies of measures were compared between subsets of the population, which were divided by occupation and contact rate. The results of this study found that nurses are in the most contact with other healthcare workers, but the least mobile. Additionally, “other” healthcare workers such as respiratory therapists are at a high likelihood for superspreading events in the hospital setting. This study also identified locations that would reach a large subset of the hospital, which could be used for hand-washing stations or other interventions. Simulations suggest targeting the most highly connected HCW occupations for vaccination, transmission reduction measures, and social distancing may lead to more effective disease containment in outbreak scenarios, and a reduction in resources needed. The results of this study can be used to inform policy decisions, and direct future research towards targeted control strategies in hospital settings.

Preface

This thesis was an expansion of the ongoing work involving hospital contact patterns through the *CONNECT* study: a multi-institutional, CIHR-funded endeavor lead by Dr. Babak Pourbohloul. Survey data collection, entry, and cleaning were previously completed by the research team. All survey data analysis and exploration included in this thesis was completed myself. Unless otherwise specified, all figures were created myself using R or Gephi.

Multi-type networks for the purpose of hospital modeling were previously researched and developed by researchers within the Division of Mathematical Modeling, University of British Columbia Centre for Disease Control, lead by Dr. Babak Pourbohloul. I developed all MATLAB code independently, with input and editing from Dr. Babak Pourbohloul. The random connection algorithm was expanded from Matlab Toolbox for Network Analysis (MIT, 2011), adding the unique components of the multi-type network, multiple degree distributions and association components, independently and originally.

Table of Contents

Abstract.....	ii
Preface.....	iii
Table of Contents	iv
List of Tables	vi
List of Figures.....	vii
Acknowledgements	ix
Dedication	ix
1. Introduction.....	1
1.1 Respiratory Infections	1
1.2 Healthcare Associated Infections.....	1
1.3 Factors That Make a Disease More Likely to Spread	2
1.4 Social Network Heterogeneity.....	4
1.5 Social Networks in the Hospital.....	5
1.6 Research Statement.....	7
1.7 Hypotheses	7
1.8 Research Goals	8
2. Network Modeling Introduction.....	9
2.1 Introduction to Mathematical Models	9
2.2 Terminology	9
2.2.1 Node.....	10
2.2.2 Edge	10
2.2.3 Degree.....	10
2.3 Network Generation.....	12
2.4 Disease Spread on the Network	13
2.5 Hospital Models to Date.....	14
3. Development of a Stylized Network	18
3.1 Introduction	18
3.2 Methods.....	19
3.2.1 Terminology.....	19
3.2.2 Network Attributes	20
3.2.3 Network Generation.....	21
3.3 Results	25
3.4 Discussion.....	30
4. Simulations on the Stylized Network	31
4.1 Introduction	31
4.2 Methods.....	31
4.2.1 Model Characteristics	31
4.2.2 Network Parameters.....	32
4.2.3 Simulation	33
4.3 Results	35
4.3.1 Transmissibility Simulations	35
4.3.2 Network Size Simulations	38
4.3.3 Initial Infection Simulations	45

4.4 Conclusions	48
4.4.1 Transmissibility Simulations	48
4.4.2 Network Size Simulations	49
4.4.3 Initial Infection Simulations	49
5. Contact Patterns and Network Development for a Large Canadian Hospital	50
5.1 Introduction	50
5.2 Methods	51
5.2.1. Survey Analysis	51
5.2.2 Network Development	52
5.3 Results	53
5.3.1 <i>CONNECT</i> Analysis	53
5.3.2 Network Parameterization	58
5.3.3 Network Generation	61
5.4 Discussion	64
6. Modeling Control Strategies in a Hospital Setting	66
6.1 Introduction	66
6.2 Methods	67
6.3 Results	70
6.3.1 Baseline Simulations	70
6.3.2 Transmission Interventions	75
6.3.3 Vaccination Strategies	76
6.3.4 Social Distancing Interventions	79
6.4 Discussion	83
7. Conclusion	85
7.1 A Tool for Network Modeling	85
7.2 Revealing Patterns of Contact and Movement in a Large Canadian Hospital	86
7.3 Evaluating Control Strategies	88
7.4 Strengths and Limitations	89
7.5 Applications of Findings	90
7.6 Future Work	91
References	92
Appendices	99
Appendix A: Network Generation Code	99
Appendix B: Simulation Code	102

List of Tables

Table 3.1 Association parameters	20
Table 3.2 Input parameters for degree distributions	25
Table 4.1 Critical transmissibility and corresponding transmission probabilities for scenarios 1 and 2	35
Table 4.2 Low, medium, and high transmissibility parameters used for simulations	38
Table 4.3 Comparison of mean outbreak size and attack rate between Scenarios 1 and 2 for network size 500.....	43
Table 4.4 Comparison of mean outbreak size and attack rate between Scenarios 1 and 2 for network size 1000.....	43
Table 4.5 Comparison of mean outbreak size and attack rate between Scenarios 1 and 2 for network size 2000.....	43
Table 4.6 Tukey pairwise comparisons for all combinations. 95% Confidence interval ..	46
Table 4.7 Tukey pairwise comparisons for all combinations at medium transmissibility. 95% Confidence interval.....	47
Table 4.8. Tukey pairwise comparisons for all combinations at high transmissibility, 95% Confidence interval.....	48
Table 5.1 Tukey pairwise comparisons for number of locations visited weekly, by HCW type.....	57
Table 5.2 Tukey pairwise comparisons for each occupation comparison	58
Table 5.3 AIC values for the statistical distribution fit	59
Table 5.4 Edge counts for each HCW combination	60
Table 5.5 Association parameters for the hospital network.....	60
Table 6.1 Transmissibility values and corresponding B for each level of efficacy	68
Table 6.2 Corresponding R_0 , transmissibility and B values for SARS-, influenza-, and measles- like illness	73
Table 6.3 Mean outbreak sizes and probability of an epidemic for the three disease scenarios tested.	75
Table 6.4 Probability of epidemic (shaded grey), or mean outbreak size (if probability of epidemic is zero). Grey shading indicates unsuccessful intervention	76
Table 6.5 Probability of an epidemic (shaded) and mean outbreak size, where probability of epidemic is zero, for preexisting immunity at three different R_0 values. Random vaccination strategy	78
Table 6.6 Probability of an epidemic (shaded) and mean outbreak size, where probability of epidemic is zero, for three different R_0 values. Percentages represent vaccination of the top percent of connected individuals in the population	79
Table 6.7 Probability of an epidemic (shaded) and mean outbreak size, where probability of epidemic is zero, for a range of social distancing intervention levels.....	80
Table 6.8 Probability of an epidemic (shaded) and mean outbreak size, where probability of epidemic is zero, for a range of cohorting levels.....	80
Table 6.9 Probability of an epidemic (shaded) and mean outbreak size, where probability of epidemic is zero, for a combination of distancing and cohorting measures.....	82

List of Figures

Figure 1.1 Degree distributions of A. Poisson or normal distributed network and B. Power law distributed network. Nodes in this visualization are referred to as firms.	11
Figure 1.2 Generated networks with degree of A. Poisson distribution and B. Power law distribution	11
Figure 3.2 Depiction of an adjacency matrix and corresponding network	22
Figure 3.3 Probability density plot of the degree distributions generated for Nurses, Physicians, and other HCWS for Scenarios 1 and 2	26
Figure 3.4. Density plot of error for Scenarios 1 and 2	27
Figure 3.5 Clustering coefficients for Scenarios 1 and 2	27
Figure 3.6 Visualization of Scenario 1 ensemble network 158. Red, blue, and green nodes represent physicians, other HCWs, and nurses, respectively	28
Figure 3.7 Visualization of scenario 2 ensemble network 675. Red, blue, and green nodes represent physicians, other HCWs, and nurses, respectively	29
Figure 4.1 Mean outbreak size over 1,000 simulations across a range of beta for Scenarios 1 and 2. Arrow indicates approximate critical value	36
Figure 4.2 Density curve of attack rates at beta values above the threshold: ranging from 0.2 to 0.1	37
Figure 4.3 Density curve of attack rates at beta values ranging from below the threshold value to just above: 0.004 to 0.016	38
Figure 4.4 Simulation results for the network size of 500, using low (A), medium (B), and high (C) transmissibility values, for scenarios 1 and 2	40
Figure 4.5 Simulation results for the network size of 1,000, using low (A), medium (B), and high (C) transmissibility values, for scenarios 1 and 2	41
Figure 4.6 Simulation results for the network size of 2,000, using low (A), medium (B), and high (C) transmissibility values, for scenarios 1 and 2	42
Figure 4.7 Comparisons of the distribution of attack rates between network sizes for high, medium, and low transmissibility scenarios	44
Figure 4.8 Comparison of the distribution in outbreak sizes for simulations with initial infection starting with a physician, nurse, and other HCW at low transmissibility. White circles indicate mean	45
Figure 4.9 Comparison of the distribution in outbreak sizes for simulations with initial infection starting with a physician, nurse, and other HCW at medium transmissibility ...	46
Figure 4.10 Comparison of the distribution in outbreak sizes for simulations with initial infection starting with a physician, nurse, and other HCW. High transmissibility	47
Figure 5.1 Locations in the hospital plotted by total visits per week, and mean hours spent for each location. Colour represents location type	54
Figure 5.2 Locations in the hospital plotted by total visits per week, and mean hours spent for each location. Coloured and plotted by location type	55
Figure 5.3 Distribution of the number of locations visited per week by HCW type	56
Figure 5.4 Distribution of reported HCW contact number between HCWs, nurses, and physicians, in a hospital	57
Figure 5.5 Histogram of the distribution of contact number, by HCW type	58

Figure 5.6 Histogram of physician degree distribution with fitted lognormal curve.....	59
Figure 5.7 Histogram of nurse degree distribution with fitted gamma curve	59
Figure 5.8 Histogram of other HCW degree distribution with fitted gamma curve	59
Figure 5.9 Contact network representing a large Canadian hospital, 2,805 nodes. Red, blue and green nodes represent physicians, other HCW and nurses, respectively	62
Figure 5.10 Density plot of the error rate over 1,000 iterations of network generation	63
Figure 5.11 Clustering coefficients over 1,000 iterations of the network generation process.....	65
Figure 6.1 Density curves of outbreak sizes for simulations with transmissibility values ranging from 0 to 0.059	71
Figure 6.2 Density curves of outbreak sizes for simulations with transmissibility values ranging from 0.141 to 0.469	71
Figure 6.3 Mean outbreak sizes over a range of transmissibility values	72
Figure 6.4 Histogram of outbreak sizes over 1,000 simulations for the low R_0 illness.....	74
Figure 6.5 Histogram of outbreak sizes over 1,000 simulations for the medium R_0 illness	74
Figure 6.6 Histogram of outbreak sizes over 1,000 simulations for the high R_0 illness....	74

Acknowledgements

I would like to thank first and foremost my supervisor, Dr. Babak Pourbohloul, for his support and guidance throughout this process. Your knowledge and insights in both public health and mathematics have allowed me to produce work that I am incredibly proud of. I appreciate you not only taking the time to take me on as a student, but also challenging me throughout the process. Thank you to my committee members, Dr. Bonnie Henry and Dr. Raymond Tellier for your stimulating discussions and valuable insights as public health professionals. An additional thank-you to the Canadian Institutes of Health Research, for providing funding for my research.

Throughout this process, I have received support and inspiration from many other individuals. Thank you to Krista English for your professional guidance and invaluable advice throughout the past two years. Thanks to Dr. David Patrick for giving our Epidemiological Methods class, at the beginning of graduate school, the most inspiring presentation. It confirmed my belief that public health is where I belong. Thank you to my incredibly supportive family: Ian, Georgina, and Gavin. I am so appreciative of you all for trusting me, believing in me, and pushing me to pursue my goals. Thank you to my smart, talented, and beautiful friends for your listening abilities, and daily inspiration to challenge myself.

1. Introduction

1.1 Respiratory Infections

Respiratory transmitted infections have the potential to become widespread, pandemic infections, and are often difficult to fully characterize and contain. Because physical contact with another infected individual or fomite isn't necessary for transmission, traditional hand-washing and sanitization methods are not sufficient for effective disease containment. With these diseases, droplet or airborne particles are transmitted through coughing, sneezing, or simply talking. Thus, casual face-to-face interaction can propagate these respiratory transmitted diseases.

Many of the significant respiratory transmitted infections of the past two decades are emergent, or re-emergent diseases, where the epidemiological characteristics are not yet discovered or commonly known. The emergence of SARS in 2003, influenza A(H1N1)pdm09 in 2009, and the resurgence of measles in many countries in 2013 have all posed challenges to infection control. Emergent infectious diseases have shown a significant increase in incidence since 1940; thus there is a potential for future disease emergence and challenges to infection control in years to come (Jones et al., 2008).

1.2 Healthcare Associated Infections

Healthcare associated infections, or nosocomial infections, are a significant burden to public health. In Canada, more than 200,000 patients acquire a nosocomial infection annually, and more than 8,000 die as a result (Zoutman et al., 2003).

Nosocomial infections can impact the burden of disease both within the hospital and the surrounding community. Chains of healthcare associated infections can serve as a propagator of disease spread, due to close contact of patients and hospital personnel, potentially vulnerable immune systems of patients, and movement of patients from hospitals to throughout the community.

Respiratory infections including SARS, MERS-CoV, and influenza have been propagated through hospital settings (Chironna et al., 2010, Salgado 2004, Assiri et al 2013). Some of the factors that likely contributed to this are close contact between patients and healthcare workers (HCWs), decreased immune function of many hospitalized patients, and aerosol/respiratory droplet producing procedures such as intubation.

Healthcare workers are disproportionately affected in many respiratory disease outbreaks. The emergence of SARS revealed the vulnerability of Canadian HCWs in outbreak scenarios. Forty-three percent of Canadian SARS cases occurred in healthcare workers (World Health Organization, 2003), leading to long-term stress and trauma among HCWs in SARS-affected hospitals. The WHO has reported a similar burden of disease for MERS-CoV, with approximately 25% healthcare worker cases between April and June 2014 (World Health Organization, 2014).

1.3 Factors That Make a Disease More Likely to Spread

The primary measure in epidemiology used to predict the likelihood of an outbreak is the basic reproductive number, or R_0 , which is the average number of secondary cases arising from an average primary case in an entirely susceptible population, assuming homogeneous population mixing (Anderson & May, 1991). This value is used to inform vaccination policy, infection prevention, and control measures for a range of diseases and scenarios.

R_0 can be estimated through seroprevalence data, epidemic growth rates, and outbreak size (Dietz, 1993). In an entirely susceptible population, diseases with values of R_0 below one are predicted to die out over time, while diseases with an R_0 above one have epidemic potential. Diseases such as seasonal influenza have an R_0 estimated at approximately 1.3 (range 0.9 to 2.1) (Chowell, Miller, & Viboud, 2007), while highly contagious diseases such as measles are estimated between 6 and 7 (Mossong & Muller, 2000) and as high as 15 (Anderson & May, 1991).

Although the value of R_0 remains informative in epidemiology, there is often a large heterogeneity in the magnitude of disease spread from population to population, despite identical R_0 (Meyers, Pourbohloul, Newman, Skowronski, & Brunham, 2005). This may be due to many factors, including differing immunity or infection control measures, as well as the structure of the population's social network.

First, the susceptibility of the population plays a role in the success or failure of the epidemic. R_e , or the net effective reproductive number, is utilized when predicting epidemics in a population which is not entirely susceptible. It is calculated by the product of R_0 and the percent of susceptible individuals in the population. Thus, as the percent of susceptibles in a population decreases, due to vaccination or natural immunity, the effective R_e decreases.

Infection control measures can also impact the heterogeneity in disease spread between populations. Interventions that decrease either duration of illness, probability of transmission, or contact rate, will reduce R_0 . Thus, the use of interventions such as early detection, and treatment of the disease, mask-wearing, and population level distancing measures such as school closures or quarantine will decrease R_0 in a population.

Finally, an individual's position in the social network can impact the magnitude of disease spread in a population. Traditional applications of R_0 assume that all individuals in a population have an equal probability of exposure, thus ignoring the potential differences in social network, or network of contacts between individuals, which can greatly impact the fate of a disease.

The outbreaks of SARS in Canada exhibited the impact of the social network on outbreak heterogeneity (Meyers et al. 2005). Through the introduction of a single case in both cities, Vancouver experienced four probable cases of SARS, while Toronto experienced approximately 375 (Campbell, 2006).

The reasons for this vast heterogeneity is likely influenced by the social networks of the index patients in each city. The index patient in Vancouver acquired SARS while traveling abroad, and upon return to Vancouver, returned to his home where only his wife resided. Limited social contact and almost immediate hospitalization prevented the spread of SARS in Vancouver much past this index patient (Skowronski et al., 2006).

In contrast, patient zero in Toronto was the matriarch of a large family, and died at home. Her son's subsequent infection was the first of five additional familial infections. Due to the index patient's large family social network, and slow identification of the disease, SARS spread to other HCWs and patients, leading to approximately 375 probable infections. (Poutanen, 2003; Campbell, 2006).

1.4 Social Network Heterogeneity

Utilizing R_0 while addressing the heterogeneity in social networks is essential for effectively predicting disease spread, particularly in populations where these contact rates vary greatly between individuals. Networks of social interactions can be observed in many subsets of the population, and measure many different types of connections including sexual contact, face-to-face interaction, and physical contact.

The study of sexually transmitted diseases first revealed the importance of social networks in epidemiology. Anderson and May's derivation of R_0 for the spread of sexually transmitted diseases in a population utilizes the average number of sexual contacts in a population (Anderson & May, 1991). The number of sexual contacts is known to vary significantly across a population, adopting a scale-free distribution where the majority of the population has few sexual contacts, and a small number of highly connected individuals preferentially attach to each-other (Liljeros, Edling, Amaral, Stanley, & Aberg, 2001). Therefore, although useful for planning control strategies and evaluating R_0 within homogenous sexual populations, this *average* fails to capture the diversity within sexual networks. Addressing the sexual network along with R_0 has led to

more efficient and targeted STI control strategies, by concentrating resources to the most highly connected subsets of the population.

1.5 Social Networks in the Hospital

Although mixing patterns and contact rates of the general population have been quantified on a large scale (Mosson et al., 2008), rates of healthcare worker contact (HCWs) are known to be significantly higher and more heterogeneous than the general population (Bernard, Fischer, Mikolajczyk, Kretzschmar, & Wildner, 2009). Studies that have attempted to quantify these rates have utilized methods ranging from observation to wearable sensors.

Shadowing of hospital employees has been utilized to capture contact network dynamics, but only in 30-minute time increments for a total of 40 hours per job category (Polgreen, Tassier, Pemmaraju, & Segre, 2010). This leaves little room for evaluation of contact heterogeneity, but revealed that intensive care nurses, residents and fellows had the highest number of HCW contacts of the job categories observed.

Utilizing radio-frequency identification (RFID) tags within a pediatric ward (Isella, Barrat, Cattuto, & Colizza, 2011), a large heterogeneity of contact frequency was found among HCWs, even within occupation type. RFID tags are useful along with other sensor technologies to track movements and interactions between healthcare workers in an automated way. Ward assistants were found to have the highest number of daily contacts, with physicians having the least. Patterns of inter-HCW interaction were analyzed, and intra-occupational (assortative) mixing was the most common: The most frequent interactions were ward assistant-ward assistant and nurse-nurse (Isella et al., 2011).

The presence of highly mobile and connected HCWs was confirmed through the study of HCW contact in a medical intensive care unit (MICU), using sensor technology (Hornbeck et al., 2012). This study was the first to confirm the heavy tailed distribution of HCW contacts, where a small number of HCWs were responsible for a large

proportion of contacts. These HCWs were determined to be peripatetic, or having a significantly large and diverse set of contacts.

Similar patterns were found using RFID tags in a geriatric ward. Only 6 HCWs were responsible for 42% of all patient contacts, suggesting a similar heavy-tailed distribution in this setting (Vanheems et al., 2013). Additionally, inter-HCW interaction was studied, again suggesting assortative mixing between HCW occupations: Nurse-Nurse interactions were the most frequent HCW contact type, followed by medical doctor-medical doctor interactions.

The methods of direct observation and wearable sensors are accurate in capturing all forms of HCW contact within the hospital: both social and work-related. Currently, these data only exists within units or wards, since sensor technology and observation isn't feasible or cost-effective to implement on a large scale. Measures of HCW contact are needed hospital-wide.

Large-scale studies have utilized electronic medical records to approximate HCW contact throughout the hospital, which provides a larger and more readily available source of data. A combination of Electronic Medical Records (EMR) and spatial hospital data was utilized to examine contact between HCWs. A heavy-tailed distribution of contacts was found, with a high level of contact heterogeneity among HCWs (Curtis, Hlady, Kanade, Pemmaraju, & Polgreen, 2013). Resident Physicians and nurses were the most frequently in the top 10% of contacts, contributing most to the heavy-tailed distribution. Although useful for quantifying on a large scale, EMR data fail to capture casual and social contacts in the hospital setting.

More research is needed at a hospital-wide level to explore these patterns of contact. The ability to inform hospital-wide policy is currently limited, as the majority of studies focus within a single ward or unit in the hospital. Despite these limitations, it is apparent that the social network of a hospital is highly heterogeneous, and that utilizing a network

approach for infection control would best capture the spread of infectious disease in this environment.

1.6 Research Statement

Healthcare associated infections remain a significant burden to public health, affecting over 200,000 Canadians yearly. Routine practices of hospital infection control are currently centered on hospital-wide precautions to reduce transmission (Public Health Agency of Canada, 2013)

Social contact heterogeneities among HCWs are not addressed in current control strategies, yet patterns of social contact may significantly affect the spread of respiratory infections within the hospital setting. Research is needed to explore the effect of targeted control strategies on disease transmission in hospital settings.

The purpose of this thesis is to develop a realistic contact network of a hospital, in order to create targeted and optimized control strategies in the hospital setting.

1.7 Hypotheses

1. Patterns of contact in the Canadian hospital setting will reflect the heavy-tailed distributions of contacts found in other healthcare settings.
2. Simulations using contact networks show a large heterogeneity in disease trajectory, reflecting the patterns seen in previous outbreaks of SARS, for example.
2. Control strategies simulated which target highly connected individuals will be more efficient than random strategies in the scenarios tested.

1.8 Research Goals

This thesis has four distinct research chapters, prefaced by an introduction to network modeling in public health.

The research goals of this thesis are to:

1. Create and validate a flexible tool for network generation in hospital settings
2. Analyze patterns of contact in a large Canadian hospital to create a realistic hospital network
3. Utilize the network to evaluate the efficacy of control strategies

2. Network Modeling Introduction

2.1 Introduction to Mathematical Models

Mathematical disease modeling has become increasingly relevant in the area of pandemic preparedness, combining the fields of mathematics, statistics, and public health in order to predict the impact of outbreaks, and the efficacy of various intervention strategies. The mathematical tools being developed have the potential to be applied to emerging infectious diseases within small communities, across Canada, and worldwide.

Network modeling is a subset of mathematical disease modeling, which simulates disease spread on networks. The benefit of network modeling is that it is able to capture non-homogenous rates of contact within populations and communities, a feature that isn't present in traditional compartmental modeling. Originally utilized in epidemiology to simulate the unique and heterogeneous dynamics of STI spread, network modeling is currently utilized to capture the spread of a variety of diseases, from bacterial infections to respiratory illnesses.

Network models can be utilized to study the patterns of transmission of a disease, as well as optimize control strategies to a population. Because pandemics are relatively rare occurrences, we are limited in our evidence supporting or refuting infection control strategies. Modeling allows for the simulations of thousands of possible scenarios, providing robust evidence to aid in policy decision-making.

2.2 Terminology

A few common terms are used in network modeling, and will be utilized throughout this thesis.

2.2.1 Node

A *node* is the representation of a person, object, or place, in the network. In the context of network epidemiology, a node in a network typically represents an individual. A network typically contains multiple nodes, representing the population being studied.

Nodes can be assigned parameters that are unique to the individual or a subpopulation within the network. In disease modeling, these can include transmissibility parameters and recovery rates. Additionally, nodes can be removed from the network, representing vaccination.

2.2.2 Edge

Nodes are connected by *edges*. These edges can represent many types of interactions, from face-to-face meetings to sexual contact, depending on the transmission of the disease to be simulated.

2.2.3 Degree

Each node has a number of edges attached to it, known as its *degree*. The distribution of degrees across the population is known as the *degree distribution*.

Degree distributions take many shapes, depending on the population and type of edge studied. Degree distributions are often approximated by probability distributions. Some of the most common include Poisson, lognormal, gamma, and power-law distributions.

Power-law and Poisson distributions are represented in Figure 1.1, with associated networks in Figure 1.2 Poisson degree distributions are indicated by a normal distribution of contacts among the population, with the majority of individuals having the average contact number in the population. Scale-free networks indicate a high number of individuals with low contact numbers, with a small number of individuals with a very high degree. These high-degree individuals in networks are often called “super-spreaders,” in the context of epidemiology.

Figure 1.1 Degree distributions of A. Poisson or normal distributed network and B. Power law distributed network. Nodes in this visualization are referred to as firms.

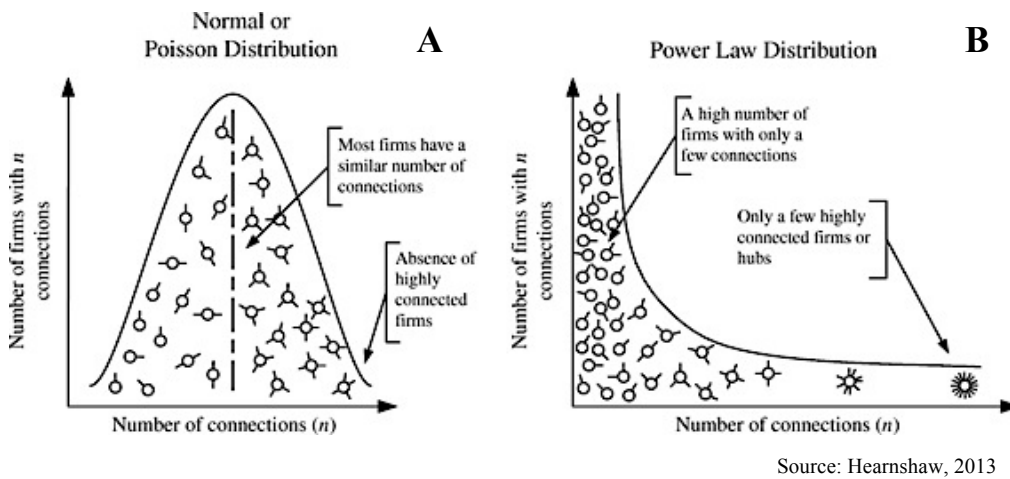
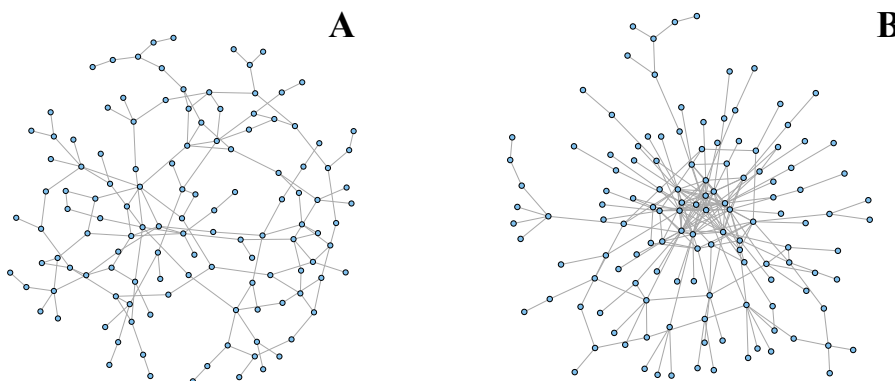


Figure 1.2 Generated networks with degree of A. Poisson distribution and B. Power law distribution.



2.3 Network Generation

Networks in epidemiological modeling can be constructed using theoretical assumptions, real data, or a combination of the two. Measures of degree, as well as association are needed.

Networks that are based primarily on theoretical assumptions of contact are called *stylized* networks. Although less applicable to real-world problems, these networks allow for a controlled environment to validate models and test theoretical assumptions.

Few studies have generated complete data of a contact network. Because we need both the measures of *degree* as well as who comes into contact with whom on an individual level, these data are difficult to obtain. Sexual health surveys and RFID technology have been the most successful at obtaining real network data (Jolly, Muth, Wylie, & Potterat, 2001; Isella et al., 2011). Although informative for small communities, these data would be difficult to obtain on a large scale.

Most studies utilize a combination of data source and theory to approximate networks. Survey, census, and direct observation data have been useful for informing network parameters. Based on the questions asked (i.e. How many individuals do you come into contact with?), these data can often be used to approximate a degree distribution. Without complete network data, we have no information on who comes into contact with whom. Thus, nodes are either connected randomly or by probabilities of association, which can be obtained from the dataset or other sources (i.e. *Who* do you come into contact with?).

Within a network, there may exist distinct types of nodes that differ in degree distribution and measures of association. In public health, these types of nodes could represent sexual orientation, occupation, or age class, for example. Approximated networks that contain multiple types of nodes are known as *multitype networks* (Newman, 2003). Degree distributions and association parameters are approximated from data independently for each type of node, and then utilized to create a network that includes all types. Multitype

networks allow for the specification of distinct degree distributions and mixing patterns for each node type, creating a network more closely aligned with a network derived directly from data.

2.4 Disease Spread on the Network

Disease spread on a network is simulated through *bond percolation*, where a disease will move through the network by edges, or bonds (Meyers, 2007). Unlike traditional compartmental models where there is equal probability of exposure throughout the population, exposure through network models require a susceptible individual to be connected to an infected individual by an *edge*. If connected, an individual can become infected by transmission probability, T . Thus, the outcome of the disease is dependent on both the transmissibility of the disease, T , and the structure of the network (Newman, 2002).

The transmission probability, T , is derived from the R_0 of the disease and network density (Meyers, 2007). When simulating intervention strategies that reduce disease transmission, such as wearing masks or hand washing, T is reduced by a corresponding percentage. This reduction in transmission can be implemented for all connections in the network (simulating a population-wide intervention), or for specific connections between nodes (simulating an intervention for high risk individuals, for example).

The structure of a network is critical to the dynamics of disease spread upon it (Keeling, 2005). These structural properties may include distribution of contacts, or the patterns in which the contacts are organized, known as *clustering*. The presence of super-spreaders in power-law distributions, for example, confers a lack of epidemic threshold; meaning transmission at even an extremely low rate could cause an epidemic (Pastor-Satorras & Vespignani, 2001).

Models simulating social distancing and vaccination confer changes to the network, and impact the outcome of disease spread. Social distancing is the reduction of social interactions in a population, which can be achieved by a variety of measures depending on the disease and population, but may include: school closures, travel restrictions, abstinence and staffing changes. Social distancing has been proposed as a control strategy to mitigate the impact of pandemic influenza (Roth & Henry, 2011). It is simulated by removing edges between nodes, either randomly or according to node type. This leads to a decrease in network density, which results in reduced disease spread on the network.

Vaccination can also be simulated on the network, a measure that prevents an individual from contracting and transmitting a disease. Thus, simulating vaccination effectively removes nodes entirely from the network, because their edges to other nodes cannot contribute to disease spread. Depending on the nodes that are vaccinated, density may increase or decrease. For example, removing a highly connected node would decrease density, while removing a node with a single connection would increase overall network density.

2.5 Hospital Models to Date

Network modeling for hospital settings have been recently developed for a range of diseases and uses. A theoretical approach (Meyers, Newman, Martin, & Schrag, 2003) first explored the application of network theory on hospital networks. The study lacked real data to parameterize the network, instead using assumptions of caregiver, ward and patient structure. The model didn't propose distinction among hospital personnel. Simulating the spread of *M. pneumoniae* demonstrated the importance of caregiver contact patterns in the control of these outbreaks.

Medical records were first explored as a data source to inform contact networks in hospital settings (Liljeros, Giesecke, & Holme, 2007). Inpatients were connected if they were on the same ward within a specified time period, varied depending on the disease studied. Although this model excluded HCW interactions, it provided valuable insight

into the non-homogenous mixing patterns of inpatients, and provided insight into effective control measures for infections such as MRSA.

Building on the use of medical records for network development, a combination of EMR data and assumptions of HCW activity were utilized to approximate the social network in a hospital setting (Ueno & Masuda, 2008). Patient records documenting assigned caregivers were utilized: Patients and their assigned physicians and nurses were connected, along with nurses of the same ward. This study added a new level of detail to the hospital network. Because the assumptions of contact are primarily determined from patient care, this study ignores patterns of HCW interaction not recorded in a patient chart, such as social interaction and meetings. With this model, it was found that medical doctors were the key spreaders of disease on the network, and strategies to reduce the travel of doctors throughout the hospital would impact epidemic size more than direct isolation of patients.

In order to capture both work related and social contacts within the hospital, direct observation was used to create a contact network (Polgreen et al., 2010). Each occupation was observed for approximately 40 hours, and the HCW's contacts were recorded, including their occupation and duration of contact. Contact was defined as coming within 0.9 meters of another HCW, approximating a distance sufficient for droplet transmission. A network was created by randomly connecting HCW nodes, using the probabilities of association between HCW types derived from the observational data. Although a much more detailed network than those derived from EMR data, the method of network generation failed to capture the heterogeneity within HCW classes, only between. The simulations on the network involved testing the efficacy of vaccination strategies that preferentially target the most highly connected HCWs, a strategy that, although difficult to implement, may lead to a higher efficacy with a lower vaccination rate.

Wearable sensor (RFID) technology has been used to capture real network information in hospital settings, collecting data at an individual resolution on *who* comes into contact with *whom* in a hospital setting. This technology was first utilized in a hospital setting to

measure HCW interactions within a pediatric ward (Isella et al., 2011). Contact in this study was defined as face-to-face interaction of approximately one meter for at least 20 seconds. Simulations were not conducted on the network, although observations from the network structure suggested that nurses play a central role in infection spread and should be prioritized.

Mote based sensoring, a similar technology, was used to create a contact network in the medical intensive care unit (MICU) (Hornbeck et al., 2012), based on physical contact for at least 30 seconds. Hand hygiene compliance was simulated on the network, and the relative effect of the compliance of either highly connected or random individuals was compared. It demonstrated that these highly connected HCWs can have a much greater effect on disease spread if noncompliant, compared to less connected HCWs.

RFID technology was again used to create a contact network of HCW contact in a geriatric unit (Vanhems et al., 2013). Contact in this study was defined as face-to-face interaction of approximately one meter for at least 20 seconds. Although no simulations were performed, the presence of peripatetic highly connected and mobile HCWs were identified.

Electronic medical records were used to create a network of patient contacts across the entire hospital (Cusumano-Towner, Li, Tuo, Krishnan, & Maslove, 2013). Patients were connected either through sharing of a room, or sharing of a HCW within a 2-hour period. Although not as effective in capturing the spread of respiratory pathogens, using EMRs were useful in this context for modeling the spread of contact-transmitted bacterial pathogens throughout the hospital. The model was unique in testing the impact of the disease starting in different parts of the hospital. It revealed parts of the hospital that can most propagate an epidemic, including the emergency department, where individuals can be transferred to many wards in the hospital after admission.

Another hospital-wide network was constructed through EMR data and the hospital layout (Curtis et al., 2013), utilizing a range of distance and time parameters to define

contact. Simulations on the network compared vaccination policies targeting HCWs based on mobility throughout the hospital or *degree*. It was found that vaccinating the most mobile individuals is as effective and more feasible, than vaccinating those with the highest *degree*.

Although sensor technologies are the most accurate in informing hospital contact networks, it is not currently feasible or cost-effective to implement a wearable sensor study on a hospital-wide scale. Thus, EMR data has been utilized to approximate these networks. Healthcare records provide evidence of formal interactions such as physician visits and procedures, but fail to capture face-to-face social interactions and meetings between HCWs. These types of informal face-to-face contacts are often sufficient for the transmission of respiratory infections through both droplets and aerosols, and therefore may be significant in the spread of respiratory-transmitted pathogens in the hospital setting. Methods to create a realistic hospital-wide contact network are needed to model the spread of respiratory pathogens. Capturing the heterogeneity in HCW contact rates, as well as informing accurate simulations, can contribute to more optimized and effective control strategies in the hospital setting.

3. Development of a Stylized Network

3.1 Introduction

Given the complex organizational structures of hospitals, contact patterns of healthcare workers in hospital settings are thought to be highly heterogeneous. Recent studies involving wearable sensors, direct observation, and electronic medical records have confirmed this, suggesting the potential for superspreaders in the hospital setting (Vanhems et al., 2013), and differences in contact patterns between occupations (Isella et al., 2011). Because of this heterogeneous environment, quantifying networks of contacts in a hospital is crucial for effective and optimized infection control.

Networks have been informed through direct observation and wearable sensors within a single ward or unit in the hospital setting (Polgreen et al., 2010; Hornbeck et al., 2012, Isella et al. 2011; Vanhems et al., 2013). Although effective in capturing the individual-level resolution of contact diversity in small settings, these methods are not currently feasible to develop hospital-wide networks. Healthcare records have been utilized to approximate hospital-wide contact networks (Liljeros et al., 2007; Ueno & Masuda, 2008; Cusumano-Towner et al., 2013; Curtis et al., 2013), but fail to capture non-recorded contact, such as social interaction or staff meetings. A new methodology is needed to create a realistic hospital contact network, which is both feasible on a large scale and accurate in capturing all work and social interaction in the hospital setting.

Multitype networks (Newman, 2003) are a type of data-informed network, which allow for the specification of both intra- and inter- group contact patterns and rates. These networks have been theoretically applied in epidemiology (Allard, Noël, Dubé, & Pourbohloul, 2009; Vazquez, 2006), but haven't yet been utilized to capture the complex network dynamics of HCWs in hospitals. Approaching a hospital contact network as a multitype network allows for the specification of a distinct contact and HCW mixing pattern for each occupation within the hospital. Rather than informing the network

directly through sensors, observation, or medical records, we can use a simple set of parameters for each occupation to produce a network of the entire hospital.

The goal of this chapter is to develop a tool that generates multitype contact networks for hospital settings. This framework will account for the heterogeneity in contact patterns and mixing both within occupations and across the hospital, and allow for the flexible application of the model to a variety of hospital settings. For demonstration and proof of concept, two stylized networks will be generated.

3.2 Methods

3.2.1 Terminology

The hospital network is composed of *nodes* and *edges*. Each *node* represents a healthcare worker within the hospital, and is categorized as a physician, nurse, or other healthcare worker (HCW). Currently, patients are excluded from the hospital network, but could be added in future iterations of this network.

An *edge* represents a connection between nodes. The connection, for the purpose of this thesis, is defined as face-to-face contact within 1 meter for at least two minutes, which is taken from the definition of direct HCW contact utilized in the *CONNECT* survey. This level of contact is sufficient to meet the minimum requirements for exposure to respiratory droplets (Bischoff, Swett, Leng, & Peters, 2013).

The *degree* of a node is the number of edges connected to it. In this network, the *degree* represents the number of contacts a healthcare worker has. The *degree distribution* is the pattern or *distribution* of the number of contacts a node has, across a population or subgroup.

3.2.2 Network Attributes

There are two primary attributes of the network generation tool that are utilized as the primary input, and distinguish it from previously developed network generation tools.

3.2.2.1 Occupational Degree Distribution

Each occupation is assigned a unique degree distribution, characterizing the contact patterns of that group. Degree distributions can incorporate the *shape* of the distribution (i.e. Poisson or scale-free), and *magnitude* (mean degree of the population), for each occupation. The ability to incorporate distinct degree distributions for each subpopulation is a primary attribute of the network generation code.

3.2.2.1 Mixing Patterns

The next attribute addressed in the network is the ability to specify mixing patterns between occupations. This determines the proportion of HCWs that will be connected to others within their occupation, and those outside their occupation. Thus, we can specify whether mixing is assortative.

For each occupation of type i (either physician, nurse, or other), there exists N_i , the sum of degrees across individuals of occupation i . The parameter p_{ij} represents the proportion of N_i connected to HCW type j . Thus, there are 9 p_{ij} combinations in a hospital network of 3 occupations: Nurses, physicians, and others.

Table 3.1 Table of association parameters.

	Incoming Connection			
Outgoing Connection		Physicians	Nurses	Other HCW
	Physicians	p_{pp}	p_{pn}	p_{po}
	Nurses	p_{np}	p_{nn}	p_{no}
	Other HCW	p_{op}	p_{on}	p_{oo}

The proportional constraints are not symmetric (p_{ij} does not equal p_{ji}). This is because N_i is different for each occupation. For example, if there is a higher total degree (N_i) for physicians, the probability of a nurse connecting to a physician is higher than a physician connecting to a nurse. Although the proportions aren't symmetric, the number of edges of type i connecting with j must equal the number of edges of type j connecting with type i . Thus the following must be satisfied prior to network generation:

$$N_i * p_{ij} = N_j * p_{ji}.$$

When deriving these parameters directly from a network, the agreement of these parameters is inherent. Because these parameters may not agree when taken from other datasets, a forcing mechanism is built into the stylized code to ensure this constraint.

3.2.3 Network Generation

Network generation code was developed in MATLAB. Code is included in Appendix A.

3.2.3.1 Inputs

A set of input parameters are required to initialize the generation function:

Population size: Numerical value for total population, and size of each occupation

Degree distribution: Including type and magnitude for each occupation.

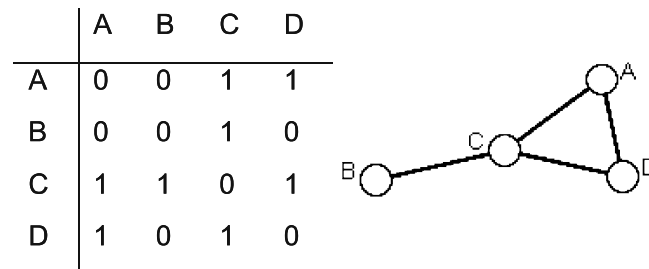
Association parameters: Estimated values of p_{ij}

3.2.3.2 Process

- a) The first step assigns a specific degree to each node of the subgroup according to the input degree distributions. Thus, individuals are given a pre-determined number of 'stubs,' or unconnected degrees.

- b) An empty adjacency matrix is created. Adjacency matrices are a method of recording network structure where row and column indices represent each node in the network, and a connection is marked by a “1” in the corresponding row and column combination of the connected individuals. An example of an adjacency matrix is found in Figure 3.2.

Figure 3.2 Depiction of an adjacency matrix and corresponding network.



Source: Merrill 2008

- c) Assigned values of p_{ij} are multiplied by N_i to calculate the number of connections between each HCW-type. Thus, before the connection process starts, there are a pre-determined number of edges designated for each connection type (physician-physician, physician-nurse, nurse-nurse, etc). A forcing mechanism ensuring $N_i * p_{ij} = N_j * p_{ji}$ is incorporated which currently prioritizes nurse-nurse and physician-physician association parameters, but can be altered depending on the association of interest. The mechanism of forcing is included in the code (Appendix A)
- d) Nodes are connected randomly. First, one of the six edge types is randomly selected (i.e. nurse-nurse, nurse-physician, etc.). Then, individuals are selected randomly within the corresponding HCW type(s). Only individuals with unconnected stubs remaining can be selected. HCW pairs that are already connected can't be connected again, ensuring that self-loops won't occur.

- e) The nodes selected to be connected are recorded in the adjacency matrix, indicated by an entry of “1” in the row/column combinations associated with the HCW. A stub is removed from each individual’s assigned degree.
- f) Steps [d-e] continue until there are no more stubs left to connect within the connection type, or all individuals with stubs remaining have already been connected.

3.2.3.3 Outputs

The final outputs are generated and saved for each process:

Adjacency Matrix: Final adjacency matrix corresponding to the hospital network.

Error: Error rate is calculated as the number of unconnected stubs, divided by assigned stubs, for the entire network

Clustering coefficient: The clustering coefficient is a measure of randomness in the network, and is calculated using a corresponding MATLAB function

3.2.3.4 Ensembles of Networks

As described in the network generation process, the method of connecting each node is inherently random. Thus, even with identical input parameters, any two iterations of the generation algorithm can result in different network structure.

The input parameters are primarily utilized in specifying two main components: *how many* other individuals a HCW comes in contact with, and *what occupations* they are in contact with. For example, a randomly selected nurse node n has the following assignment: 2 physicians, 5 nurses, and 1 HCW. Thus, we know how many contacts of each occupation a HCW has, but we do not assign the specific HCW. This is where the randomness of the

process enters: Within each connection assignment, an individual can become connected with any other HCW of the assigned occupation, given they have unconnected stubs corresponding. Nurse node n , for example, can be connected with any 2 physicians in the network, as long as those physicians have available stubs for a nurse connection.

Therefore, although the process is random, the network will retain its assigned parameters, regardless of network structure.

To examine the potential variation in the network due to the random generation process, steps [b-g] of the code are repeated over 1,000 iterations, and the results stored as an *ensemble* of networks in 3-d matrix form. Step [a] is not repeated to ensure consistent degree distribution input across networks.

3.2.3.5 Scenario Input

Two hospital network scenarios were created as a demonstration and proof of concept. Input parameters were based on the distribution of non-admin hospital personnel across three major Canadian hospitals (Johnston 2012), and a total hospital size of 500. Thus, the networks include 70 physicians, 310 nurses, and 120 other HCWs. The association parameters of p_{ij} were assigned as $p_{nn}=0.4$, $p_{dd}=0.05$, and $p_{on}=0.4$, corresponding to nurse-nurse, doctor-doctor, and other-nurse connections, respectively. The degree distribution parameters are summarized in Table 3.2. Two contrasting sets of degree distributions were chosen, and are referred to as Scenarios 1 and 2. Scenario 1 contains degree distributions of the same shape (Poisson), but differ in parameter value lambda (λ), the average rate of contact. Scenario 2 contains degree distributions that differ in both shape and rate of contact. Physicians were assigned a power law distribution with an exponential cutoff, with power law scaling parameter alpha (α), minimum value x_{min} , and exponential scale value lambda (λ). Nurses were assigned a Poisson distribution with average

rate λ). Other HCWs were assigned a lognormal distribution, with mean parameter μ , and standard deviation value σ .

Table 3.2 Input parameters for degree distributions.

	Scenario 1	Scenario 2
Physicians	Poisson ($\lambda=25$)	Power law ($\lambda=0.027$, $\alpha=0.4$, $x_{\min}=1$)
Nurses	Poisson ($\lambda=8$)	Poisson ($\lambda=8$)
Other HCW	Poisson ($\lambda=13$)	Lognormal ($\mu=2.4$, $\sigma=0.5$)

3.3 Results

The network generation code successfully created two ensembles of 1,000 networks corresponding to Scenarios 1 and 2, reflecting the assigned degree and association parameters. Density plots of the generated degree distributions used in all iterations of the generation process are visualized in Figure 3.3.

Error over the 1,000 iterations of the network can be visualized in Figure 3.4. Scenario 1 had a significantly lower mean error rate of 0.002 compared to 0.021 for Scenario 2 ($t=-212.58$, $p<0.001$), though mean error rates for both scenarios were below 5%.

Average clustering coefficients across the network were calculated, and can be visualized in Figure 3.5. Overall, these values were very low. A mean of 0.03 corresponds to Scenario 1, while a mean of 0.04 was calculated for Scenario 2.

A random selection of one network from each scenario was visualized in Gephi (Figures 3.6 and 3.7). These are ensembles 158 of 1,000 from Scenario 1 and 675 of 1,000 from Scenario 2. The size of the node corresponds to its *degree*, while the colour corresponds to occupation.

The visualization of Scenario 1 ensemble network 158 contains 500 nodes and 2833 edges, with a mean degree of 11.33. Within occupations, the mean degree of physicians, nurses, and others corresponds to 24.96, 7.85, and 12.38, respectively.

The visualization of Scenario 2 ensemble network 675 contains 500 nodes and 2747 edges with a mean degree of 10.99. Within occupations, the mean degree of physicians, nurses, and others corresponds to 24.27, 7.69, and 11.75, respectively.

Figure 3.3 Probability density plot of the degree distributions generated for Nurses, Physicians, and other HCWS for Scenarios 1 and 2.

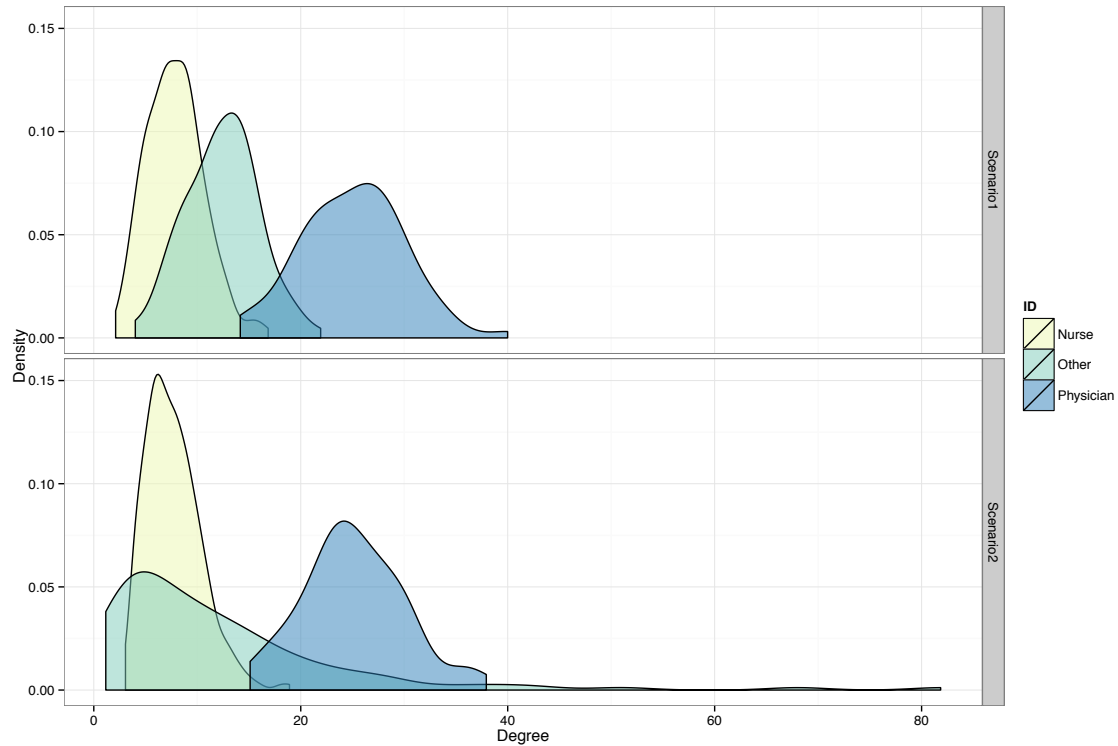


Figure 3.4. Density plot of error for Scenarios 1 and 2.

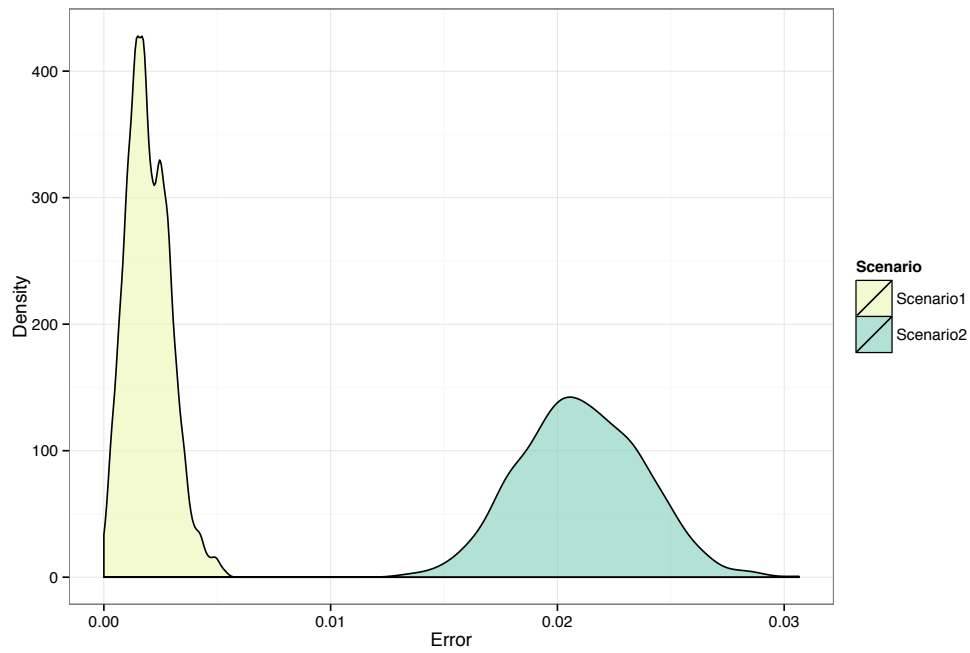


Figure 3.5 Clustering coefficients for Scenarios 1 and 2.

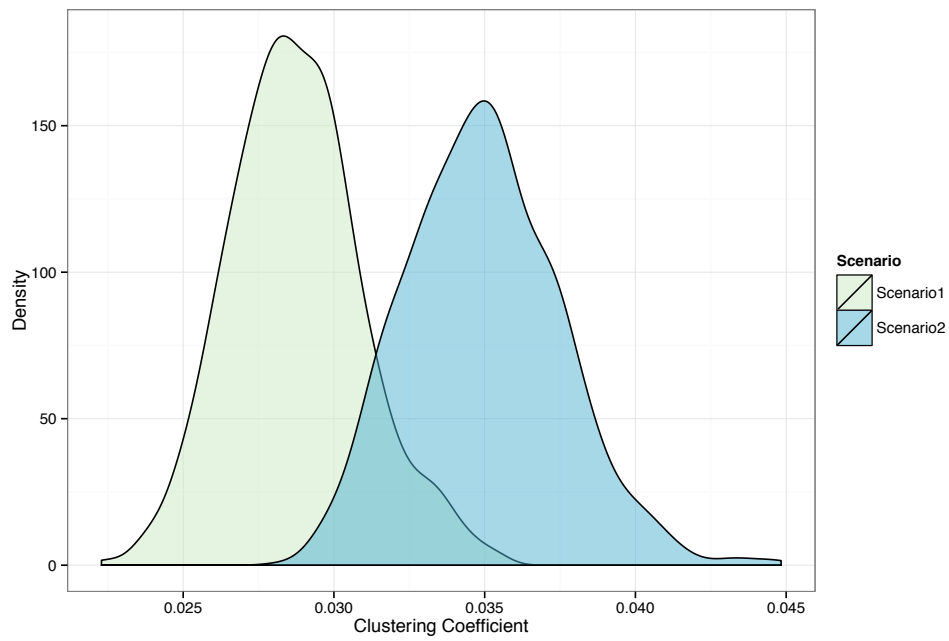


Figure 3.6 Visualization of Scenario 1 ensemble network 158. Red, blue, and green nodes represent physicians, other HCWs, and nurses, respectively.

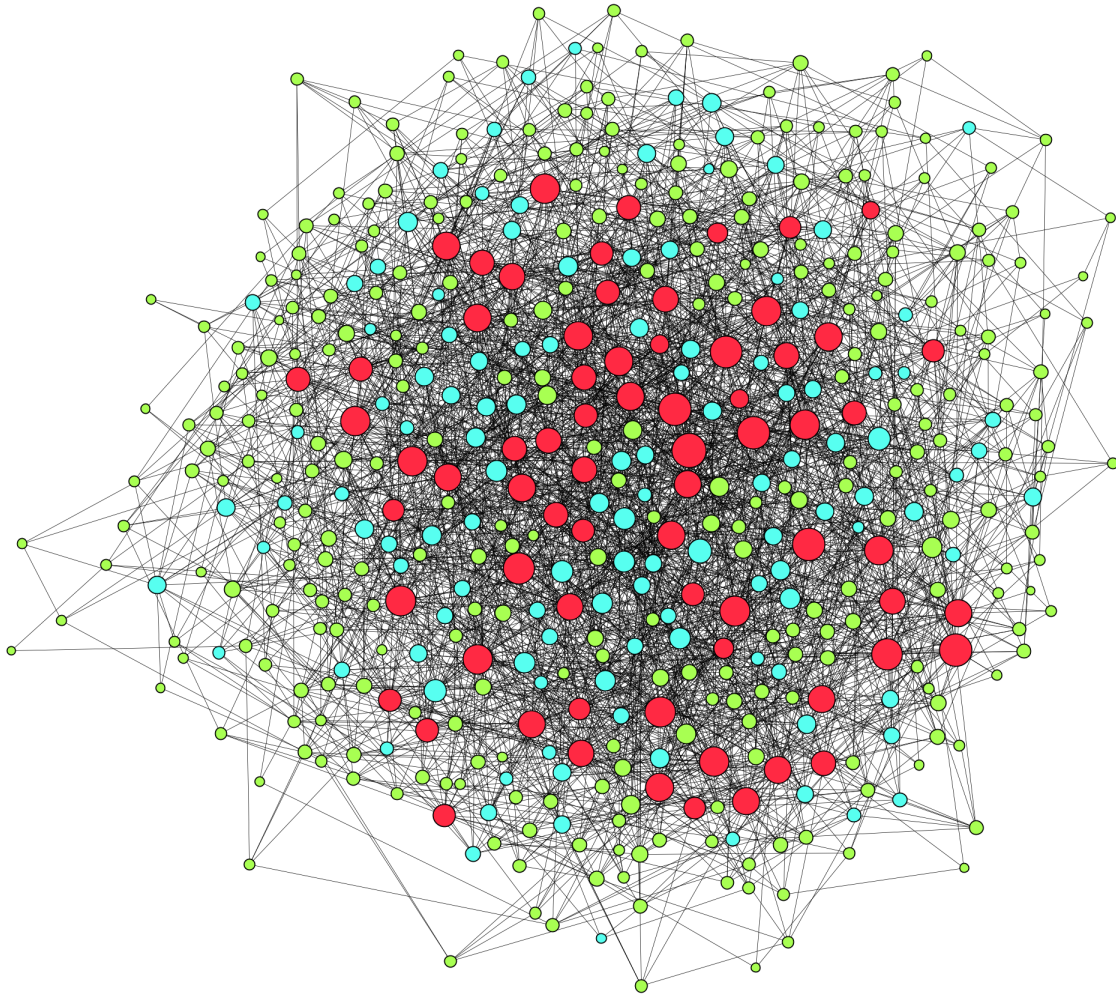
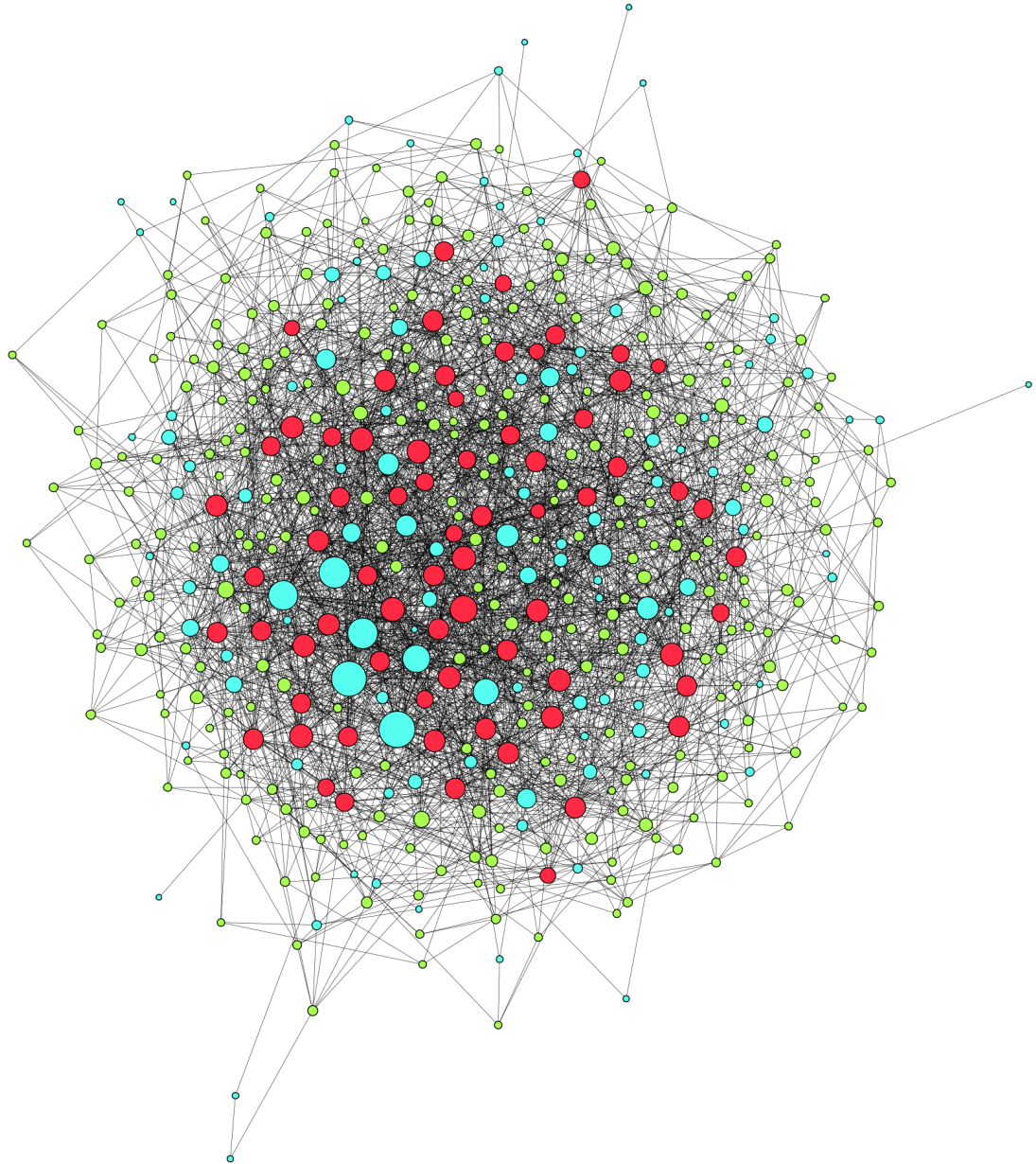


Figure 3.7 Visualization of scenario 2 ensemble network 675. Red, blue, and green nodes represent physicians, other HCWs, and nurses, respectively.



3.4 Discussion

Capturing the heterogeneous contact networks of healthcare workers in hospital settings is a challenge. Current high-resolution networks are only available for a single unit or ward, while hospital-wide networks only account for work related interactions and ignore any social or informal contact.

The network generation process developed in this chapter effectively produces multi-type networks of a hospital based on input parameters of degree and association. These parameters are flexible and easily extracted from pre-existing data sources. Many studies already provide potential input parameters, which can also be derived by survey, sensor, or observational data.

The error rate was below 5% for both scenarios, indicating that the code is effective in using assigned degree distributions as input. The significantly higher error in Scenario 2 is likely due to the scale free distribution of physicians, which includes extreme degree values of 80 or more (Figure 3.3). The low clustering coefficients for the networks additionally indicate that the randomness of the algorithm is intact.

This tool can be applied to create hospital networks of any size and combination of input parameters, given the availability of data. With the increase in surveys, sensor, and RFID technologies exploring the social interactions of the hospital, there is likely an abundance of scenarios to apply it.

4. Simulations on the Stylized Network

4.1 Introduction

Simulations of disease spread on networks have been utilized to understand transmission patterns, plan vaccination strategies, and optimize infection control. Simulations are able to demonstrate and predict a range of all possible outcomes, providing a valuable data source for policy decision-making.

This chapter simulates disease spread on the previously developed stylized hospital contact networks. Using the stylized networks, this chapter provides a contrast between networks of different structure, but similar mean degree. A range of scenarios and parameters were tested, confirming theoretical thresholds and the accuracy in the simulation process.

4.2 Methods

4.2.1 Model Characteristics

Simulations on the network will utilize an S-I-R model, where an individual node moves through three disease states of susceptible, infected, and recovered. All individuals start in the susceptible class. If susceptible, individuals may become infected by a diseased neighbor with a probability of the disease transmissibility, T . Once infected, the node is moved to the infected class, for disease duration D . All infected individuals are contagious for the duration of illness, and can infect neighbours according to the transmissibility value, T . Once infected for the disease duration D , individuals move to the recovered state, where they are not susceptible to another infection. Individuals remain in the recovered state for the remainder of the simulation, indicating permanent immunity after recovery (although this is a short-term simulation). Natural and disease-related deaths are not simulated in this model, the former because this is a short-term simulation, and the latter because we do not need this metric for our analyses, and recovered individuals exhibit similar (non-infectious) behavior as those who have died.

S-I-R models can be approximated to many diseases including influenza and measles, by altering model parameters and assumptions of immunity and death rate.

The mechanism of disease spread on the network is explained through *percolation theory*. Unlike traditional compartmental models where there is an equal probability of exposure to an infected individual throughout the population, exposure through network models require a susceptible individual to be connected to an infected individual by an *edge*. This means that the outcome of the disease is dependent on both the transmissibility of the disease, T , and the structure of the network.

4.2.2 Network Parameters

A primary parameter which determines the fate of an outbreak is the transmissibility T , which is the probability that an infected individual will effectively transmit the disease to a susceptible individual, given they have contact (Meyers, 2007).

R_0 is related to transmissibility as a function of the degree distribution (1), where k and k^2 represent the degree and squared degree, respectively.

$$R_0 = T (\text{mean } (k^2) / \text{mean } (k) - 1) \quad (1)$$

Based on the structure of the networks, there exists a critical transmissibility, T_c , which represents the threshold transmissibility past which a disease can spread throughout the network, and under which it cannot effectively spread. T_c is a function of the mean degree and mean squared degree (2)

$$T_c = \text{mean } (k) / \text{mean } (k^2) - \text{mean } (k) \quad (2)$$

The transmissibility, T , represents the probability of infection for the duration of illness. Because the duration of illness lasts more than 1 day (or timestep), in most scenarios,

another parameter is used to reflect the probability of infection per day, B . This value is a function of the duration of infection D and the transmissibility value T (3).

$$T = 1 - (1 - B)^D \quad (3)$$

4.2.3 Simulation

Simulation code was developed in MATLAB and can be found in Appendix B.

4.2.3.1 Inputs

Adjacency matrix: Directly imported. Can be created using network generation algorithm or other source

Population sizes: Numerical value for each subgroup, reflecting what corresponds to the adjacency matrix

Disease duration: The number of timesteps (days, in this case) that an infected individual is infectious for

Initial case: Whether the initial infection starts from a physician, nurse, or other HCW, or if this is a random selection

Beta: Probability of infection per timestep. Calculated from the transmissibility T value and disease duration

4.2.3.2 Process

- a) Adjacency matrix and population sizes are used to create lists of “neighbors” or contacts of each node in the network
- b) The initial case is randomly selected based on input parameters, and the percolation process begins.
- c) Within each timestep:
 - a. Neighbours are re-evaluated for susceptibility based on new infections
 - b. Infected nodes infect their susceptible neighbors with probability B
 - c. Infected individuals move to the recovered stage if they have been infected for the disease duration.
- d) The process runs until the number of infected individuals reaches 0

4.2.3.3 Outputs

Outbreak size: Total number of individuals who became infected in the population

Additional Outputs: The simulation process can produce a variety of outputs, from a simple measure of attack size, to individual infection trajectories. Outputs are customized for each research question.

4.3 Results

The calculated critical network parameters for scenarios 1 and 2, generated in the previous chapter, are summarized in Table 1. A disease duration $D = 6$ days was assigned for all simulations.

Table 4.1 Critical transmissibility and corresponding transmission probabilities for Scenarios 1 and 2.

	Critical Transmissibility	Corresponding B
Scenario 1	0.070	0.012
Scenario 2	0.063	0.011

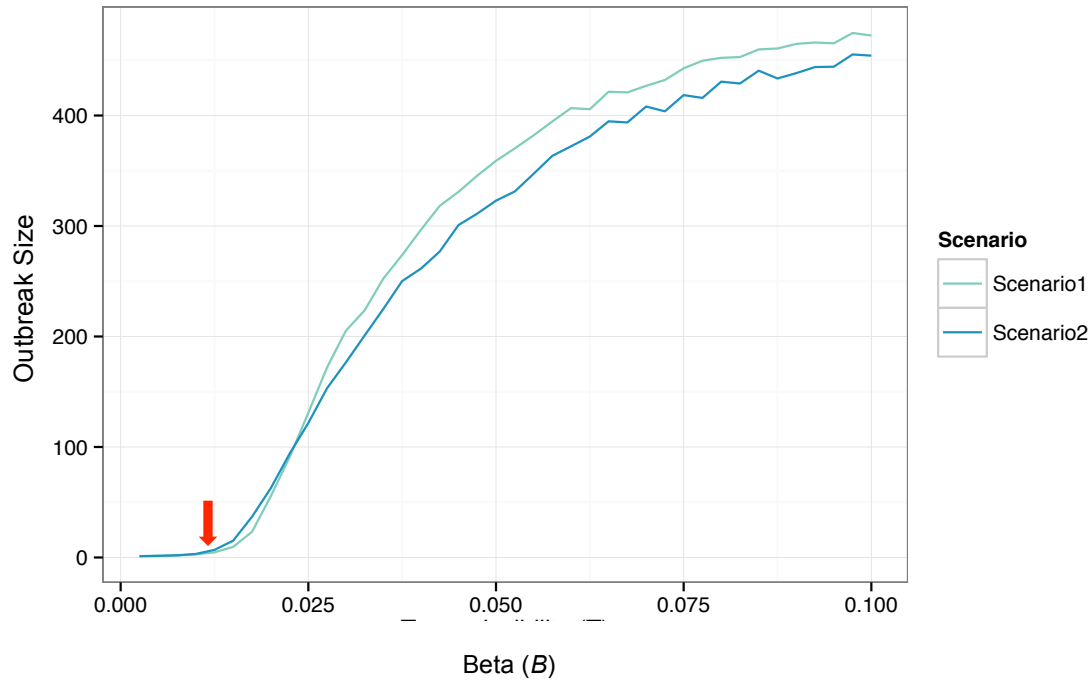
Three simulations were conducted, exploring transmissibility, network size, and initial infection.

4.3.1 Transmissibility Simulations

The behavior of disease spread across a range of beta (B) values was tested. Transmissibility values were selected across a gradient ranging from 0 to 0.1, representing values below, at, and above the critical threshold. Simulations were run 1,000 times at each value.

The mean outbreak size for each B value is plotted for both Scenarios 1 and 2 in Figure 4.1. As transmissibility increases, the mean outbreak size increases, approaching the total population size of 500 in both scenarios. A sharp increase in outbreak size occurs above the critical transmissibility value for each scenario. This demonstrates the significance of this threshold value in a disease's ability to effectively spread throughout a contact network.

Figure 4.1 Mean outbreak size over 1,000 simulations across a range of beta for Scenarios 1 and 2. Arrow indicates approximate critical value.



A density curve showing the distribution of outbreak sizes over 1,000 simulations for each B value was plotted for each scenario. Curves were plotted on the same graph, and divided into high values (Figure 4.2) and low values (Figure 4.3).

Figure 4.2 demonstrates two primary characteristics of the disease simulations. First, when the epidemic fails to spread at all past the first individual, this is indicated with an outbreak size of 1. In every simulation, there is always a probability that the epidemic fails to spread beyond the first or second individual. Thus, there is a spike in density at the low outbreak size for all B values.

Additionally, Figure 4.2 shows that as B increases above the threshold, there is less variation in outbreak size. This is because when B increases, the premature dieout of an epidemic is less likely due to the high probability of transmission at every connection. In other words, the disease is more likely to spread throughout the entire network rather than only affect a portion of it. A larger variation in outbreak size can be seen as

transmissibility decreases in this figure: values are not below threshold in this figure, so an outbreak is possible, but smaller outbreaks or disease die out are more likely at these values.

Figure 4.3 contrasts with Figure 4.2 in that, below the threshold, variation in outbreak size increases with B . This is because the transmissibility values in Figure 3 are below or approaching the threshold transmissibility value. As transmissibility approaches the critical value, there is a higher likelihood of disease transmission past the initial one or two infections. With decreasing transmissibility, simulations show a high density of outbreak sizes of low values, demonstrating the high likelihood of disease burnout upon introduction to the network.

Figure 4.2 Density curve of outbreak sizes at beta values above the threshold: ranging from 0.2 to 0.1.

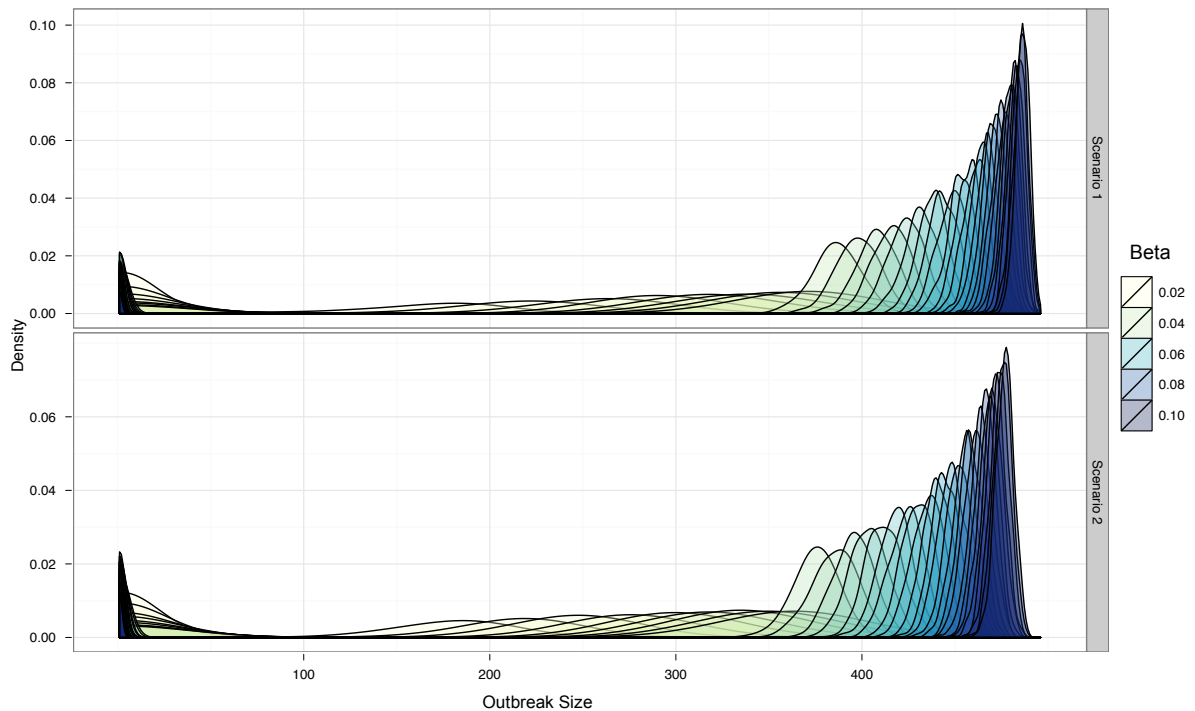
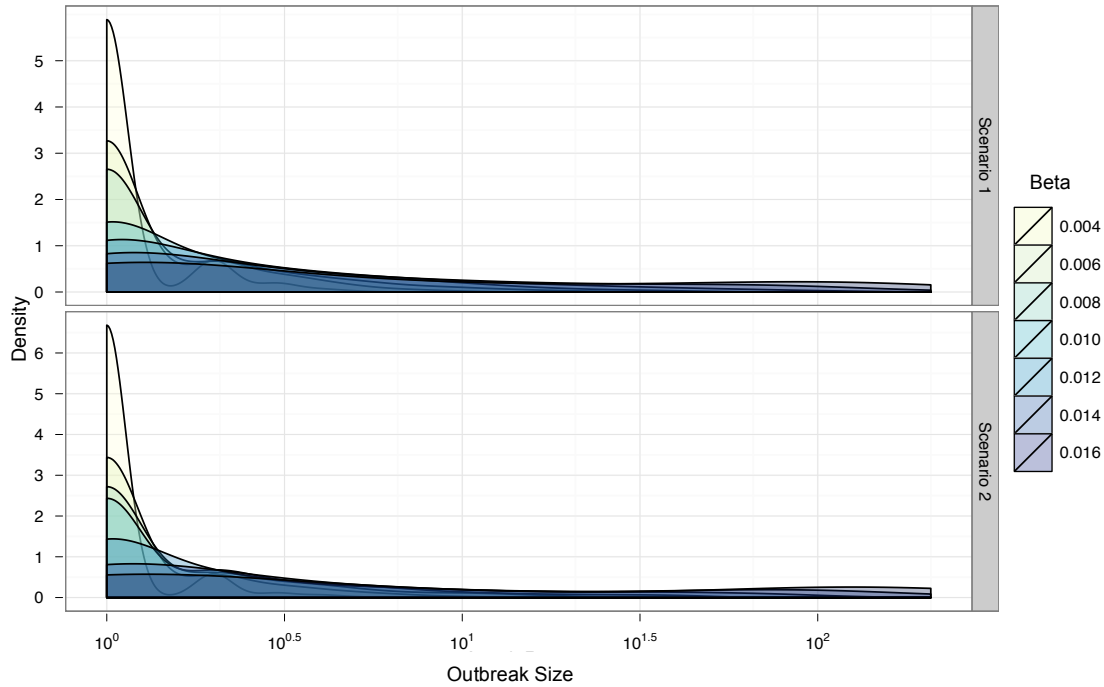


Figure 4.3 Density curve of outbreak sizes at beta values ranging from below the threshold value to just above: 0.004 to 0.016.



4.3.2 Network Size Simulations

In the next set of simulations, the effect of network size on attack rate was explored. Three contrasting transmissibility values were used, corresponding to low, medium, and high R_0 . Summarized in Table 4.2, these values were calculated from network parameters of scenarios 1 and 2.

Table 4.2 Low, medium, and high transmissibility parameters used for simulations.

	B	T		R_0
Low	0.01	0.03	Scenario 1	0.42
			Scenario 2	0.47
Medium	0.02	0.10	Scenario 1	1.43
			Scenario 2	1.59
High	0.05	0.25	Scenario 1	3.78
			Scenario 2	4.18

Using the network generation algorithm from Chapter 1, additional networks were created of sizes 1000 and 2000, preserving the subgroup proportions and parameters set in the original networks. Simulations were run 1,000 times for each of the high, medium, and low transmissibility levels. Histograms representing the range of outbreak sizes for network sizes 500, 1000, and 2000 can be found in Figures 4.4, 4.5, and 4.6, respectively.

Overall these results show a similar distribution of outbreak size across network sizes, with consistently low outbreak sizes at a low transmissibility, high outbreak sizes at high transmissibility (with a small probability of burnout), and medium-range and highly variable outbreak sizes with medium transmissibility.

Distributions were compared side-by-side as mean attack rate, which is calculated as a percentage of infected individuals of the total population (Figure 4.7). Statistically, these comparisons were evaluated. There was no significant difference in mean attack rate between network sizes of 500, 1000, and 2000 in the high transmissibility simulations: Scenario 1 (ANOVA, $f=1.90$, $p=0.15$), Scenario 2 (ANOVA, $f=0.30$, $p=0.74$). Similar results were found for the medium transmissibility simulations: Scenario 1 (ANOVA, $f=0.52$, $p=0.59$) Scenario 2 (ANOVA, $f=0.26$, $p=0.77$). At low transmissibility, significant differences between mean attack rate were found between all network sizes for both Scenarios 1 (ANOVA, $f=776.34$, $p<0.0001$) and 2 ($f=614.8$, $p<0.0001$).

The effect of network structure on attack rate was investigated. Comparisons between scenarios 1 and 2 are summarized in tables 4.3, 4.4, and 4.5 for network sizes of 500, 1000, and 2000, respectively. At medium and high transmissibility levels, network structure significantly affects mean attack rate. This effect is not seen at low transmissibility levels.

Figure 4.4 Distribution of outbreak sizes for the network size of 500, using low (A), medium (B), and high (C) transmissibility values, for scenarios 1 and 2.

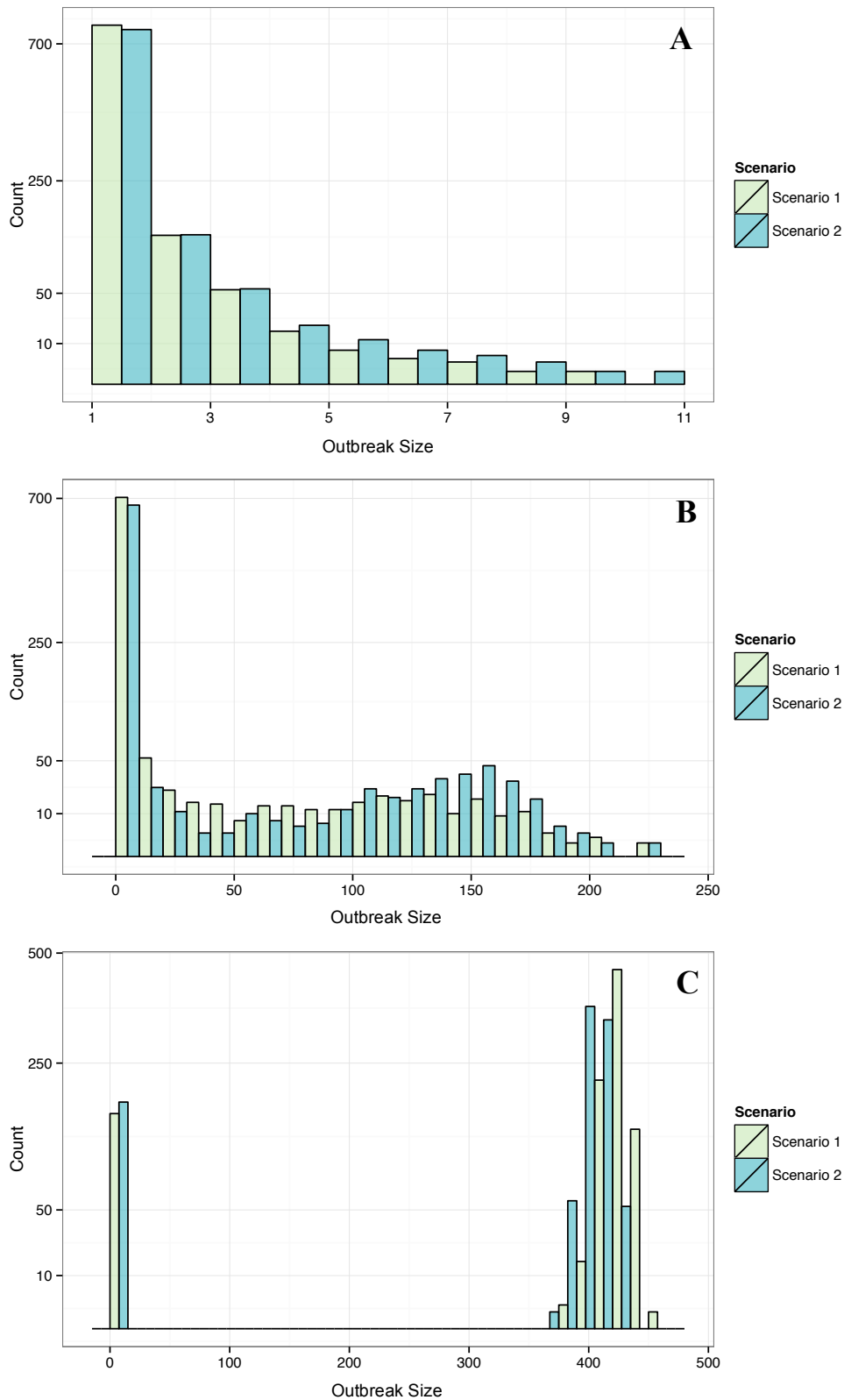


Figure 4.5 Distribution of outbreak sizes for the network size of 1,000, using low (A), medium (B), and high (C) transmissibility values, for scenarios 1 and 2.

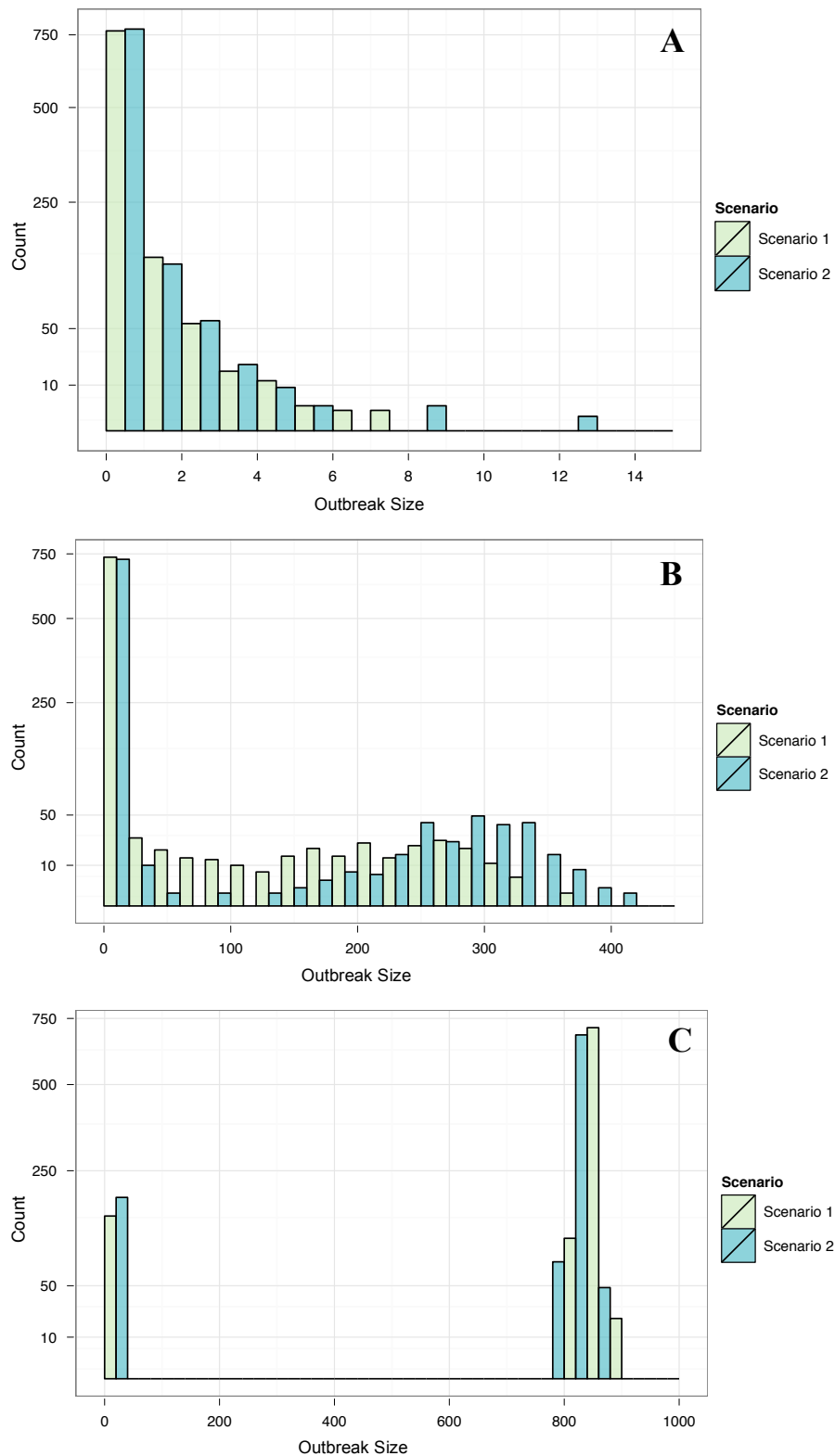


Figure 4.6 Distribution of outbreak sizes for the network size of 2,000, using low (A), medium (B), and high (C) transmissibility values, for scenarios 1 and 2.

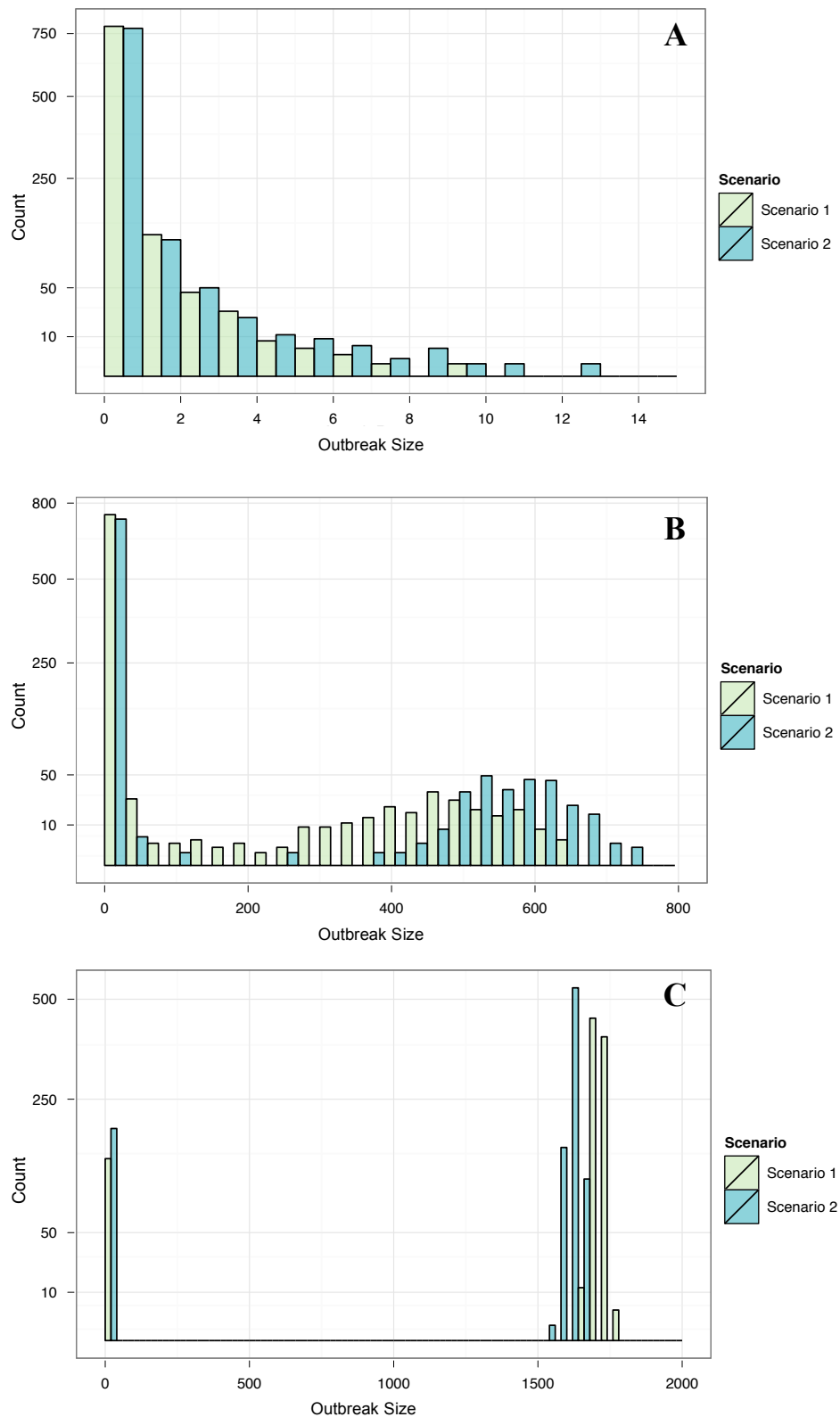


Table 4.3 Comparison of mean outbreak size and attack rates between Scenarios 1 and 2 for network size 500.

	Low T	Medium T	High T
Scenario 1	1.37 (0.27%)	26.33 (5.27%)	355.37 (71.07%)
Scenario 2	1.46 5 (0.29%)	39.92 (7.98%)	330.71 (66.14%)
p-value	0.056 (t = -1.915)	<0.0001 (t = -5.499)	<0.001 (t = 3.524)

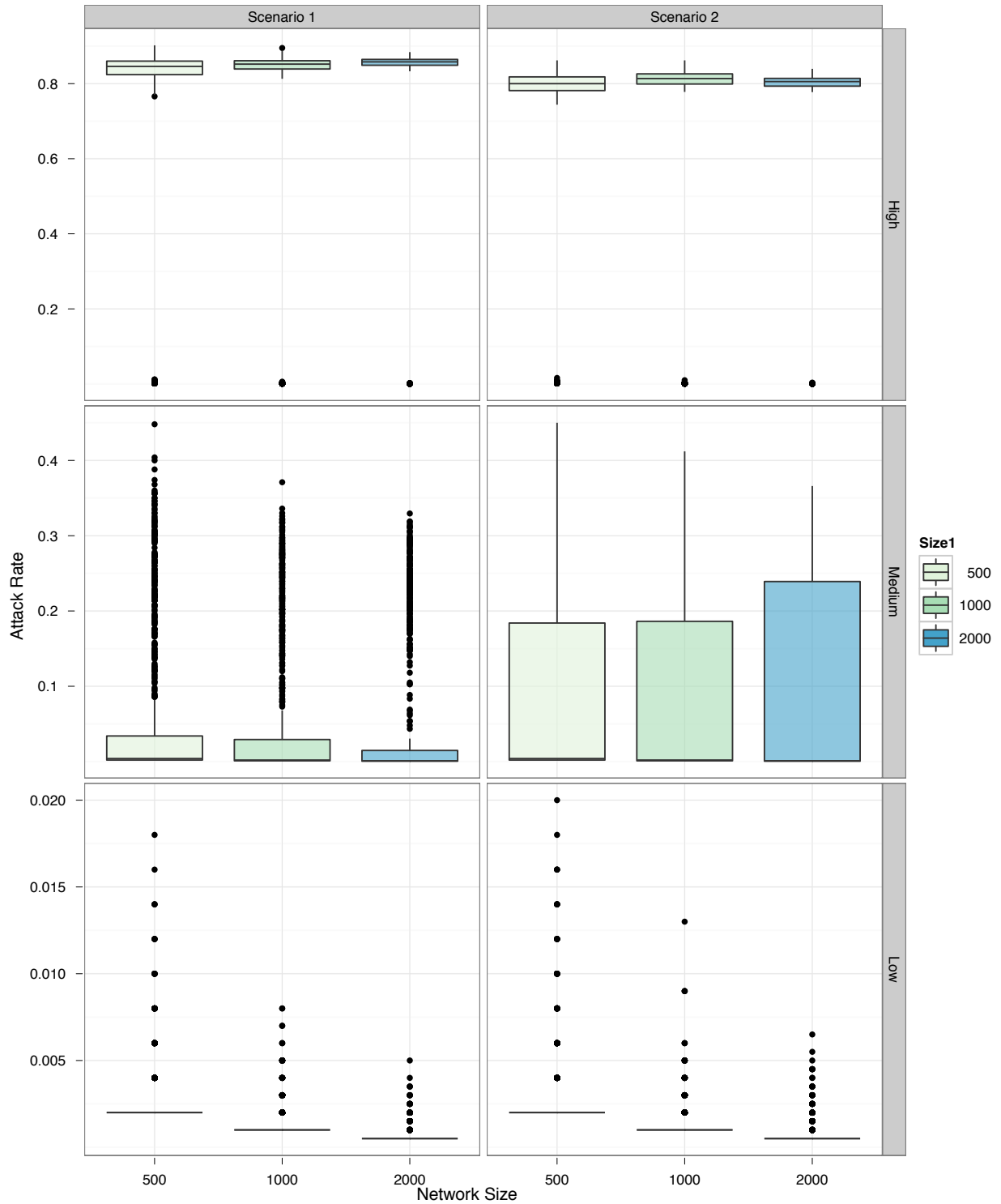
Table 4.4 Comparison of mean outbreak size and attack rates between Scenarios 1 and 2 for network size 1000.

	Low T	Medium T	High T
Scenario 1	1.39 (0.14%)	48.39 (4.84%)	724.00 (72.40%)
Scenario 2	1.40 (0.14%)	77.10 (7.71%)	663.05 (66.31%)
p-value	0.905 (t = -0.120)	<0.0001 t = -5.810)	<0.0001 (t = 4.337)

Table 4.5 Comparison of mean outbreak size and attack rates between Scenarios 1 and 2 for network size 2000.

	Low T	Medium T	High T
Scenario 1	1.39 (0.069%)	101.92 (5.10%)	1475.0 (73.75%)
Scenario 2	1.50 (0.07%)	151.71 (7.59%)	1305.67 (65.28%)
p-value	0.037 (t = -2.09)	<0.0001 (t = -4.985)	<0.0001 (t = 6.111)

Figure 4.7 Comparisons of the distribution of attack rates between network sizes for high, medium, and low transmissibility scenarios.



4.3.3 Initial Infection Simulations

Because each occupation has a distinct degree distribution, the impact of a disease starting with physicians vs. nurses, for example, may be significant. The impact of initial infection occupation on outbreak size was tested for both networks, at the previously set transmissibility levels of low, medium and high.

Each simulation was repeated 1,000 times and the results are summarized in Figures 4.8, 4.9, and 4.10 for low, medium and high transmissibility levels, respectively. Violin plots with an overlaid boxplot were used to visualize the distribution in outbreak sizes for each trial.

Mean outbreak size significantly differed between initial infection occupations of physician, nurse, and healthcare worker for both scenarios 1 and 2 (ANOVA, $p < 0.0001$). Tukey pairwise comparisons between each HCW combination revealed significant differences between all pairwise comparisons (Table 4.6).

Figure 4.8 Comparison of the distribution in outbreak sizes for simulations with initial infection starting with a physician, nurse, and other HCW at low transmissibility. White circles indicate mean.

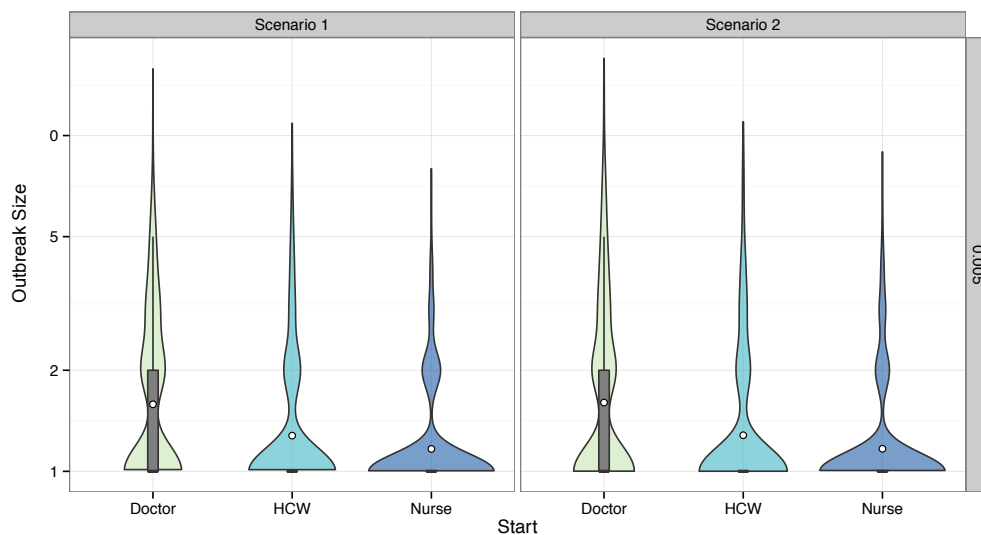


Table 4.6 Tukey pairwise comparisons for all combinations. 95% Confidence interval.

	Pair	P-Value
Scenario 1	Other HCW- Doctor	<0.0001
	Nurse- Doctor	<0.0001
	Nurse- Other HCW	<0.0001
Scenario 2	Other HCW- Doctor	<0.0001
	Nurse- Doctor	<0.0001
	Nurse- Other HCW	<0.001

At medium transmissibility levels, mean outbreak sizes were also significantly different according to occupation start, for both scenarios 1 and 2 (ANOVA, $p < 0.0001$). Tukey pairwise comparisons between HCW-types revealed significant differences between each comparison for both scenarios (Table 4.7).

Figure 4.9 Comparison of the distribution in outbreak sizes for simulations with initial infection starting with a physician, nurse, and other HCW at medium transmissibility.

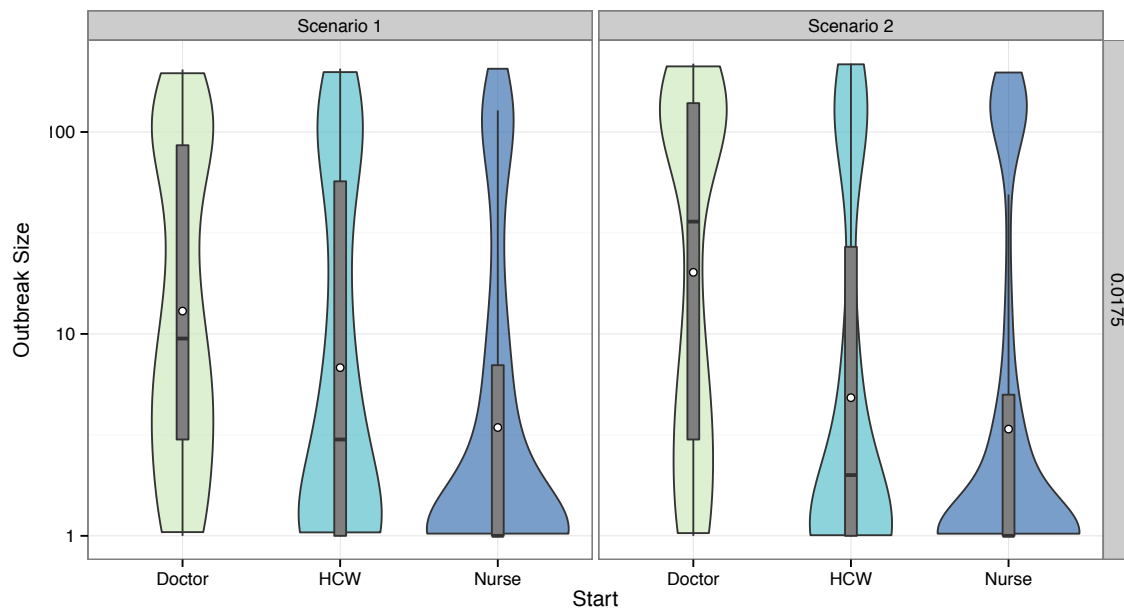


Table 4.7 Tukey pairwise comparisons for all combinations at medium transmissibility. 95% Confidence interval.

	Pair	P-Value
Scenario 1	Other HCW-Doctor	<0.001
	Nurse-Doctor	<0.0001
	Nurse- Other HCW	<0.0001
Scenario 2	Other HCW-Doctor	<0.0001
	Nurse-Doctor	<0.0001
	Nurse- Other HCW	<0.01

At high transmissibility, mean outbreak size significantly differed between outbreaks starting with physicians, nurses, and healthcare workers for both Scenarios 1 and 2 (ANOVA, $p < 0.0001$). Tukey pairwise comparisons between each HCW-type revealed significant differences between all pairwise comparisons (Table 4.7), except for the other-Nurse comparison for Scenario 2.

Figure 4.10 Comparison of the distribution in outbreak sizes for simulations with initial infection starting with a physician, nurse, and other HCW. High transmissibility.

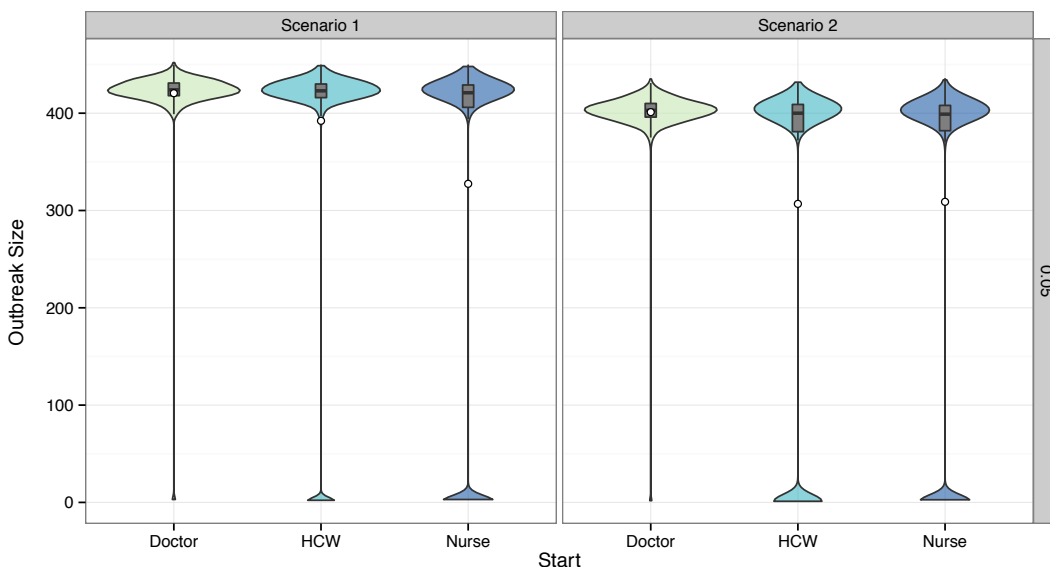


Table 4.8. Tukey pairwise comparisons for all combinations at high transmissibility. 95% Confidence interval.

	Pair	P-Value
Scenario 1	Other HCW-Doctor	<0.0001
	Nurse-Doctor	<0.0001
	Nurse- Other HCW	<0.0001
Scenario 2	Other HCW-Doctor	<0.0001
	Nurse-Doctor	<0.0001
	Nurse- Other HCW	0.941

4.4 Conclusions

The stylized simulations in this chapter demonstrate the ability of the network generation algorithm to produce networks that contrast in disease trajectory, allowing for the accurate capture of disease dynamics in different hospital settings. These simulations also confirmed the importance of disease transmission, network structure, and initial infection on the outcome of disease introduction in a population. Further, the findings were robust across different network sizes in the majority of scenarios, indicating that findings for one hospital may be applied to other settings, given similar subgroup degree characteristics.

4.4.1 Transmissibility Simulations

The simulations were effective in validating the percolation process created in MATLAB. The critical threshold of transmission calculated from the network aligned with observed critical values. These simulations also demonstrated the importance of the critical threshold for epidemic spread.

4.4.2 Network Size Simulations

The next simulations confirmed that that results may be robust in the case of uncertain or changing population size. Additionally, conclusions for a small hospital may be applicable to hospitals of a larger size, assuming similar degree distribution characteristics for subgroups, and a transmissibility above the critical threshold.

The significance of network structure on disease trajectory was also demonstrated. Scenarios 1 and 2 have similar mean degrees and critical transmissibility values, but differ in network structure. Scenario 1 is generated from three Poisson distributions, while Scenario 2 includes scale free, Poisson, and lognormal distributions. It was found that at medium and high transmissibility values, these networks had a significantly different mean outbreak size. Thus, it is important to obtain accurate input degree distribution data in the network generation process, as it significantly affects disease outcomes for medium and high transmissibility scenarios.

4.4.3 Initial Infection Simulations

Finally, it was found that the disease trajectory and success in spreading throughout the population is significantly affected by the starting occupation in most scenarios tested. At all transmissibility levels and both scenarios, the mean outbreak size was significantly higher when the disease started with physicians, compared to starting with other occupations. Physicians were assigned the highest mean degree among the subgroups. Therefore, it is apparent that the highly connected nature of the physician subgroups significantly affects the disease trajectory, if the outbreak starts with individuals within this subgroup.

5. Contact Patterns and Network Development for a Large Canadian Hospital

5.1 Introduction

Understanding the movement and contact patterns within hospital settings allows for more targeted and effective infection control interventions, and aids in the generation of contact networks for disease modeling and simulation.

Current studies have utilized electronic medical records to examine spatial movement throughout the hospital. Although these methods may provide information on a subset of hospital interaction, they fail to capture social or casual movement, such as visits to the cafeteria or meetings (Liljeros et al., 2007; Ueno & Masuda, 2008; Cusumano-Towner et al., 2013; Curtis et al., 2013).

Contact patterns for healthcare workers have also been examined through RFID tags, mote-based sensing, and direct observation (Polgreen et al., 2010; Hornbeck et al., 2012; Isella et al. 2011; Vanhems et al., 2013), suggesting the potential for superspreaders in the hospital setting (Vanhems et al., 2013), and differences in contact patterns between occupations (Isella et al., 2011). These studies are currently only within a single ward or unit, and thus are limited in generalizability to an entire hospital setting.

In addition to contact, it is important to note the patterns of movement throughout the hospital. This may reveal locations that can more readily propagate infection spread during outbreak scenarios, and may be targets in control strategies.

In order to better understand contact patterns in a hospital setting, this research chapter utilizes survey results from a large Canadian hospital. Both spatial movement and contact rates are analyzed in order to identify either locations or patterns in contact that may be targeted in infection control. Additionally, these observations are used to develop a

realistic contact network of a hospital, to be utilized for further disease modeling and simulation.

5.2 Methods

Healthcare workers at three major Canadian hospitals completed a questionnaire regarding hospital movement and contact as a part of a major research project, named the *CONNECT* study. A total of 3,048 HCWs responded to the survey, of approximately 8,000 contacted. Respondents chose one of 15 occupations, which were further aggregated into four groups: Nurses, Physicians, other HCWs, and Admin/support staff. 54% of responses were from the largest of the three hospitals. Because locations vary between hospitals, and staffing procedures likely differ, the largest hospital was chosen to conduct the movement and contact analysis for this study.

For location-based analyses, all occupation groups were included. For the contact analysis and network generation process, only the healthcare workers (nurses, physicians and others) were included.

Methods were divided into two sections: Survey analysis and network development.

5.2.1. Survey Analysis

5.2.1.1 Location Data

Each respondent recorded the hours and minutes spent per location in the hospital, per week. Locations were categorized as inpatient, outpatient, mixed, no patients, or public area. Data for a total of 234 locations in the largest hospital were analyzed in the hospital for frequency of visits and length of stay.

5.2.1.2 Contact Data

To examine HCW-HCW interactions, survey data for the self-reported average daily number of HCW contacts was utilized. A contact, from the questionnaire, is defined as spending two minutes within 1 meter of another HCW. All non-admin occupations were analyzed.

5.2.2 Network Development

The model parameters of degree distribution and association values, explained in Chapter 3, will be approximated from the survey data. The multi-type network generation algorithm (Chapter 3) will then be used to create a contact network of the hospital.

5.2.2.1 Degree Distribution

Fitting of the data was done by maximum likelihood methods using the *fitdistr* function in R. Weibull, exponential, lognormal, gamma were tested for each distribution, and the model which minimized the Akaike Information Criteria (AIC) was chosen. The AIC is a standard metric used to compare the fit of data to a distribution, and the model with the lowest value is considered to be the best fit.

5.2.2.2 Association Values

Because the survey does not ask which type of HCW an individual comes in contact with, there are not sufficiently detailed data to inform the association values, directly. We can infer approximate values from the hospital demographic proportions, and an assumption that all HCWs may come into contact with all other HCWs.

5.2.2.3 Network Generation

The association values and degree distribution are then used in the multi-type network generation algorithm (Chapter 3), and an ensemble of 1,000 networks is generated. Error rates and clustering coefficients are recorded for the ensemble.

5.3 Results

5.3.1 *CONNECT* Analysis

5.3.1.1 Location Analysis

For each location, the number of respondents who visited per week, and mean hours per week spent in the location, were plotted (Figure 5.1), and colored by location type.

It is apparent that a significant portion of hospital movement occurs in public spaces. The four most frequently visited locations represent the cafeteria, lobby café, and coffee shops, though the length of time of these visits is shorter relative to other locations visited. Additionally, inpatient locations have a greater number of visits per week compared to outpatient locations (Figure 5.2).

For all HCWs, the number of locations visited per week was compared between occupations (Figure 5.3). A significant difference was found between the number of locations visited for physicians, nurses, other HCW, and Admin/support staff ($p < 0.0001$). Nurses were found to be significantly lower in number of locations visited compared to physicians, other HCW and Admin/support staff. Tukey pairwise comparisons are summarized in Table 5.1, revealing that nurses are significantly less mobile throughout the hospital compared to other HCWs.

Figure 5.1 Locations in the hospital plotted by total visits per week, and mean hours spent for each location. Color represents location type.

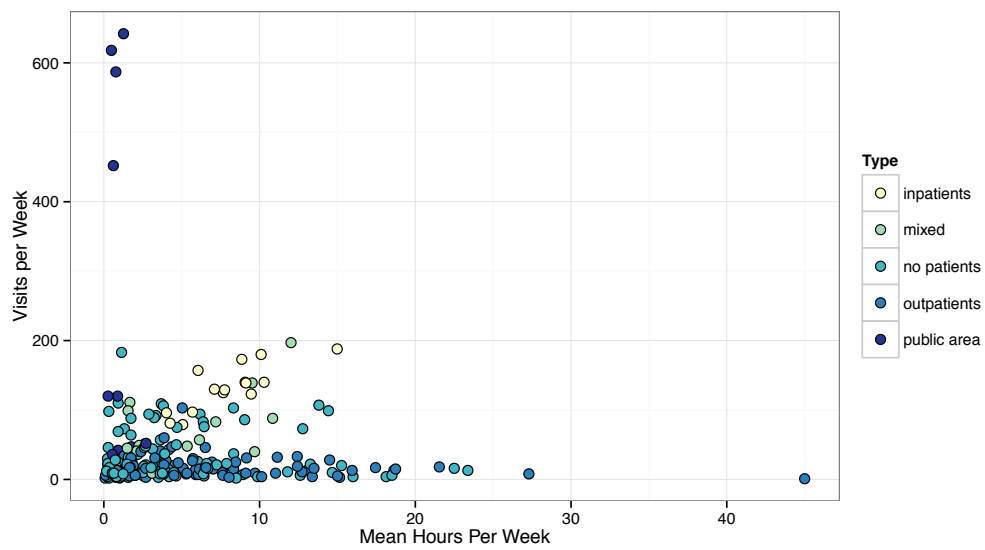


Figure 5.2 Locations in the hospital plotted by total visits per week, and mean hours spent for each location. Coloured and plotted by location type.

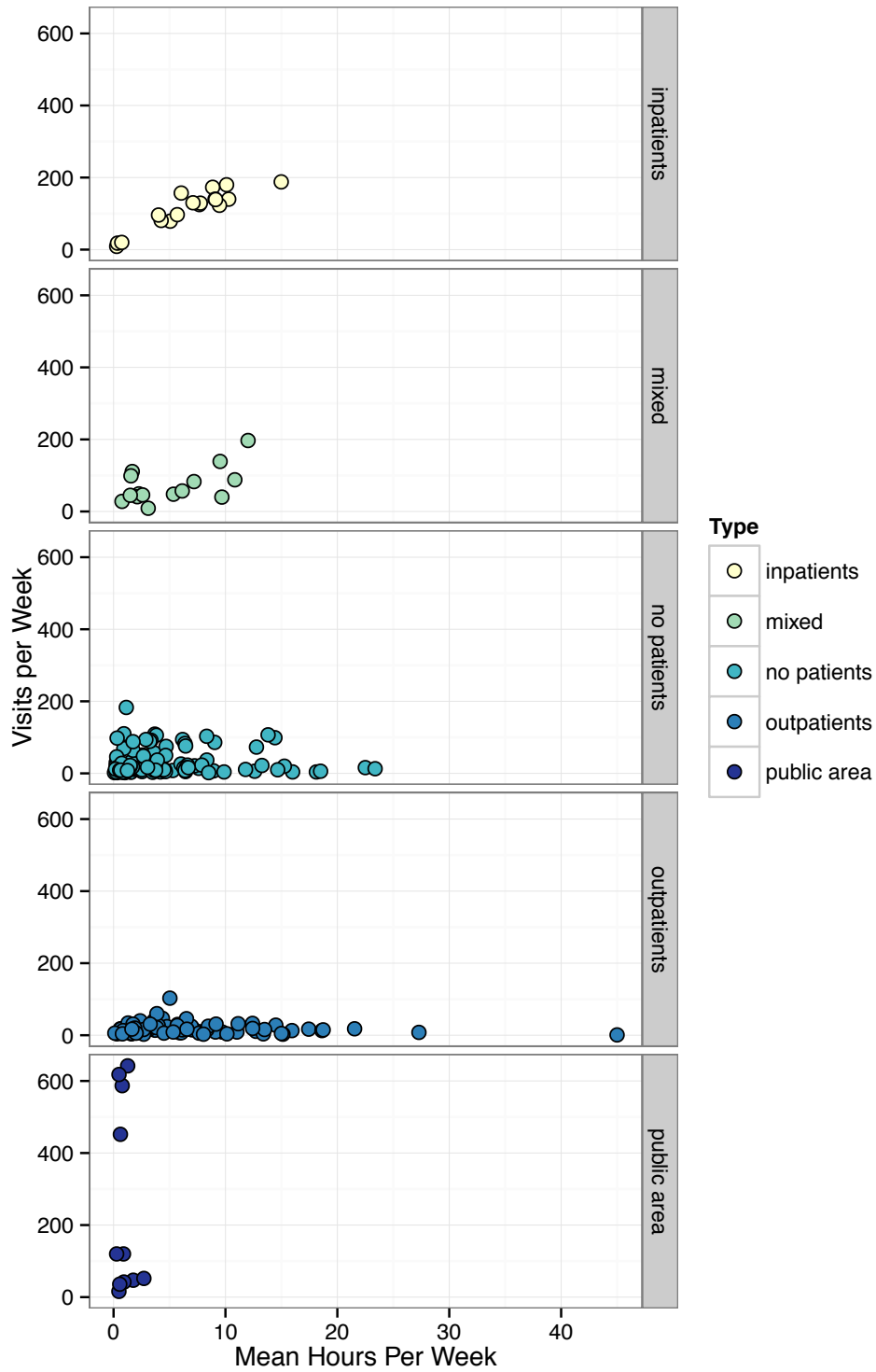


Figure 5.3 Distribution of the number of locations visited per week, by HCW type.

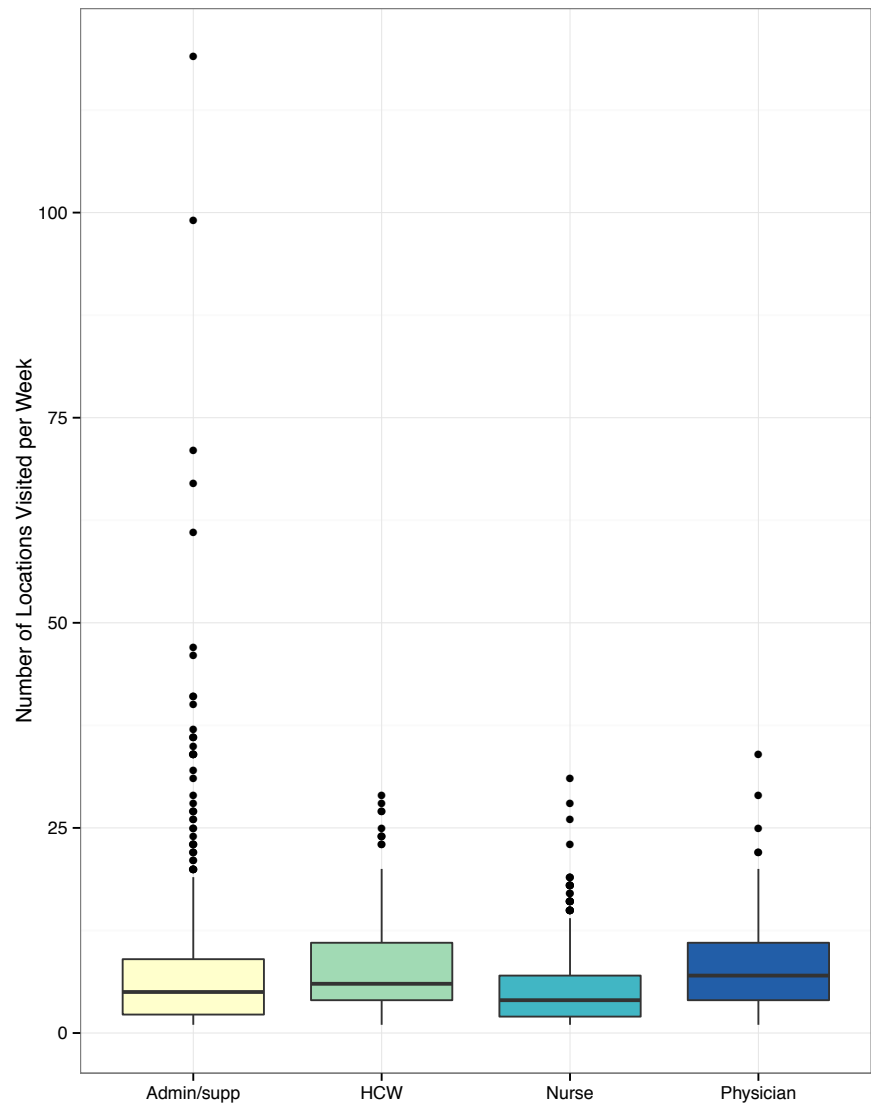


Table 5.1 Tukey pairwise comparisons for number of locations visited weekly, by HCW type.

Comparison	P-value
Other HCW- Admin	0.96
Nurse- Admin	<0.001
Physician- Admin	0.84
Nurse- Other HCW	0.001
Physician- Other HCW	0.98
Physician- Nurse	0.002

5.3.1.2 Contact Analysis

The patterns of between-HCW contact were analyzed from the survey data, for all non-admin occupations. When compared by occupation (Figure 5.4), physicians were found to have significantly lower mean HCW contact rates than both nurses and other HCWs, while nurses and HCWs don't significantly differ in mean occupation (ANOVA, $p < 0.01$, Table 5.2).

Figure 5.4 Distribution of reported HCW contact number between HCWs, nurses and physicians.

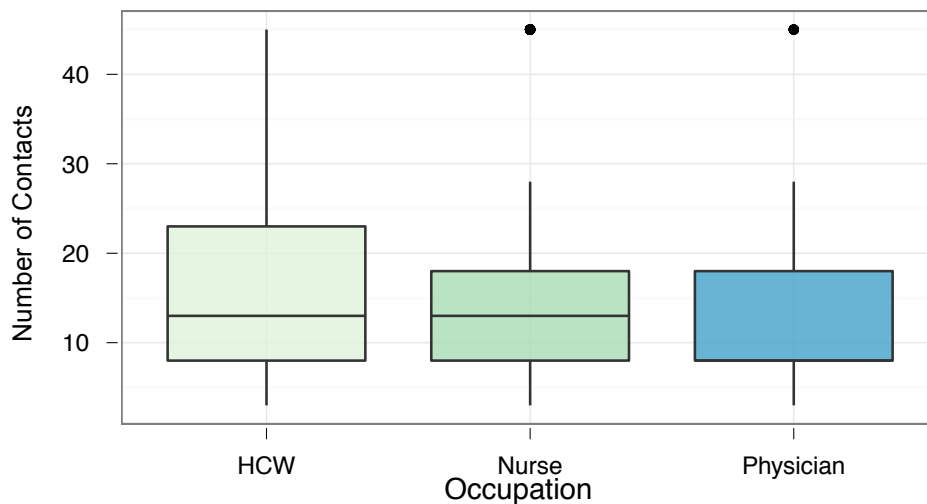


Table 5.2 Tukey pairwise comparisons for each occupation comparison.

Comparison	P-value
Nurse- Other HCW	0.39
Physician- Other HCW	0.002
Physician-Nurse	0.01

5.3.2 Network Parameterization

5.3.2.1 Degree Distribution

The distribution of contact number, or degree distribution, for each occupation is visualized in Figure 5.5. Four different statistical distributions were tested for each occupation, and AIC values were generated (Table 5.3). The best statistical fit for each occupation was plotted alongside the histogram (Figures 5.6, 5.7, 5.8). Physicians were best fit with a lognormal distribution, while the gamma distribution fit both other HCWs and nurses.

Figure 5.5 Histogram of the distribution of contact number, by HCW type.

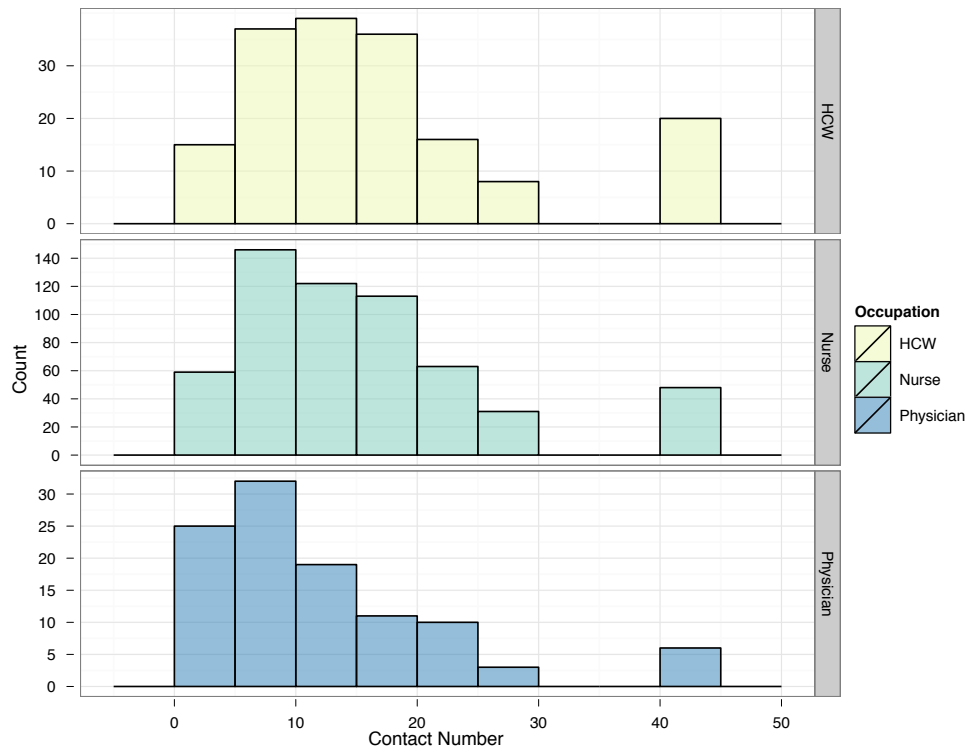


Table 5.3 AIC values for the statistical distribution fit.

	Physician	Nurse	Other HCW
Lognormal	732.63	4225.17	1270.11
Weibull	742.57	4235.49	1274.13
Exponential	754.98	4409.43	1322.36
Gamma	738.03	4212.32	1267.12

Figure 5.6 Histogram of physician degree distribution with fitted lognormal curve.

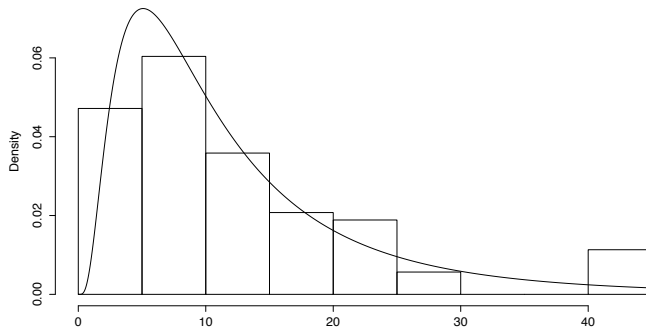


Figure 5.7 Histogram of nurse degree distribution with fitted gamma curve.

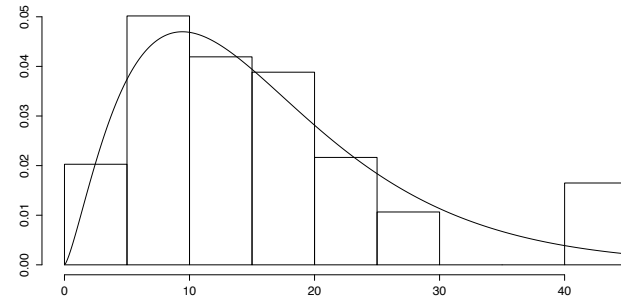
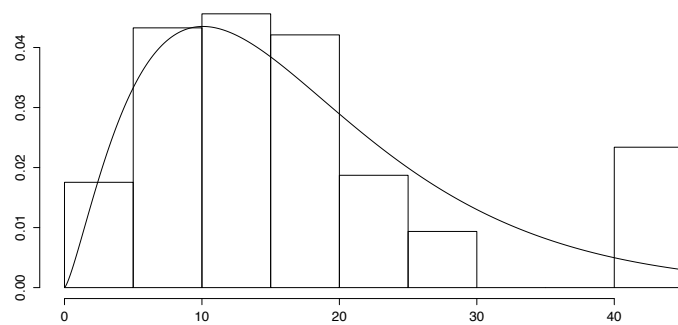


Figure 5.8 Histogram of other HCW degree distribution with fitted gamma curve.



5.3.2.2 Association Parameters

The demographics for the hospital studied indicate an estimated population of 650 physicians, 1405 nurses, and 750 other HCWs. From these values, all possible connections between each combination of occupation were calculated and summarized in Table 5.4.

Edge counts are divided by the total number of edges involving each occupation, which generates the association values. These values are described as the probability that an individual of occupation (row) is connected with an individual of occupation (column), rather than another occupation (Table 5.5). This matrix is asymmetric because the total number of edges involving each occupation differs. The sum of all outgoing association values is 1, because the edge must connect to one of the three occupations.

Table 5.4 Edge counts for each HCW combination.

	Physicians	Nurses	Other HCW	Total
Physician	210925	913250	487500	1611675
Nurses	913250	986310	1053750	2953310
HCW	487500	1053750	280875	1822125

Table 5.5 Association parameters for the hospital network.

	Physicians	Nurses	Other HCW	Total
Physician	0.13	0.57	0.30	1.0
Nurses	0.31	0.33	0.36	1.0
HCW	0.27	0.58	0.15	1.0

5.3.3 Network Generation

Utilizing the developed parameters and the network generation algorithm (Appendix A), a contact network was created. The network contains 2,805 nodes, reflecting the hospital size of non-admin personnel. The average degree for the entire network was 15.9. By occupation, the average degrees for nurses, physicians, and other HCWs were 16.1, 13.1, and 18.1, respectively. A visualization of the network was created in Gephi (Figure 5.9).

Error and clustering coefficients were evaluated over 1,000 iterations of the network generation process. Error in the process was below the significant threshold, at 0.003 (Figure 5.10), which indicates that the degree and association parameters from the survey data are in agreement. Clustering coefficients were also low at 0.008, confirming that the randomness of the connection algorithm was preserved (Figure 5.11).

Figure 5.9 Contact network representing a large Canadian hospital, 2,805 nodes.
Red, blue and green nodes represent physicians, other HCW and nurses,
respectively.

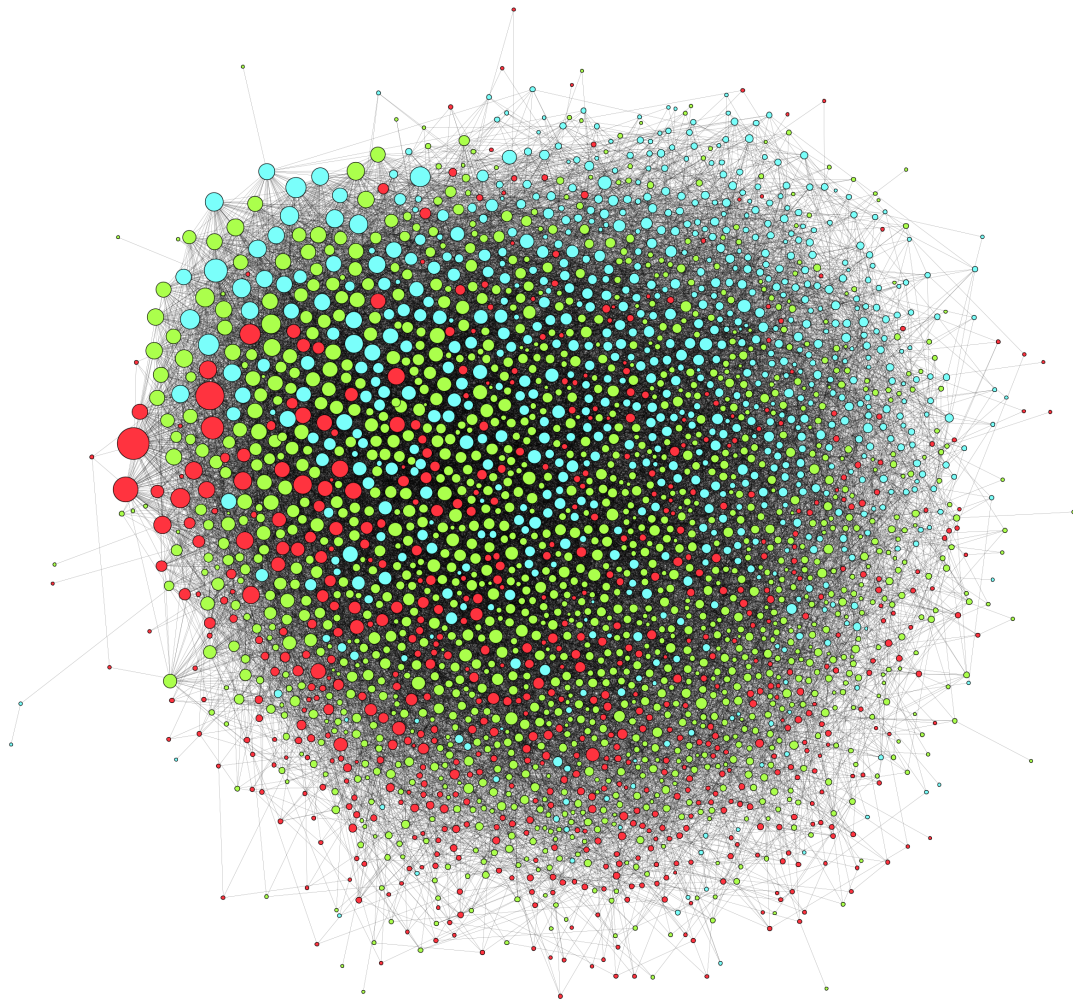


Figure 5.10 Density plot of the error rate over 1,000 iterations of the network generation process.

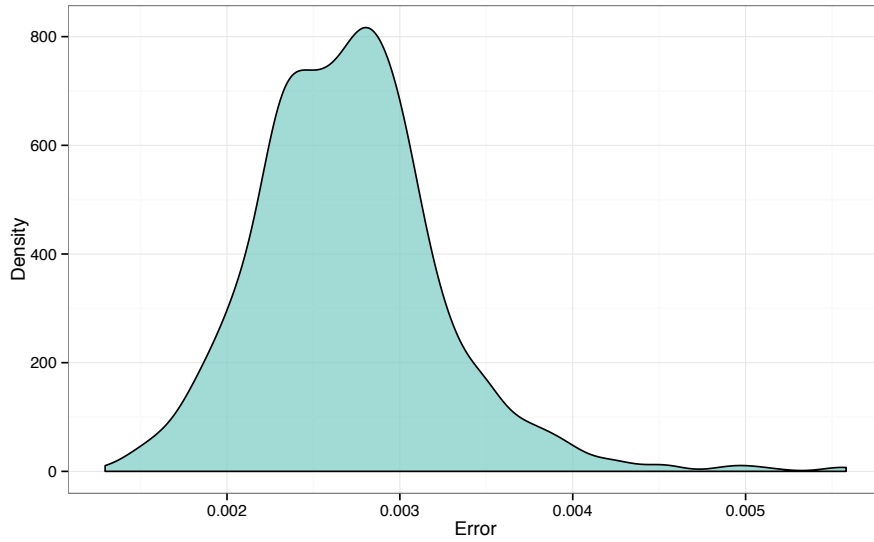
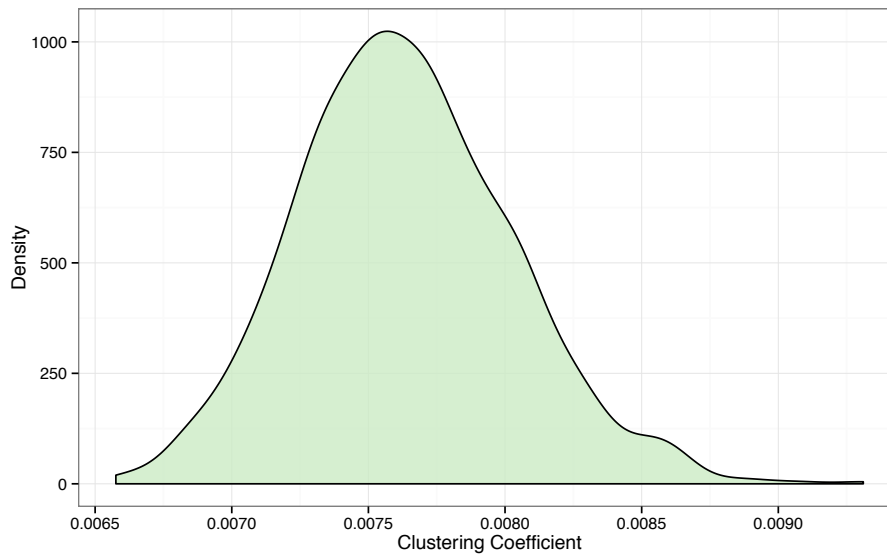


Figure 5.11 Clustering coefficients over 1,000 iterations of the network generation process.



5.4 Discussion

The analysis of the *CONNECT* data revealed patterns of movement and contact in major Canadian hospitals.

First, the most frequent locations visited by HCWs in the hospital were public spaces. The cafeteria, lobby café, and coffee shops made up the four most frequently visited locations. Although the length of time of these visits is shorter relative to other locations visited, the large proportion of the hospital that mixes in these locations may propagate disease spread in outbreak scenarios. In outbreak scenarios, cutting off these non-essential movements may aid in reducing infection transmission throughout the hospital.

Inpatient locations were found to have a greater number of visits per week compared to outpatient locations. Because of higher acuity of illness in inpatient settings, a greater diversity of HCWs may be in contact with the patient. This indicates that there may be a higher risk of disease spread in inpatient settings compared to outpatient settings. Staff cohorting in inpatient areas in outbreak scenarios would reduce the number of visits per location, per week, reducing the potential for disease propagation by HCWs.

The analyses of movement and HCW contact showed that nurses may be important for transmission within ward settings, but are not key spreaders throughout the hospital, due to their low number of average locations visited. Other HCWs may be the most important for superspreading events, having a high mobility throughout the hospital, as well as a high HCW-contact rate. Occupations that fall in this category include respiratory therapists and patient care attendants. Physicians, although mobile throughout the hospital, have an overall low HCW contact rate, in comparison.

The differences in HCW-contact were quantified in the fitted degree distributions, and were utilized in the network generation process. Degree distributions revealed gamma and lognormal distributions, indicating that all HCW types have a heavy right-tailed

distribution, meaning there are highly connected individuals in each occupation group, but the majority of connections are of lower degree. Thus, the presence of “superspreaders” in the hospital setting is possible for each of the HCW occupations, but the majority of HCWs have a lower number of connections. Targeting the highly connected individuals in each HCW-type for either contact-reduction or transmission-reduction measures may lead to efficient infection control in outbreak scenarios, though identifying these individuals may be challenging.

These parameters were effectively utilized to create a contact network of the hospital setting, representing the face-to-face interactions between subsets of the HCW population. This network can be utilized for modeling disease transmission in this Canadian hospital, and optimize and evaluate control measures.

6. Modeling Control Strategies in a Hospital Setting

6.1 Introduction

Evidence-based decision making for hospital infection control is a challenge, particularly for respiratory transmitted diseases. Respiratory diseases such as SARS and influenza have been propagated through hospital settings, but measurements of disease outcome and control measures are not consistent, and the incidence of these outbreaks is relatively rare (Varia et al., 2003; Voirin, Barret, Metzger, & Vanhems, 2009; Lee et al., 2003). Because of this, we currently lack evidence to supplement policy decision-making, and must rely on intuitive conclusions based on a handful of outbreak scenarios.

Mathematical modeling can be useful to aid in decision making for hospital infection control and outbreak planning. Because simulations can be run on a range of scenarios, the effect of an intervention can be studied on multiple diseases. Additionally, all possible outcomes of the disease trajectory can be explored due to repeated simulations, providing decision-makers with robust results that capture the full spectrum of possible outcomes.

Modeling respiratory infections to aid hospital decision-making has been utilized for vaccine planning (Polgreen et al., 2010), but more complex interventions involving transmission reduction and social distancing measures have yet to be studied.

Using the network model developed for a large Canadian hospital (Chapter 5), the spread of three contrasting diseases will be studied. Control strategies involving transmission reduction interventions, vaccination, and social distancing will be evaluated and optimized.

6.2 Methods

Simulations were run using the CONNECT hospital contact network developed in Chapter 5. This hospital network contains 2,805 HCW nodes: 650 physicians, 1405 nurses, and 750 other HCWs. The average degree of the network is 16. “Contact” between HCWs represents two minutes spent within one meter of each other, which is sufficient interaction for exposure to respiratory droplets and aerosolized particles (Bischoff et al., 2013). Association and connection parameters were derived from survey results, and quantify differences in contact between subgroups of the hospital including nurses, physicians and other HCWs.

First, simulations over a range of disease transmissibility values were run on the entire network without intervention, to demonstrate the critical transmissibility and general characteristics of an epidemic at baseline. For clarity in comparisons, duration of infection was kept constant at $D = 6$ days. The initial infection was selected randomly from all the individuals in the network.

To demonstrate the contrast in infection spread between diseases, parameters were chosen that represent a range of diseases. R_0 s and corresponding transmissibility values were assigned to approximate SARS-like illness ($R_0=0.9$) (Meyers et al., 2005), influenza- like illness ($R_0=1.5$) (Fraser et al., 2009), and measles- like illness ($R_0=5$) (Mossong & Muller, 2000). Although the R_0 of 0.9 corresponding to SARS is in the lower range of estimates of the disease (Meyers et al., 2005), we will assume a disease with R_0 of 0.9 as a SARS-like illness for contrast.

Control strategies were then evaluated for these diseases. Three types of interventions were tested in this study: Transmission-reducing interventions, vaccination, and social distancing. All simulations were run 1,000 times to capture the range in possible outcomes.

First, transmission intervention strategies were simulated on the network. Because this network models face-to-face interactions, the most realistic intervention to simulate is the wearing of masks. This model does not include physical contact, thus transmission measures such as hand washing cannot be accurately simulated.

Estimates of the efficacy of surgical and N95 masks for filtering particles from infected individuals vary. Measures range from as low as 25% when surgical masks have become ineffective or misused (Weber et al., 1993), to upwards of 90%, in scenarios where employees are fitted and correctly using N95 masks (Qian, Willeke, Grinshpun, Donnelly, & Coffey, 1998).

The reduction in transmission due to mask wearing was calculated by multiplying the baseline transmissibility rate by the mask efficacy rate, and then calculated the corresponding B value (Table 6.1). This value is used for interactions within the population or between sub-groups targeted for intervention.

Table 6.1 Transmissibility values and corresponding B for each level of efficacy.

R_0	T	B No intervention	B 25%	50	75	90
0.9	0.040	0.007	0.005	0.003	0.002	0.001
1.5	0.066	0.011	0.008	0.006	0.003	0.001
5	0.220	0.041	0.030	0.019	0.009	0.004

Next, vaccination strategies were simulated on the network. This is achieved by removing the targeted nodes and all connected edges from the network, because vaccinated individuals cannot become infected or transmit the disease. Two strategies were tested. The first selected random individuals from the network, both across the entire hospital, and within occupations.

An alternate strategy was defined as prioritizing individuals with the highest *degree*, or number of contacts to other HCWs. This strategy was tested for subgroups and the entire

hospital, at identical coverage levels. Thus, targeted vaccination meant immunizing the most connected 25%, 50%, 75% and 90% of HCWs.

Social distancing is simulated in the hospital model by randomly removing edges between HCWs. Because the targeted social interactions are likely non-specific to the HCW's occupation, or contact number, disconnections are randomly chosen across all individuals in the network. The simulated distancing measures range from removing 5% to 75% of the total edges in the network. The low social distancing measure of 5% is likely feasible through education during the flu season on limiting casual social interactions. More extreme distancing measures that may reach the highest distancing measures include cancelling in-person meetings, limiting cafeteria and coffee visits, and discouraging social hospital activities.

A more targeted approach to altering the social network, known as staff cohorting, was simulated. This occurs when HCWs who are usually assigned to work in multiple locations in the hospital are restricted to working in a single location. This type of intervention would theoretically reduce a fraction of the edges of the targeted individuals. To simulate this measure, we assume that the highest connected HCWs are the individuals who work in multiple locations, and thus will be targeted for staff cohorting. A previous study (Hornbeck et al., 2012) determined that the individuals with a high degree in the hospital network were the most mobile, simply due to the diversity in contacts. Individuals in the network with a degree greater than 30 were assumed to work in multiple locations, for the context of this simulation. These individuals were targeted for the staff cohorting simulation, and were reduced in degree by a proportion. The reduction in degree aims to simulate the reduction in contacts an individual would have if staff were cohorted to work in reduced locations. In total, 292 individuals had a degree greater than 30, approximately 10% of the population.

6.3 Results

The critical transmissibility value for the *CONNECT* network was found to be ($T=0.044$), and corresponding probability of infection was ($B=0.008$).

6.3.1 Baseline Simulations

First, simulations without intervention were conducted for a range of transmissibility values from $T=0$ to $T=0.47$, to examine the outbreak size both above and below the critical threshold of $T=0.044$.

Density curves of outbreak size for each simulation were plotted in Figures 6.1 and 6.2. Figure 6.1 contains density curves of the lower range of transmissibility values tested. As transmissibility approaches zero, outbreak size becomes increasingly concentrated at 1, meaning little to no disease spread occurs at these transmissibility values. Figure 6.2 contains density curves for transmissibility values above the critical threshold. As the transmissibility value increases above the threshold, outbreak size increases in number and density, approaching infecting the total hospital size of 2,805.

Mean outbreak size for each simulation was plotted in Figure 6.3. With increasing transmissibility, the mean outbreak size increases. Below the critical value of 0.044, little to no infection spread will occur. Once transmissibility is above the critical value of 0.0440, mean outbreak size increases and approaches the total hospital size of 2,805.

Figure 6.1 Density curves of outbreak sizes for simulations with transmissibility values ranging from 0 to 0.06.

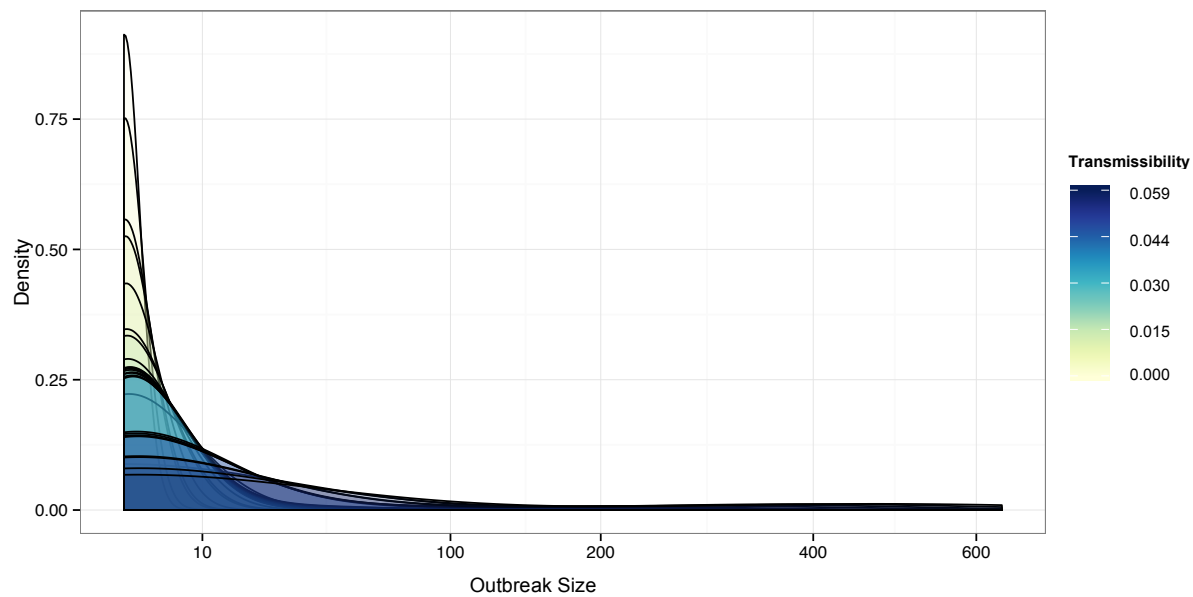


Figure 6.2 Density curves of outbreak sizes for simulations with transmissibility values ranging from 0.14 to 0.47.

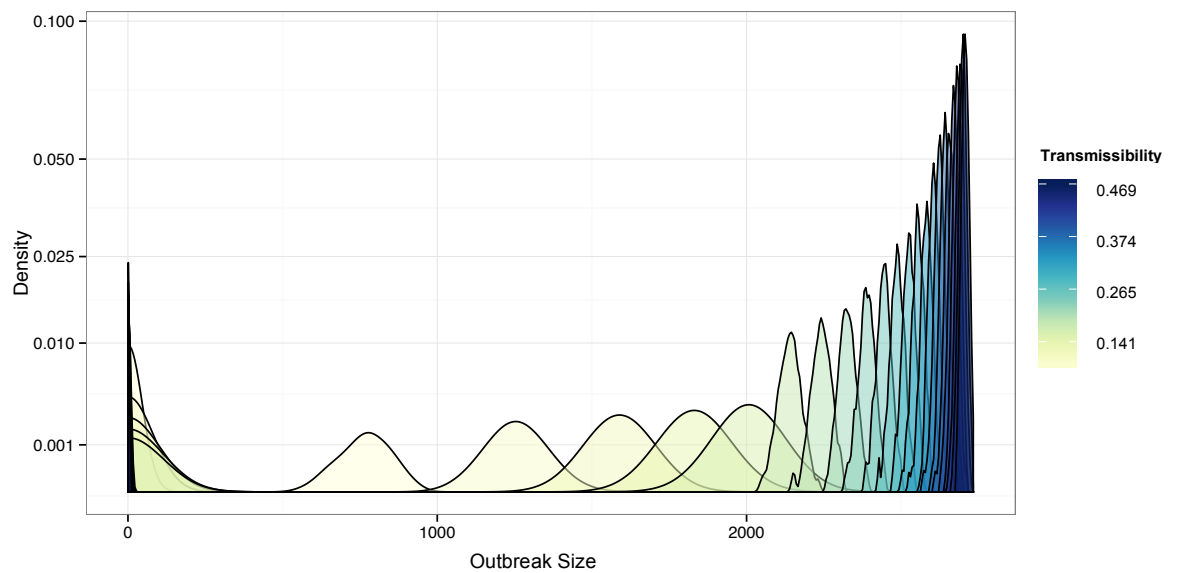
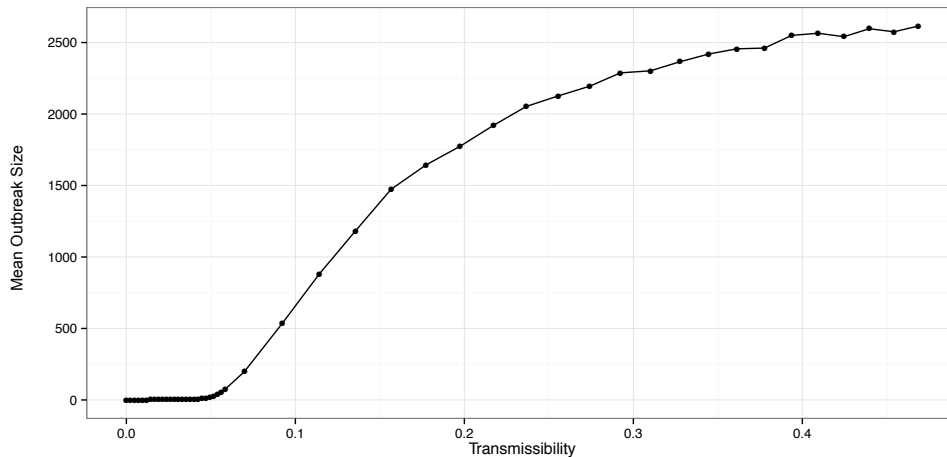


Figure 6.3 Mean outbreak sizes over a range of transmissibility values.



Simulations corresponding to the low, medium and high R_0 illnesses were explored in depth. Each scenario was run for 1,000 iterations and histograms of the results were summarized in Figures 6.4, 6.5, and 6.6, respectively. Significant differences were found between mean outbreak size of the low, medium and high R_0 illnesses (ANOVA, $F = 3656$ $p < 0.0001$).

For the low R_0 illness, the majority of outbreak sizes were below 10 (Figure 6.4). Out of 2,805 individuals in the network, the disease didn't spread to greater than 200 individuals. The mean outbreak size for the simulations was 4.2 individuals (Table 6.3). Because the R_0 is below the critical threshold of 1, epidemic scenarios on the network are not possible on a large-scale.

The medium R_0 illness demonstrated outbreak sizes that were divided: approximately 75% of the scenarios showed spread similar to that of the low R_0 illness, while the remaining scenarios showed a medium-level spread of illness between 450 and 800 individuals in the network. The probability of an epidemic of the medium R_0 illness > 200 individuals in the network is 0.26, with a mean outbreak sizes of 162 (Table 6.3). The just-above critical threshold R_0 of 1.5 for the medium R_0 illness means that there is still a high likelihood of little to no disease spread, but there is an ability to propagate infection through a large fraction of the network.

Finally, simulations of the high R_0 illness show a low incidence of little to no disease spread (Figure 6.6), and the majority of simulations spreading past 2,000 infections. The mean outbreak size across simulations was 1902. The probability of an epidemic of the high R_0 illness >200 individuals in the network is 0.81 (Table 6.3). The high R_0 and corresponding transmissibility of the measles-like illness means that with each interaction simulated in the network, there is a much higher chance of disease spread compared to the less transmissible diseases, which is reflected in the elevated mean outbreak size and epidemic probability.

With these simulations, there are no interventions simulated, thus the chances of an epidemic are very high. In the following simulations, interventions will be compared.

Table 6.2 Corresponding R_0 , transmissibility and B values for SARS-, influenza-, and measles- like illness.

	R_0	T	B
SARS-like	0.9	0.0396	0.0067
Influenza-like	1.5	0.0660	0.0113
Measles-like	5	0.2200	0.0406

Figure 6.4 Histogram of outbreak sizes over 1,000 simulations for the low R_0 illness.

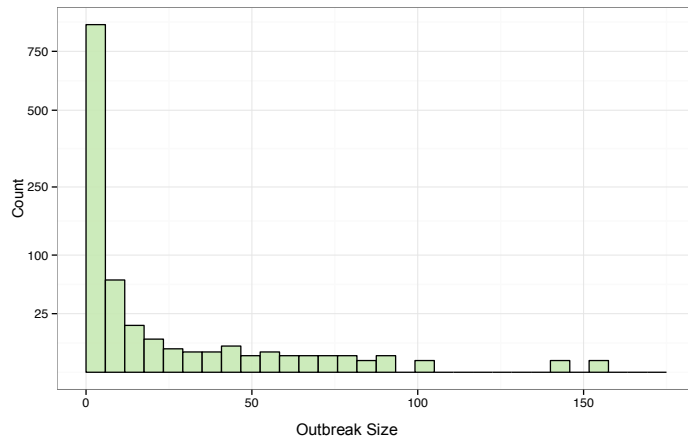


Figure 6.5 Histogram of outbreak sizes over 1,000 simulations for the medium R_0 illness.

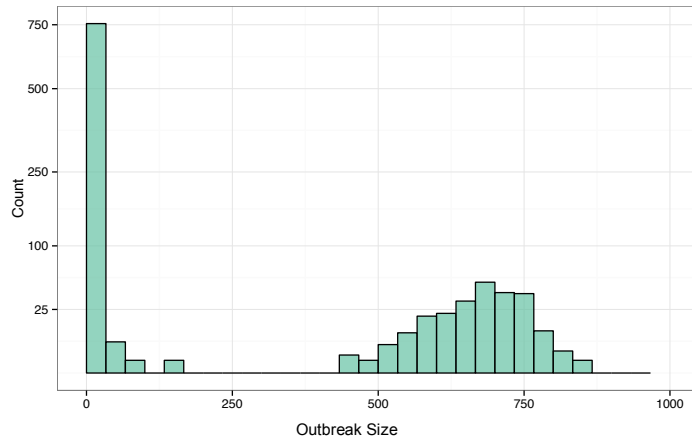


Figure 6.6 Histogram of outbreak sizes over 1,000 simulations for the high R_0 illness.

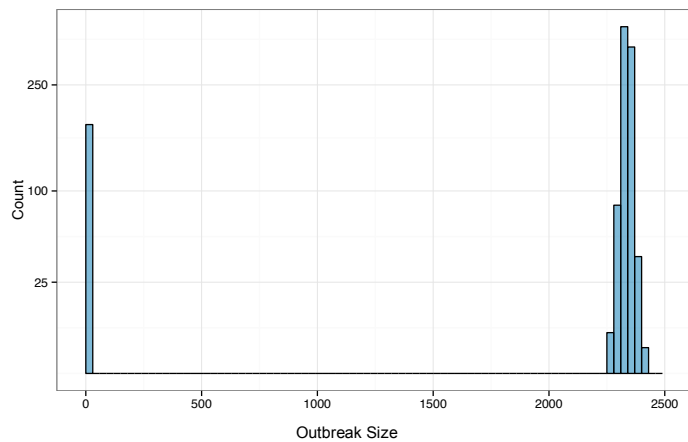


Table 6.3 Mean outbreak sizes and probability of an epidemic for the three disease scenarios tested.

R_0	Mean outbreak size (no intervention)	Probability of an Epidemic >200
0.9	4.2	0
1.5	162.4	0.26
5	1901.5	0.81

6.3.2 Transmission Interventions

First, transmission intervention strategies were simulated on the network. Results summarized in Table 6.4 show that for mask efficacies of 50% and greater, targeting nurses has the same impact as targeting the entire hospital in preventing a hospital outbreak, for the medium R_0 illness. The low R_0 illness has no probability of an epidemic regardless of the intervention. Successful intervention for the high R_0 illness only occurs when all HCWs are targeted with a 90% efficacy.

Table 6.4 Probability of epidemic (shaded grey), or mean outbreak size (if probability of epidemic is zero). Grey shading indicates unsuccessful intervention.

$R_0=0.9$	25% Efficacy	50	75	90
All HCWs	2.02	1.43	1.16	1.06
Physicians	3.15	2.18	2.20	1.97
Nurses	2.22	1.58	1.29	1.16
Other HCW	2.48	2.20	1.84	1.68

$R_0=1.5$	25	50	75	90
All HCWs	0.04	2.38	1.38	1.13
Physicians	0.16	0.12	0.07	0.07
Nurses	0.07	3.20	1.58	1.34
Other HCW	0.14	0.06	0.01	0.001

$R_0=5$	25	50	75	90
All HCWs	0.73	0.53	0.11	1.60
Physicians	0.81	0.74	0.70	0.65
Nurses	0.76	0.60	0.39	0.20
HCW	0.80	0.72	0.66	0.61

6.3.3 Vaccination Strategies

Vaccination strategies were simulated and compared on the *CONNECT* network. Because we do not have data on pre-existing vaccinations or immunity of this population, the vaccination coverage will be instead referred to as “preexisting immunity”, meaning that the percentages indicated are not for vaccination coverage, but immunity at the start of an epidemic.

The first vaccination strategy tested targets individuals randomly, either within subgroups or the hospital as a whole. At preexisting immunity rates of 75% or greater, it is just as

effective to target nurses, compared to targeting the entire hospital, for the medium R_0 illness. Hospital-wide vaccination was required to prevent disease spread for the high R_0 illness, at a rate of 75% or greater.

Next, the degree-based strategy was tested, which targets those with the highest degree, or number of contacts to other HCWs. Results indicate that prioritizing HCWs based on their connectivity is more effective than random vaccination for disease containment (Table 6.6). For the medium R_0 illness, vaccinating the highest connected nurses was effective at preventing an epidemic at coverage rates as low as 25%. With the random targeting strategy, nurses would have to be vaccinated at 75% or greater to confer the same results (Table 6.5). With hospital-wide vaccination, only 25% coverage is required to prevent an epidemic using the targeted strategy, compared to 50% for random vaccination.

Table 6.5 Probability of an epidemic (shaded) and mean outbreak size, where probability of epidemic is zero, for preexisting immunity at three different R_0 values. Random vaccination strategy.

$R_0=0.9$	25% Coverage	50	75	90
All HCWs	1.90	1.45	1.16	1.08
Physicians	3.10	2.97	2.40	2.60
Nurses	2.40	1.92	1.50	1.28
Other HCW	3.40	2.63	2.16	1.96

$R_0=1.5$	25% Coverage	50	75	90
All HCWs	0.03	2.34	1.37	1.11
Physicians	0.20	0.16	0.11	0.09
Nurses	0.15	0.01	2.52	1.77
Other HCW	0.17	0.12	0.04	0.01

$R_0=5$	25% Coverage	50	75	90
All HCWs	0.75	0.58	19.07	1.76
Physicians	0.84	0.80	0.80	0.80
Nurses	0.78	0.73	0.58	0.47
Other HCW	0.81	0.79	0.75	0.76

For the high R_0 illness, targeted vaccination of subgroups was not sufficient for disease containment. This is likely because the transmissibility is high enough that all occupations must be partially covered. Hospital wide, targeted vaccination was still more effective than randomized vaccination. Vaccinating the top 50% of connected individuals was sufficient to prevent epidemic spread throughout the hospital, compared to 75% for the random vaccination.

Table 6.6 Probability of an epidemic (shaded) and mean outbreak size, where probability of epidemic is zero, for three different R_0 values. Percentages represent vaccination of the top percent of connected individuals in the population.

$R_0=0.9$	25% Coverage	50	75	90
All HCWs	1.25	1.09	1.04	1.00
Physicians	2.19	2.30	2.24	2.30
Nurses	1.46	1.29	1.22	1.23
Other HCW	1.95	1.90	1.90	1.90

$R_0=1.5$	25% Coverage	50	75	90
All HCWs	1.68	1.20	1.06	1.01
physicians	0.09	0.09	0.08	0.08
Nurses	2.63	1.64	1.38	1.42
Other HCW	0.02	0.003	0.002	0.01

$R_0=5$	25% Coverage	50	75	90
All HCWs	0.42	2.33	1.19	1.03
Physicians	0.79	0.78	0.77	0.78
Nurses	0.72	0.45	0.30	0.26
Other HCW	0.78	0.72	0.71	0.72

6.3.4 Social Distancing Interventions

The next type of intervention to be evaluated was social distancing, achieved by removing edges within the network structure. A large fraction of HCWs visit public spaces such as the cafeteria, and coffee shops (Chapter 5), indicating that many of the contacts that occur within the hospital may be social and non-essential to hospital functioning. In an outbreak scenario, measures to limit non-essential social interactions in hospital settings may lead to a decrease in network density, and thus reduced disease propagation.

Results from the simulations show that generalized social distancing measures are not sufficient at preventing epidemic spread for the high R_0 illness, even in the most extreme scenario of 75%. As expected, epidemic spread failed to occur in the low R_0 illness, regardless of intervention. For the medium R_0 illness, social distancing measures of 50% were sufficient to prevent epidemic spread of the disease

Table 6.7 Probability of an epidemic (shaded) and mean outbreak size, where probability of epidemic is zero, for a range of social distancing intervention levels.

	5% Distancing	10%	25%	50%	75%
$R_0=0.9$	3.51	3.12	2.04	1.51	1.17
$R_0=1.5$	0.21	0.18	0.04	2.45	1.28
$R_0=5$	0.83	0.79	0.74	0.57	0.14

A more targeted approach to altering the social network, known as staff cohorting, was simulated by reducing highly connected individuals in the network by a percentage. Results were summarized in Table 6.8. For the medium R_0 illness, cohorting at 75% successfully prevented an epidemic. No cohorting measure was sufficient to prevent an epidemic for the high R_0 illness.

Table 6.8 Probability of an epidemic (shaded) and mean outbreak size, where probability of epidemic is zero, for a range of cohorting levels.

	5%	10%	25%	50%	75%
$R_0=0.9$	3.45	3.06	2.66	2.08	1.88
$R_0=1.5$	0.24	0.23	0.16	0.02	5.94
$R_0=5$	0.86	0.81	0.84	0.82	0.82

The social distancing and staff cohorting interventions may be used together in an outbreak scenario to maximize efficacy. This was tested for each combination of interventions levels (Table 6.9). When utilized together, cohorting at 50% and social

distancing at 25% is sufficient to prevent epidemic spread of the medium R_0 illness. These levels are more feasible than 75% cohorting and 50% distancing, which would be necessary if utilized separately.

Table 6.9 Probability of an epidemic (shaded) and mean outbreak size, where probability of epidemic is zero, for a combination of distancing and cohorting measures.

$R_0=0.9$	5% Cohorting	10%	25%	50%	75%
5% Distancing	3.49	3.02	2.17	2.00	1.82
10%	2.62	2.70	2.28	1.94	1.67
25%	1.96	1.89	1.77	1.55	1.47
50%	1.48	1.39	1.35	1.31	1.28
75%	1.15	1.150	1.12	1.14	1.14

$R_0=1.5$	5% Cohorting	10%	25%	50%	75%
5% Distancing	0.19	0.16	0.10	0.01	4.66
10%	0.15	0.12	0.06	0.002	3.88
25%	0.03	0.02	0.003	3.65	2.28
50%	2.34	2.03	1.92	1.66	1.62
75%	1.36	1.30	1.32	1.28	1.22

$R_0=5$	5% Cohorting	10%	25%	50%	75%
5% Distancing	0.81	0.83	0.81	0.80	0.77
10%	0.79	0.78	0.80	0.78	0.75
25%	0.75	0.74	0.72	0.73	0.70
50%	0.54	0.53	0.50	0.47	0.42
75%	0.09	0.08	0.03	0.001	3.41

6.4 Discussion

The results from this study show that occupation-targeted strategies for infection control are often equally as effective as general hospital control measures, and require fewer resources, particularly for the medium R_0 illness.

For the medium R_0 illness, targeting nurses for mask-wearing is equally as effective in preventing an outbreak compared to targeting the entire hospital, assuming the masks can prevent transmission by at least 50%. This is a reduction of 1,400 individuals when targeting only nurses, leading to more concentrated resources and supply costs.

If random vaccination for the medium R_0 illness achieves 75% immunity levels, it is equally as effective to target nurses ($n=1405$) as the entire hospital ($n=2805$), which means a savings of approximately 1,000 vaccines. When prioritizing HCW vaccination based on their number of contacts, only 25% of nurses need to be vaccinated to prevent epidemic spread, which is equally as effective as targeting the entire hospital at 25%. Although vaccinating 25% vs. 75% of the population is cost effective, gathering the data necessary to prioritize individuals based on contact rate is very resource intensive.

Occupation targeted strategies were not effective for the high R_0 illness. Masks must achieve a 90% reduction in infection transmission, and be worn by the entire hospital, to prevent an epidemic. Additionally, both random and contact-targeted vaccination strategies must achieve 75% and 50% hospital-wide immunity, respectively, to prevent epidemic spread. Although vaccination for many high R_0 diseases, such as measles, is often a condition of employment for hospital personnel, these results demonstrate the control challenges an emerging and highly infectious disease may pose in a hospital setting.

Social network interventions of social distancing and staff cohorting were effectively simulated. When implemented separately, high levels of compliance for both measures were required to prevent epidemic spread of the medium R_0 illness, and the interventions

were not effective in preventing an epidemic for the high R_0 illness. When used in combination, more feasible levels of 50% cohorting and 25% distancing were sufficient to prevent epidemic spread of the medium R_0 illness.

Distancing of 25% may be achieved by measures such as cancelling face-to-face meetings, limiting cafeteria and coffee shop visits, and discouraging causal social contact in the face of an outbreak. Staff cohorting at 50% may be achieved by altering staffing procedures to prevent HCWs from working in multiple wards.

Regardless of intervention strategy, the low R_0 illness did not spread to epidemic levels in the hospital. This is due to the below-threshold R_0 . Although local spread of low R_0 diseases such as SARS is possible, as demonstrated in outbreaks in Toronto and Hong Kong (Varia et al., 2003, Voirin et al., 2009), illnesses with $R_0 < 1$ are incapable of spreading to a significant proportion of individuals in the hospital network. Additionally, many of the HCW cases of SARS were directly acquired from patients, a factor not addressed in the current simulations.

Given challenges to vaccination compliance, prioritizing individuals by occupation or degree may lead to more effective control strategies. Further research on how to efficiently target highly connected individuals for intervention strategies is needed.

7. Conclusion

The results of this thesis build upon previous work exploring the contact patterns of healthcare workers in hospital settings, and propose a novel approach to creating a hospital-wide contact network. Additionally, control strategies including social distancing measures, not previously simulated in the hospital setting, were explored. The concluding chapter of this thesis will review the principal findings in the context of current literature, discuss strengths and limitations, propose applications of the findings, and suggest future work.

7.1 A Tool for Network Modeling

Contact networks in the hospital setting are a challenge to capture. Direct observation and sensor technologies have been effective in measuring face-to-face contact of healthcare workers in the hospital setting, but have only been used within single units or wards. Attempts to create hospital-wide networks have utilized electronic medical records (Liljeros et al., 2007; Ueno & Masuda, 2008; Cusumano-Towner et al., 2013; Curtis et al., 2013), but fail to capture off-the-record interactions such as social contact and staff meetings.

The results of this research successfully created a tool for network generation in a hospital setting. Measures of *degree* and *association* are utilized as inputs, allowing for the capturing of all social and work-related interactions, as well as differences in mixing between occupations in the hospital. Error rates were low for the generation process, even in the extreme scenarios used in Chapter 3.

The input parameters of *degree* and *association* are commonplace in literature involving social networks and contact patterns. For example, Isella *et al.* (2010) used sensor networks to measure both degree distribution and mixing patterns among the ward. These metrics can be directly used as input in the generation algorithm. Thus, this tool can

easily be applied to create networks of other hospitals, and potentially, other communities of interests.

7.2 Revealing Patterns of Contact and Movement in a Large Canadian Hospital

Self-reported survey data from a large Canadian hospital were analyzed, with questions regarding HCW contact rate and hospital movement patterns.

It was found that physicians had significantly less HCW contact than nurses and other HCWs, where contact from the survey was defined as within one meter for two minutes or more. This agrees with a study by Isella *et al.* (2011), which found physicians to have the least number of contacts of the occupations surveyed, although measurements were only within a pediatric ward. Contact in this study was defined as within 1.5 meters for 20 seconds or more. A contrasting study by Polgreen *et al.* (2010) found that nurses, resident physicians and fellows had the highest number of HCW contacts of the job categories observed. With this study, contact was defined as within 0.9 meters, but had no minimum time component. This suggests that physicians may have more frequent short interactions, and less contact of extended periods. In another study by Curtis *et al.* (2013), resident physicians and nurses were found to be the most frequently in the top 10% of contacts, again conflicting with our findings, though these assumptions of contact were based on movement patterns from electronic medical records, and fail to capture casual social contacts.

Nurses visited significantly less locations in a typical week, but had an overall elevated contact rate, compared to other occupations. Although there is no current literature on HCW movement patterns to compare with, this finding is intuitive: nurses are often assigned to a single ward or unit for patient care. The “other” HCW category may be the most significant for superspreading events, having a high mobility throughout the hospital, as well as a high HCW contact rate. Occupations that fall in this category include respiratory therapists and patient care attendants.

The degree distributions for each occupation demonstrated a heavy-tailed distribution, with a small number of HCWs accounting for a large proportion of contacts. A similar distribution was found in a medical intensive care unit (Hornbeck et al., 2012). The heavy-tailed distribution of HCW contacts was again observed in hospital-wide networks by Vanhems *et al.* (2013), and Curtis *et al.* (2013). The distribution of the hospital setting implies that a select few HCWs are responsible for many of the contacts, and thus transmission within hospitals. Our results indicate that these “super-spreaders” may occur in any occupation, thus methods of identifying these individuals for prioritized vaccination or targeted control strategies would be effective.

Location analyses showed that public spaces were visited the most frequently per week, including the cafeteria, lobby café, and coffee shops, although for a shorter length of time. Although again intuitive, this finding highlights a potential vulnerability of hospitals: given the vast overlap of HCWs at these locations, these locations could be significant facilitators of disease spread between otherwise unconnected wards or units. Approaching these high-traffic areas as sites for interventions such as hand-washing stations or mask distribution may be effective, as a large and diverse subset of the hospital population would be reached. Additionally, the interaction of HCWs with the general public in these spaces is likely, revealing a potential vulnerability for the propagation of infections circulating within hospitals to the surrounding community.

Finally, it was found that inpatient locations have a greater number of visits per week compared to outpatient locations, indicating there may be a higher risk of disease spread in inpatient settings compared to outpatient settings. During outbreak scenarios, staff cohorting in inpatient areas can reduce the number of visits per location, per week, reducing the potential for enhanced disease propagation by HCWs who work in inpatient settings.

7.3 Evaluating Control Strategies

Control strategies were evaluated on the Canadian hospital network created from survey data, and diseases with three contrasting R_0 s were compared. R_0 s and corresponding transmissibility values were assigned to approximate SARS-like illness ($R_0=0.9$) (Meyers et al., 2005), influenza-like illness ($R_0=1.5$) (Fraser et al., 2009), and measles-like illness ($R_0=5$) (Mossong & Muller, 2000). Considering hospital-wide, occupation-based, and degree-based intervention strategies, the most resource efficient and effective strategies were identified.

It was found that degree-based vaccination strategies may be more effective than random strategies for both the medium and high R_0 illnesses. These findings complement previous work involving hospital vaccination strategies. Polgreen *et al.* (2010) tested a series of vaccination strategies: random vaccination, degree-based vaccination, as well as vaccination of the occupation with the highest degree. Similar to our findings, the degree-based strategy was the most effective. Curtis *et al.* (2013) compared both mobility-based and degree-based vaccination strategies with a random vaccination strategy, finding again the degree-based strategy to be most effective.

Our conclusions align with current literature in that, although degree-based strategies are the most effective, it isn't currently feasible to gather these data on an individual level. A more realistic approach may be targeting a highly connected occupation, such as nurses, which was found in our study to be effective for the medium R_0 illness, and supported through simulations by Polgreen *et al.* (2010). Another approach not tested in our simulations, but found to be effective and feasible (Curtis et al., 2013), is targeting HCWs with the most mobility, rather than degree.

Transmission intervention simulations yielded findings that suggest targeting the most connected occupations leads to similar efficacy with a reduction in resource use. These strategies were not effective for the high R_0 illness. Hornbeck *et al.* (2012) simulated the

effects of hand-washing noncompliance in a hospital setting, but literature has not yet tested transmission-reduction measures in targeted populations.

Social network interventions of social distancing and staff cohorting were also simulated. Levels of 50% cohorting and 25% distancing, used in combination, were required to prevent the epidemic spread of the medium R_0 illness. Social distancing measures were not sufficient to prevent the epidemic spread of the high R_0 illness. Distancing of 25% may be achieved through measures such as cancelling face-to-face meetings, limiting cafeteria and coffee shop visits, and discouraging causal social contact in the face of an outbreak. Staff cohorting at 50% may be achieved by altering staffing procedures to prevent HCWs from working in multiple wards.

7.4 Strengths and Limitations

The network generation algorithm is a strength of this study, in that it is able to capture the diversity in contact rates both within and between occupations, while remaining flexible in its applications. A previous studies by Polgreen *et al.* (2010) was flexible in its ability to use a range of association values for interactions between occupations, but was unable to capture diversity in degree distribution within occupations, only between. Other studies included intra-occupational degree diversity, but these studies were taken directly from observational or sensor data, which makes them unable to be applied to other hospital settings (Hornbeck *et al.*, 2012; Isella *et al.* 2011; Vanhems *et al.*, 2013). The strength of the current study is that we are able to capture the intra-occupational variation in degree distribution, while remaining flexible in the application to other hospitals. Another strength of this study was its attempts to simulate social distancing transmission reduction interventions, which haven't previously been simulated in the hospital setting.

A primary limitation of this study is that patients are not included in the current network. Although patients aren't included in some recently developed hospital networks (Curtis *et al.*, 2013), the inclusion of patients would further refine the patterns of transmission in the hospital setting, particularly because infections often begin with patients in these settings.

Secondly, because the rates of contact between HCWs are self-reported, there is a potential for a non-differential reporting bias in the results of this study. Because healthcare workers have long shifts, with many different types of contacts, they may be more likely to underreport their contacts. This would lead to a sparser graph, and consequently different conclusions regarding intervention efficacies. Confirming the findings of contact patterns through a hospital-wide RFID or observational study would validate the findings further.

The final limitation identified in this study is the lack of detailed association data between healthcare workers. Because we did not ask in the survey, “*who* do you come into contact with,” we do not know which types of HCWs interact, or if there is assortative mixing. Although previous studies have suggested a non-random mixing of HCW occupations (Isella et al. 2011; Vanhems et al., 2013), the current study was unable to capture this through the survey data analyzed.

7.5 Applications of Findings

These findings may aid in policy decisions for hospital control planning. Particularly, in the face of limited resources, this study suggests that targeting particular subgroups of the hospital setting is just as efficient as targeting the entire hospital. Additionally, it suggests that altering the usage of public spaces, or targeting these as intervention sites, may be most effective. The flexibility of this model means that it can be tailored to address a specific disease or scenario in a question, making it a potentially valuable tool in policy decision making.

The multitype network tool developed in this study can be applied to capture contact networks in many different populations. This can include other hospitals, as well as other communities. This tool could be applied to other hospitals in an efficient way, given the availability of input data of *association* and *degree*. Additionally, the tool developed may be useful when applied to sexual networks, for example, where different subgroups may

represent different sexual orientations, with distinct degree and association components for each.

7.6 Future Work

Future work includes adding patients to the model. As explained previously, this would increase the simulation's accuracy and account for all types of HCW infection.

Additionally, as identified in Chapter 5, a large subset of the HCW population spends time in public hospital spaces. Utilizing this intersection between the community and hospital as a way of modeling the community as a whole would be interesting future work.

Finally, refining the occupations further would lead to a more detailed network. For example, this would mean dividing the "Other HCW" category into respiratory therapists, physiotherapists, social workers, etc. The current diversity in contact rates within occupations may be explained by differences between more fine-grained subgroups within these occupation categories. Further, this may reveal particular occupations within the current categories that contain the majority of superspreaders, which would provide a more targeted intervention population for degree based control strategies.

References

- Allard A, Noël PA, Dubé LJ, Pourbohloul B. Heterogeneous bond percolation on multitype networks with an application to epidemic dynamics. *Phys Rev E Stat Nonlin Soft Matter Phys.* 2009 Mar;79(3 Pt 2):036113.
- Anderson RM, May RM *Infectious Diseases of Humans: Dynamics and Control.* Oxford: Oxford University Press 1991.
- Assiri A, McGeer A, Perl TM, Price CS, Al Rabeeah AA, Cummings DA, Alabdullatif ZN, Assad M, Almulhim A, Makhdoom H, Madani H, Alhakeem R, Al Tawfiq JA, Cotten M, Watson SJ, Kellam P, Zumla AI, Memish ZA; KSA MERS-CoV Investigation Team. Hospital outbreak of Middle East respiratory syndrome coronavirus. *N Engl J Med.* 2013 Aug 1;369(5):407-16.
- Bansal S, Pourbohloul B, Meyers LA. A comparative analysis of influenza vaccination programs. *PLoS Med.* 2006 Oct;3(10):e387.
- Bernard H, Fischer R, Mikolajczyk RT, Kretzschmar M, Wildner M. Nurses' contacts and potential for infectious disease transmission. *Emerg Infect Dis* 2009 15: 1438–1444.
- Bischoff WE, Swett K, Leng I, Peters TR. Exposure to influenza virus aerosols during routine patient care. *J Infect Dis.* 2013 Apr;207(7):1037-46.
- Campbell A. The SARS Commission: Spring of Fear, Final Report. 2006 Dec
<<http://www.health.gov.on.ca/en/common/ministry/publications/reports/campbell06/Vol1Chp1.pdf>>.
- Chironna M, Tafuri S, Santoro N, Prato R, Quarto M, Germinario CA. A nosocomial outbreak of 2009 pandemic influenza A(H1N1) in a paediatric oncology ward in Italy, October-November 2009. *Euro Surveill.* 2010 Jan 7;15(1).

Chowell G, Miller MA, Viboud C. Seasonal influenza in the United States, France and Australia: transmission and prospects for control. *Epidemiol Infect.* 2007;136:852–864.

Curtis DE, Hlady CS, Kanade G, Pemmaraju SV, Polgreen PM, et al. Healthcare Worker Contact Networks and the Prevention of Hospital-Acquired Infections. *PLoS ONE* 2013 8(12): e79906.

Cusumano-Towner M, Li DY, Tuo S, Krishnan G, Maslove DM. A social network of hospital acquired infection built from electronic medical record data. *J Am Med Inform Assoc.* 2013 May 1;20(3):427-34.

Dietz K. The estimation of the basic reproduction number for infectious diseases. *Stat Methods Med Res.* 1993;2(1):23-41.

Fraser C, Donnelly CA, Cauchemez S, Hanage WP, Van Kerkhove MD, Hollingsworth TD, Griffin J, Baggaley RF, Jenkins HE, Lyons EJ, Jombart T, Hinsley WR, Grassly NC, Balloux F, Ghani AC, Ferguson NM, Rambaut A, Pybus OG, Lopez-Gatell H, Alpuche-Aranda CM, Chapela IB, Zavala EP, Guevara DM, Checchi F, Garcia E, Hugonnet S, Roth C; WHO Rapid Pandemic Assessment Collaboration. Pandemic potential of a strain of influenza A (H1N1): early findings. *Science.* 2009 Jun 19;324(5934):1557-61.

Hearnshaw EJ, Wilson MM. A complex network approach to supply chain network theory. *International Journal of Operations & Production Management.* 2013;Vol. 33 Iss: 4, pp.442 – 469.

Hornbeck T, Naylor D, Segre AM, Thomas G, Herman T, Polgreen PM. Using sensor networks to study the effect of peripatetic healthcare workers on the spread of hospital-associated infections. *J Infect Dis.* 2012 Nov 15;206(10):1549-57.

Isella L, Romano M, Barrat A, Cattuto C, Colizza V, et al. Close Encounters in a Pediatric Ward: Measuring Face-to-Face Proximity and Mixing Patterns with Wearable Sensors. PLoS ONE 2011 6(2): e17144.

Jolly AM, Muth SQ, Wylie JL, Potterat JJ. Sexual networks and sexually transmitted infections: a tale of two cities. J Urban Health. 2001 Sep;78(3):433-45.

Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P. Global trends in emerging infectious diseases. Nature 2008 451:990-994.

Keeling M. The implications of network structure for epidemic dynamics. Theoretical Population Biology 2005 67(1):1-8.

Lee N, Hui D, Wu A, Chan P, Cameron P, Joynt GM, Ahuja A, Yung MY, Leung CB, To KF, Lui SF, Szeto CC, Chung S, Sung JJ. A major outbreak of severe acute respiratory syndrome in Hong Kong. N Engl J Med. 2003 May 15;348(20):1986-94.

Liljeros F, Edling CR, Amaral LA, Stanley HE, Aberg Y. The web of human sexual contacts. Nature. 2001 Jun 21;411(6840):907-8.

Liljeros F, Giesecke J & Holme P (2007) The Contact Network of Inpatients in a Regional Healthcare System. A Longitudinal Case Study, Mathematical Population Studies: An International Journal of Mathematical Demography, 14:4, 269-284.

Merrill J, Caldwell M, Rockoff ML, Gebbie K, Carley KM, Bakken S. Findings from an organizational network analysis to support local public health management. J Urban Health. 2008 Jul;85(4):572-84.

Meyers L. Contact network epidemiology: Bond percolation applied to infectious disease prediction and control. Bulletin of the American Mathematical Society 2007 44.1: 63-86.

Meyers LA, Newman ME, Martin M, Schrag S Applying network theory to epidemics: control measures for *Mycoplasma pneumoniae* outbreaks. *Emerg Infect Dis*. 2003 9(2):204-10.

Meyers LA, Pourbohloul B, Newman ME, Skowronski DM, Brunham RC. Network theory and SARS: predicting outbreak diversity. *J Theor Biol*. 2005 Jan 7;232(1):71-81.

Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, Massari M, Salmaso S, Tomba GS, Wallinga J, Heijne J, Sadkowska-Todys M, Rosinska M, Edmunds WJ. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med*. 2008 Mar 25;5(3):e74.

Mossong J, Muller CP. Estimation of the basic reproduction number of measles during an outbreak in a partially vaccinated population. *Epidemiol Infect*. 2000 Apr;124(2):273-8.

Newman ME. Spread of epidemic disease on networks. *Phys Rev E Stat Nonlin Soft Matter Phys*. 2002 Jul;66(1 Pt 2):016128.

Newman, M. E. J. Mixing patterns in networks. 2003 *Phys. Rev. E* 67, 026126.

Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free networks. *Phys Rev Lett*. 2001 Apr 2;86(14):3200-3.

Polgreen PM, Tassier TL, Pemmaraju SV, Segre AM. Prioritizing healthcare worker vaccinations on the basis of social network analysis. *Infect Control Hosp Epidemiol*. 2010 Sep;31(9):893-900.

Poutanen SM, Low DE, Henry B, Finkelstein S, Rose D, Green K, Tellier R, Draker R, Adachi D, Ayers M, Chan AK, Skowronski DM, Salit I, Simor AE, Slutsky AS, Doyle PW, Krajden M, Petric M, Brunham RC, McGeer AJ; National Microbiology Laboratory, Canada; Canadian Severe Acute Respiratory Syndrome Study Team. Identification of

severe acute respiratory syndrome in Canada. *N Engl J Med*. 2003 May 15;348(20):1995-2005.

Public Health Agency of Canada. Routine practices and additional precautions for preventing the transmission of infection in healthcare settings. 2013 <http://www.ipac-canada.org/pdf/2013_PHAC_RPAP-EN.pdf>.

Qian Y, Willeke K, Grinshpun SA, Donnelly J, Coffey CC. Performance of N95 respirators: filtration efficiency for airborne microbial and inert particles. *Am Ind Hyg Assoc J*. 1998 Feb;59(2):128-32.

Roth DZ, Henry B. Social Distancing as a pandemic influenza prevention measure. 2011 <https://cdn.metricmarketing.ca/www.nccid.ca/files/Evidence_Reviews/H1N1_3_final.pdf>.

Salgado CD, Giannetta ET, Hayden FG, Farr BM. Preventing nosocomial influenza by improving the vaccine acceptance rate of clinicians. *Infection control and hospital epidemiology*. 2004. Vol 25: 11 923-928.

Skowronski DM, Petric M, Daly P, Parker RA, Bryce E, Doyle PW, Noble MA, Roscoe DL, Tomblin J, Yang TC, Krajden M, Patrick DM, Pourbohloul B, Goh SH, Bowie WR, Booth TF, Tweed SA, Perry TL, McGeer A, Brunham RC. Coordinated response to SARS, Vancouver, Canada. *Emerg Infect Dis*. 2006 Jan;12(1):155-8.

Ueno T, Masuda N. Controlling nosocomial infection based on structure of hospital social networks. *J Theor Biol*. 2008 Oct 7;254(3):655-66. doi: 10.1016/j.jtbi.2008.07.001.

Vanhems P, Barrat A, Cattuto C, Pinton JF, Khanafer N, Régis C, Kim BA, Comte B, Voirin N. Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PLoS One*. 2013 Sep 11;8(9):e73970.

Varia M, Wilson S, Sarwal S, McGeer A, Gournis E, Galanis E, Henry B; Hospital Outbreak Investigation Team. Investigation of a nosocomial outbreak of severe acute respiratory syndrome (SARS) in Toronto, Canada. CMAJ. 2003 Aug 19;169(4):285-92.

Vazquez A. Spreading dynamics on heterogeneous populations: multitype network approach. Phys Rev E Stat Nonlin Soft Matter Phys. 2006 Dec;74(6 Pt 2):066114.

Voirin N, Barret B, Metzger MH, Vanhems P. Hospital-acquired influenza: a synthesis using the Outbreak Reports and Intervention Studies of Nosocomial Infection (ORION) statement. J Hosp Infect. 2009 Jan;71(1):1-14.

Weber A, Willeke K, Marchioni R, Myojo T, McKay R, Donnelly J, Liebhaber F. Aerosol penetration and leakage characteristics of masks used in the health care industry. Am J Infect Control. 1993 Aug;21(4):167-73.

World Health Organization. Infection and prevention and control of epidemic- and pandemic- prone acute respiratory diseases in health care: WHO Interim Guidelines. 2007

<http://apps.who.int/iris/bitstream/10665/69707/1/WHO_CDS_EPR_2007.6_eng.pdf>.

World Health Organization. Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003. 2003

<http://www.who.int/csr/sars/country/table2004_04_21/en/index.html>.

World Health Organization. Middle East respiratory syndrome coronavirus (MERS-CoV) summary and literature update—as of 11 June 2014. 2014

<http://www.who.int/csr/disease/coronavirus_infections/MERS-CoV_summary_update_20140611.pdf?ua=1>.

Zoutman, D., et al., The state of infection surveillance and control in Canadian acute care hospitals. *Am J Infect Control*, 2003. 31: p. 266-73.

Appendices

Appendix A: Network Generation Code

```
% Network_Generation_Final
% Matlab pseudocode that uses the inputs of degree distribution, population
% sizes, and association parameters to create a network or ensemble of
% networks with error and clustering coefficients

% A
degree=csvread('DegreeList.csv'); %reads CSV of degree distributions

numd=650; %number of doctors
numn=1405; %number of nurses
numhwcw=750; %number of other HCW

doctors= %assigns degree distribution of doctors
nurses= %assigns degree distribution of nures
hwcw= %assigns degree distribution of
%other HCW

% B
adj=zeros(length(degree)); %creates adjacency matrix%subtract stubs from
%each individual

% C
% association parameters
pnn=0.4;
pnd=0.2;
pno=0.3;
pdn=0.3;
pdd=0.05;
pdo=0.75;
pon=0.4;
pod=(1/3);
poo=(1/3);

%calculate assigned stubs of each connection type
dd=round(sum(doctors)*pdd);
nn=round(sum(nurses)*pnn);
on=round(sum(hwcw)*pon);
no=on;
nd=sum(nurses)-nn-no;
dn=nd;
do=(sum(doctors)-dn-dd);
od=do;
oo=(sum(hwcw)-od-on);

%D-E
%starting the network generation process.

for i=1:n %n is the number of repetitions, if an ensemble is being created.
while %there are stubs left to connect, unless they have already been
    %connected, loop through this process:

type= randsample(1:6,1); %choose a random connection type (dd, nn, oo, do,
```



```

%dn, or on)

if type==1 & %there are stubs left
    %find nodes that have remaining stubs
    n1=%select a random nurse
    n2=%select random doctor
    if %n1 and n2 have already been connected, restart
    %add entries to adjacency matrix
    %subtract stubs from each individual

elseif type==2 %there are stubs left
    %find nodes that have remaining stubs
    n1=%select a random nurse
    n2=%Select a random other hcw
    if %n1 and n2 have already been connected, restart
    %add entries to adjacency matrix
    %subtract stubs from each individual

elseif type==3 & %there are stubs left
    %find nodes that have remaining stubs
    n1=%select a random dr
    n2=%Select a random other hcw
    if %n1 and n2 have already been connected, restart
    %add entries to adjacency matrix
    %subtract stubs from each individual

elseif type==4 & %there are stubs left
    %find nodes that have remaining stubs
    n1=%select a random doctor
    n2=%select another random doctor
    if %n1 and n2 have already been connected, restart
    %add entries to adjacency matrix
    %subtract stubs from each individual

elseif type==5 & %there are stubs left
    %find nodes that have remaining stubs
    n1=%select a random nurse
    n2=%select another random nurse
    if %n1 and n2 have already been connected, restart
    %add entries to adjacency matrix
    %subtract stubs from each individual

elseif type==6 & %there are stubs left
    %find nodes that have remaining stubs
    n1= %Select a random other hcw
    n2= %Select a random other hcw
    if %n1 and n2 have already been connected, restart
    %add entries to adjacency matrix
    %subtract stubs from each individual
end

```

```
end

%save the ensemble of networks
%calculate error
%calculate clustering coefficient
end
```

Appendix B: Simulation Code

```
% Simulation_Code_Final
% Matlab pseudocode that uses the inputs of adjacency matrix, population
% size, beta, disease duration, and initial case to run disease simulations
% on the network.

%A
adj=csvread('adjacencymatrix.csv'); %reads adjacency matrix created in
%network generation process

numd=650; %number of doctors
numn=1405; %number of nurses
numhcw=750; % number of other HCWs

beta=0.01;
Duration=6;

for %all individuals in the population
    neighbors=%neighbors are extracted from adjacency matrixx;
end

%B
Init_case=%randomly select a first to infect

%C
while %there are infected individuals in the populatoin
    for %each infected node
        %select neighbors and infect at probability Beta

        newrecovereds= %find those who have been infected for 6 timesteps
            %remove them from the infected pool and add to recoverds
    end
end

attack(q)=Recovered(t);
probepid(o)=length(find(attack>200))
```