MODELING EPIDEMICS ON NETWORKS OF CONNECTED COMMUNITIES

by

Isaac Gilbert Freedman

Submitted to the Undergraduate Faculty of

Arts and Sciences in partial fulfillment

of the requirements for the degree of

BPhil in Physics

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

Dietrich School of Arts and Sciences

This thesis was presented

by

Isaac G. Freedman

It was defended on

April 9, 2015

and approved by

Daniel Boyanovsky, PhD, Professor, Department of Physics and Astronomy

James A. Mueller PhD, Associate Professor, Department of Physics and Astronomy

John J. Grefenstette PhD, Professor Department of Health Policy and Management

Visitng Reviewer: Murat Demirbas PhD, Associate Professor, University at Buffalo, Department of Computer Science and Engineering

Thesis Advisor: Hasan Guclu, Assistant Professor, Department of Health Policy and Management

Modeling Epidemics on Networks of Connected Communities

Isaac G. Freedman, BPhil

University of Pittsburgh, 2015

Advances in the fields of mathematics, physics, epidemiology, and computing have led to an incredibly productive period of epidemic modeling. Here I will present the findings of several computational studies aimed at understanding how epidemics spread across networks. I investigate specifically how epidemics spread across networks consisting of two weakly connected sub-networks (communities) with varying internal connectivities, vaccination probabilities, and probabilities of social distancing. I find that, on average, epidemics may spread across communities even for a single cross connection, that crossing over is characterized by multiple time delayed epidemic waves that result in increased epidemic duration. I develop a novel mathematical characterization of networks consisting of an arbitrary number of weakly connected communities and derive a relationship between the reproductive number ($\mathcal{R}_0$) of an epidemic and the Mean Squared Displacement (MSD) of the epidemic, when the spread is viewed as the progression of multiple forward-biased random walkers. Finally, I propose a new compartmental Susceptible Exposed Infected Quarantined Recovered (SEIQR) model for the 2014 Ebola Virus Disease (EVD) outbreak based on differential equations. I extend this model to an immigration SEIQR (iSEIQR) model with a constant rate of immigration and demonstrate homologous behavior in the form of multiple infection waves between a dynamic single community network model with a constant immigration of possible exposed individuals and the

two community models discussed elsewhere in this work. The applications of two community network models are discussed, especially in the context of understanding and mitigating regional and transnational epidemic spread. Pharmaceutical and non-pharmaceutical interventions, such as targeted vaccination, public health education (i.e. avoidance), quarantine, and travel restrictions are explored and some mathematical and physical applications of modeling weakly coupled sub-networks are described. Finally, several possible extensions to this work are listed and discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE AND ACKNOWLEDGMENTS

I would like to thank my thesis advisor Dr. Hasan Guclu and my entire BPhil committee for their time and guidance while writing this thesis, especially Dr. Murat Demirbas for serving as my External Examiner. I would also like to thank Drs. Walter Goldburg and Hanna Salman for several helpful conversations. Finally, I would like to thank Dr. David Hornyak, Karen Billingsley, and the University of Pittsburgh University Honors College for their assistance in organizing the defense of this thesis.

1.0 INTRODUCTION

In the real world, surfaces have friction, projectiles experience air-resistance, and no potential well is ever infinite. As Benoit B. Mandelbrot said in his seminal 1982 work, “Clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line.” [1] Approximate models and simplifying assumptions that smooth out complex and chaotic systems are the foundation of all modern science. However, for practical applications the limitations and validity of these models are of the utmost importance. Models of infectious diseases, their limitations, and regions of validity are the subject of this thesis. I will investigate approximate models for infectious diseases, especially as they spread within and across networks consisting of two communities. I will also propose a mathematical model characterizing how infectious diseases diffuse across heterogeneous networks of an arbitrary number of communities in general, and I will propose several models for studying infectious diseases, both on networks and in the continuous case with differential equations.

1.1 THE KERMACK-MCKENDRICK SIR MODEL

The utility of modeling infectious diseases has been recognized since the days of Isaac Newton and Thomas Malthus. While Malthus proposed one of the first models of population spread, which bears his name, the first popular differential equation model for modeling the spread of

infectious diseases was one of several introduced by Kermack and McKendrick in 1927. In modern formulation, the Kermack-McKendrick Model (henceforth called the KM Model) is a compartment-based, Susceptible Infectious Recovered/Removed (SIR) model consisting of a system of three coupled Ordinary Differential Equations (ODEs), one equation for each compartment (1.1.1), where an overdot denotes a derivative with respect to time [2].

$$\begin{cases} \dot{S} = -\beta SI \\ \dot{I} = \beta SI - \lambda R \\ \dot{R} = \lambda I \end{cases}$$

(1.1.1)

While relatively simple, this model is nevertheless the archetype and forbearer of modern differential equation based infectious disease models. The KM model assumes complete immunity after a single infection, which occurs with a probability proportional to β, and the *R* compartment is treated as the sum of individuals who cease to be infected, either through recovery or death with a probability proportional to λ. The interpretation of the solutions of (1.1.1) is that for a given population, the size of the susceptible population *S* decreases when an individual becomes infected. The probability of infection is proportional to the product of the susceptible and infected populations at a given time, *S* and *I*. This makes sense intuitively, as the number of individuals who become infected should increase both as there are more susceptible individuals to infect and more infected individuals to cause infections. A schematic view of the compartmental flow of the population is shown in *Figure 1*.

Figure 1. Schematic representation of the flow of (1.1.1). Susceptible individuals become Infected individuals with a rate proportional to β and Infected individuals become Recovered/Removed with a rate proportional to λ .

Once an individual is infected, they may become removed with a probability proportional the size of the removed population *R*. In a real epidemic, removed would correspond to any state where an infected individual is no longer infectious–i.e., recovery with permanent immunity or death. However, as will be discussed in later chapters for the case of Ebola, death is not a guarantee that an individual is no longer infectious. A typical outbreak trajectory is shown in *Figure 2*.

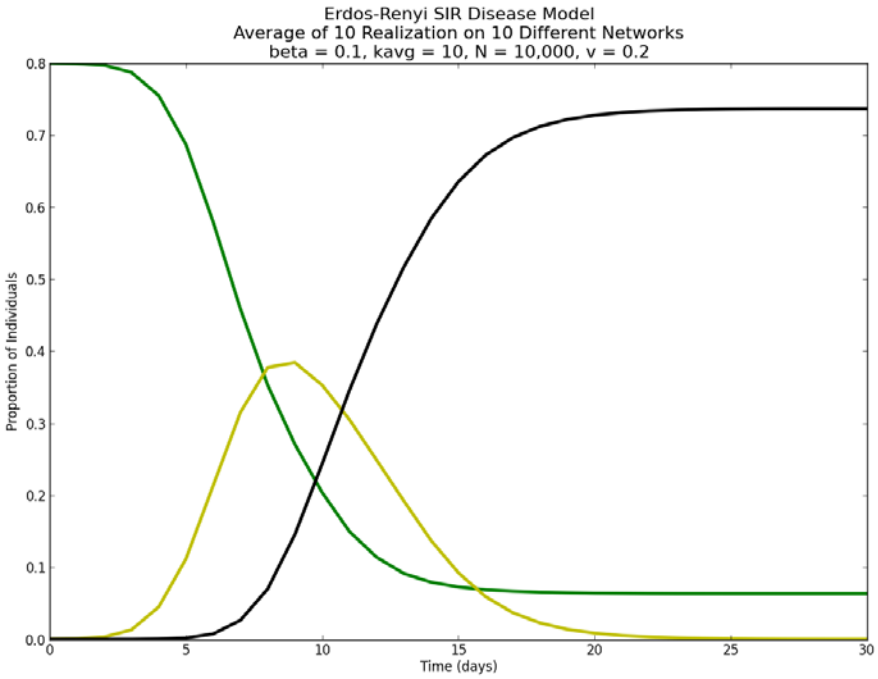

Figure 2. Typical SIR outbreak trajectories for Eq. (1.1.1). All three subpopulation sizes over time, the I-Curve is shown in yellow. The size of the Susceptible and Recovered/Removed populations are shown in green and black, respectively.

Several key assumptions went into forming (1.1.1). While each of these simplifying assumptions allows the KM model to be easily analyzed and understood mathematically, they are also what lead to its limitations as a model of real-world epidemics. The first assumption is that the population can be modeled continuously, i.e., there can and will be a *fractional* number of susceptible, infected, and recovered individuals at each infinitesimal time step. While this results in smooth and continuously differentiable equations, it is not true at the individual scale and only holds for large population number N. In this thesis, I will study primarily network models of epidemics, which contain a discrete number of individuals, so each subpopulation must have an integer number of individuals.

The second assumption of the KM model is each individual is in contact with every other individual for all time. I will refer to this as the *uniform mixing assumption*. In the KM Model,

the uniform mixing assumption provides that every individual is equally likely to be infected by the infectious population, with a rate proportional to β at each time step. In network models, and indeed in nature, an individual can *only* infect individuals in their particular *contact network*, or list of individuals with which they come in contact. The consequences of the uniform mixing assumption are profound and cases where the assumption does not hold will be a primary topic of this thesis. By assuming uniform and total mixing, the KM model ignores the spatial component of epidemic spreads. As epidemics spread out from *patient zero*, or the *index case*, they form population waves [3]. Containing disease outbreaks to a particular region or nation is a primary concern of modern public health and has played a crucial role in international policy and foreign aid for outbreaks such as the 2004 SARS outbreak [4], the 2009 H1N1 pandemic influenza outbreak [5], and the 2015 Ebola Virus Disease (EVD) outbreak [6]. Studying the spatial spread of epidemics across networks and between weakly connected communities will be the experimental focus of this thesis.

Another assumption of the KM model is the simplified disease progression–that is, from susceptible to infected to recovered or removed. This assumption greatly simplifies the biology of many diseases and many extensions to the KM model have been proposed that model different disease progressions. For example, SIS models (Susceptible Infectious Susceptible) have been used to model the spread of sexually transmitted diseases, which can and are often contracted multiple times by individuals in at-risk populations. The SEIR model, which contains an Exposed compartment *E*, will be discussed further in *Chapter 3* as a means to account for a latent period when an individual is infected, but not yet infectious.

The final assumption of the KM model I will discuss is homogeneity of population. Although outbreaks modeled by (1.1.1) are not studied spatially, there can still be heterogeneity

of host population susceptibility and even infectivity. Indeed, the original investigation of Kermack and McKendrick included variable infectivity for different ages. In fact, the KM model as stated in (1.1.1) is the homogenized reduction of a more complicated model proposed by Kermack and McKendrick that took into account the age distribution of the infected population. However, a vast number of extensions of the KM model have been proposed that include multiple populations of infected individuals [7]. One such model is the SEIQR model, which contains a Quarantined compartment *Q* and which will be discussed in further detail in *Chapter 3* in the context of the 2015 EVD outbreak [8].

The KM Model has several important properties common to many other differential equation and network models. First, this model has a conserved quantity, which we will call population size N [9]. This can be seen by relabeling each compartment X_0, X_1, X_2, such that $S = X_0$, $I = X_1$, and $R = X_2$ then from (1.1.1) we can write

$$\sum_i \dot{X}_i = 0 \tag{1.1.2}$$

where the sum is implicitly over all i compartments. Integrating, this leads to

$$\sum_i X_i = N, \tag{1.1.3}$$

where we have labeled the constant of integration N, the total population of the system. We will call (1.1.3) the conservation condition. For conservative systems, *N* will be a conserved quantity,

i.e., a constant for all time, while for non-conservative systems we can write *N* as a function of time such that $N = N(t)$.

Along with being conservative, (1.1.1) is not analytically solvable. This is a trait the KM model has in common with all but a few special cases of differential equation models for epidemics. Although not analytically solvable, the KM model can be analyzed via linearization and other mathematical methods, as well as being numerically solvable. Linearizing (1.1.1) leads to the threshold for outbreak of this SIR model

$$\mathcal{R}_0 = \frac{\beta}{\lambda}$$

(1.1.4)

where it can be shown that for $\mathcal{R}_0 < 1$ no outbreak occurs and for $\mathcal{R}_0 > 1$ an outbreak occurs.

This threshold of the KM SIR Model is identified as one of the most important and thoroughly researched epidemiological quantities, the *basic reproductive number* or *basic reproductive rate*. The basic reproductive number $\mathcal{R}_0$ can be defined as the number of new cases caused by a single infected individual in a virgin (i.e., unexposed) population. Although, $\mathcal{R}_0$ has units of $\frac{individuals}{unit\ time}$, i.e. of a rate, it is more appropriately considered a constant characteristic of a particular outbreak that is only considered over infection durations. In the KM model $\mathcal{R}_0$ is given by (1.1.4), so $\mathcal{R}_0$ serves as a threshold for the epidemic. While $\mathcal{R}_0 > 1$ is a common heuristic for the infectivity threshold an infectious agent must reach to be capable of an outbreak, it is strictly

true only for particular models, such as the KM model. The calculation of $\mathcal{R}_0$ is notoriously difficult for infectious agents and populations that do not obey the simplifying assumptions of the KM model. Some such calculations particularly on networks containing one or more connected communities will be the topic of much of *Chapter 2*. For real-world epidemics, $\mathcal{R}_0$ values are often calculated using various statistical methods. Several $\mathcal{R}_0$ estimations for recent and historical epidemics are given in *Table 1* and range from $\mathcal{R}_0 = 1.5 - 2.5$ for the case of the 2014 EVD outbreak to $\mathcal{R}_0 = 12 - 18$ for airborne Measles.

Table 1. Basic Reproductive Numbers for recent and historical outbreaks.

Disease	$\mathcal{R}_0$ **Range** $\left(\frac{individuals}{unit\ time}\right)$
Measles	$12-18$ [10]
Smallpox	$5-7$ [10]
Polio	$5-7$ [10]
HIV/AIDS	$2-5$ [10]
SARS	$2-5$ [11]
Influenza (1918 Pandemic Strain)	$2-3$ [12]
Ebola Virus (2014 Outbreak)	$1.5-2.5$ [13]

Finally, one trait common to analyzing the KM model and other compartmental models, be they differential equation models or discrete network models, is the relevant quantities measured. For this thesis, the Infected subpopulation will be reported throughout the outbreak (***Fig. 2***) and will be referred to as the *Infection Curve* or *I-Curve,* but neither the Susceptible nor Recovered/Removed populations will be explicitly shown or analyzed for most cases. This parsimonious plotting is the result of two factors: 1) the most relevant quantity for policy makers and researchers alike is the *strain* on the healthcare system caused by the infection. This strain is related almost entirely to the number of afflicted individuals, rather than to the number of healthy

or removed individuals. And, 2) in the vast majority of compartmental models, the size of each subpopulation depends on the size of all of the other subpopulations, so plotting a single subpopulation–i.e., Infected individuals–encodes information about the individuals who are Susceptible and Recovered/Removed, provided information about the total population is given. A final useful measureable quantity that should be discussed is the proportion of individuals who were ever infected, or conversely who are recovered ($R(t \to \infty)$), at the end of an epidemic. This is called the *Attack Rate* (*AR*) of an epidemic and is a measure of the extent to which the epidemic affected the population.

1.2 NETWORK MODELS OF INFECTIOUS DISEASES

Networks analysis is a growing and powerful field that has helped advance mathematics, physics, biology, medicine, and many other disciplines. In this section, I will describe the basic properties of networks, especially those relevant to this investigation. I will also discuss how the KM model (**1.1.1**) and other infectious disease models can be adapted to networks and some benefits of doing so.

Networks, also known as *Graphs* in mathematics, are structures consisting of two objects: *edges* and *nodes* [14]. The *degree* or *connectivity* of a node is the number of edges attached to that node and for *simple graphs*, every edge attaches exactly two nodes, precluding *self-attachements*. A network is normally indicated by the number of edges and nodes it contains, viz. $G(E,V)$ where E is the set of edges and V is the set of *vertices* or nodes. Graphs can be

directed, where the edges have a given direction. An example of a directed network is a *transmission network*, where each node represents an infected individual; the *tail* of each edge points to the infector and the *head* of the edge points to the newly infected individual. Graphs can also be *undirected*, where the edges are symmetric and indicate mutual connection. Two examples of an undirected network are a *social network* and a *contact network*. In both cases each node represents a person and each edge represents mutual social or physical contact, respectively.

In this thesis, several graph *structures* or *topologies* will be used to understand the behavior of epidemics on real-world contact networks. Here, I will present those structures used throughout this work and discuss some of their limitations and advantages. *Erdős-Rényi* (ER) networks, named for Paul Erdős and Alfréd Rényi, are a class of *random graph* where a set of edges is distrubted amongst a set of nodes by randomly and independently picking to which nodes to attach. This random picking results in a Poissonian distribution of number of nodes with a given degree and the distribution tends towards a Gaussian ($Fig. 3a$) as $|V| \to \infty$. Erős-Rényi networks is one type of *random* network and will be used as a control group that should closely model well-mixed populations.

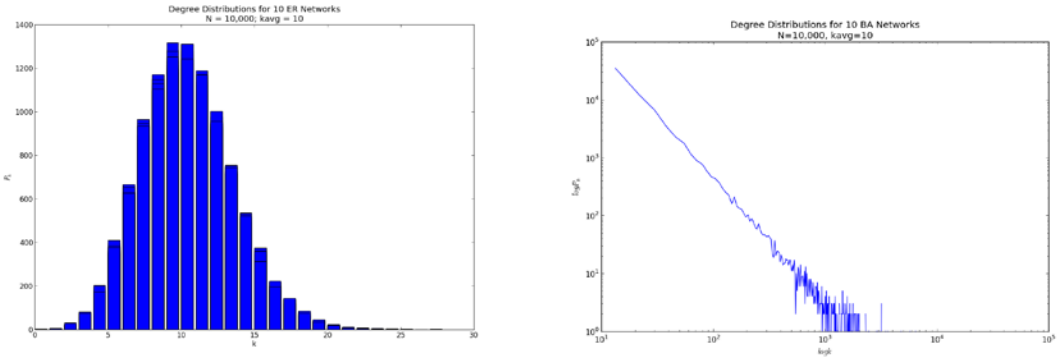

Figure 3. Degree distributions for two network topologies. a) Erős-Rényi (ER) random networks. A Poissonian distribution. b) Barabási-Albert (BA) random networks. An approximately linear fit when plotted on a loglog scale.

Barabási-Albert (BA) networks are a class of *scale-free* networks, which have a degree distribution largely independent of the individual network's size, i.e., a scale-free degree distribution which is logarithmic in P_k, the number of individuals of a given degree k ($Fig. 3b$).

Several real-world networks have been shown to exhibit scale-free properties, including the internet [15] and some social networks. Networks closely modeled by BA networks contain a few *hubs* or very well connected individuals, while the vast majority of individuals have only a few connections. It has been shown that scale-free networks are extremely robust to random *attacks*, cutting individual edges or vaccinating individuals, but are suceptible to attacks that target the hubs. The policy implication of this finding is that targeted interventions, such as vaccination, that focus on hubs, such as healthcare workers, can be much more efficient than random interventions.

The spread of epidemics on both random (ER) and scale-free (BA) networks will be studied. In $Chapter\ 2$, epidemic spread on single community networks will be discussed to give a point of reference for more complex networks. Later in $Chapter\ 2$, epidemic spread on

multiple communities, as well as on vaccinated communities and communities with avoidance behaviors will be discussed. *Chapter 3* will discuss characterizing networks consisting of multiple communities as well as calculating basic reproductive numbers on networks with various community structures. Finally, in *Chapter 4* a novel SEIQR differential equation model for Ebola Virus Disease (EVD) will be presented, discretized, and applied to networks. Immigration will then be accounted for as a linearly increasing population.

2.0 TWO COMMUNITY SYSTEMS

Epidemic spread on populations consisting of multiple communities of individuals is well studied. Indeed, the Kermack-McKendrick model of 1927 was initially proposed to study a population of individuals consisting of many different age groups, each of which responded differently to the infection, for example by making β a function of age [2]. Such systems on continuous epidemic models are common in the literature and closely model many real-world diseases [7]. However, comparably little research has been published on the subject of *networks* consisting of two heterogeneous communities [16]. Of the sparse literature that has been published, nearly none has investigated both random and scale-free networks along with the SIR extensions of vaccination or asymptomaticity and social distancing ("staying home"). Both of these cases are relevant to public health policy and real-world epidemics as well as interesting to the physical study of networks. One important example of differential vaccination across communities that may be weakly interacting are differential vaccination rates in different racial groups as well as, in keeping with the tradition of Kermack and McKendrick, in different age groups [17-19].

In this section, I will describe several novel results for an epidemic spreading on a single network and on a network containing two complex communities with varying average degree, size, and even vaccination rates. In the future, these results could be useful to understanding the

spread of epidemics between different age groups, racial groups, or even international populations. This last case will be further discussed in *Chapter 4*.

2.1 SINGLE COMMUNITY NETWORKS

Prior to investigating two community systems, the spread of an epidemic on a single community was examined. The literature is replete with examples of epidemics spreading on single communities. Indeed, in the study of complex networks the multiple-community perspective is seldom taken, so much of the literature exclusively relates to single community networks. In this section (*Ch. 2.1*), I will describe the results for several computational experiments carried out on so-called single community networks with either random (ER) or scale-free (BA) construction. These experiments are meant to elucidate the behavior of individual communities, to better understand how the two community systems discussed in the next section behave.

To begin, a baseline value for the average connectivity or degree $\langle k \rangle$ was found for both random (ER) and scale-free (BA) network structures. Note that unless otherwise stated, in this work the brackets $\langle \dots \rangle$ indicate the average value of some quantity over an ensemble. The infection parameters were $\beta = 0.1$ and the infection duration or average period $p = 4\ days$. *Figure 4* shows the infection curves and attack rates for four representative values of $\langle k \rangle$ on random (ER) networks. The infection curves for $\langle k \rangle = 10$ were most similar to those found for

SIR models, so this value was chosen for the value of $\langle k \rangle$ that ensured greatest correspondence between the network and continuous SIR models. Similar results were found for scale-free networks, with the same value of $\langle k \rangle = 10$. For the remainder of this thesis, the value $\langle k \rangle = 10$ will be used for all networks unless otherwise stated.

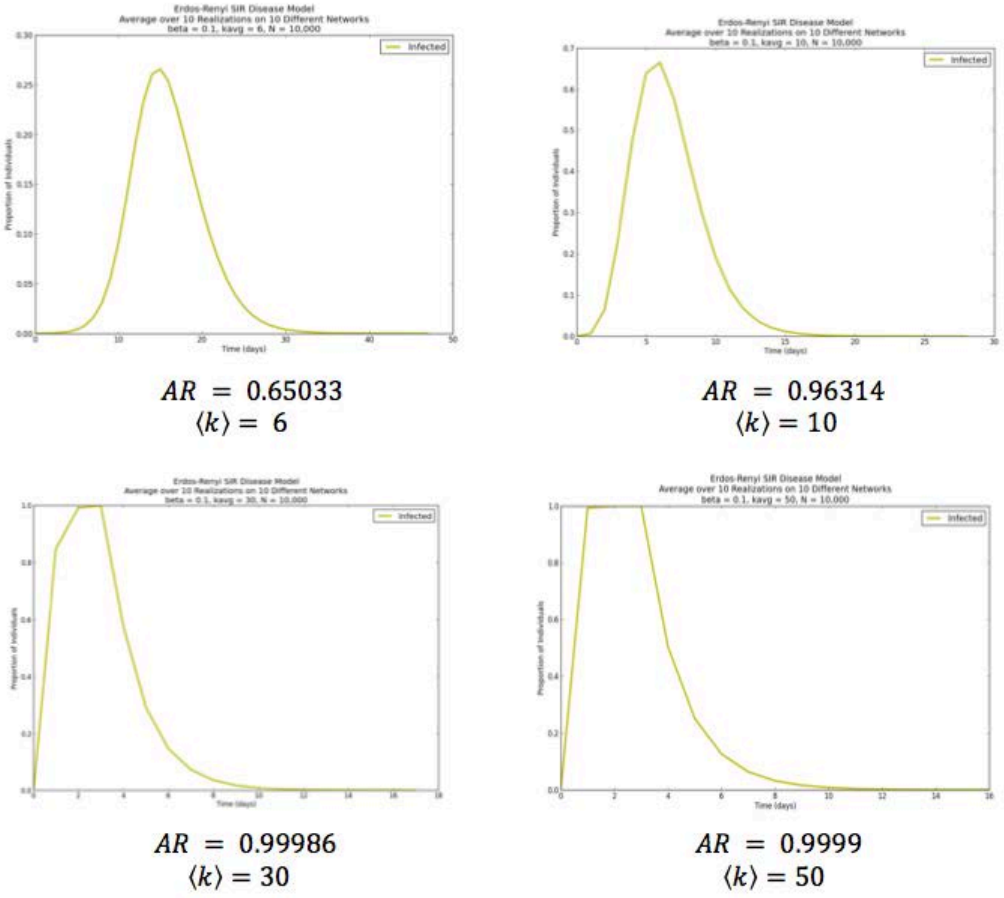

Figure 4. Infection curves on random (ER) networks for various values of $\langle k \rangle$. The infection curve for $\langle k \rangle = 10$ most resembles that of continuous SIR models.

A second conclusion was drawn for the case of a network containing a single community. The *transmission network*, or the directed network containing individuals who were infected

along with information about who infected whom was studied. *Figure 5* shows how the degree distribution for the transmission network on a random contact network closely resembles the structure of its parent contact network. An identical result was found for the case of scale-free networks.

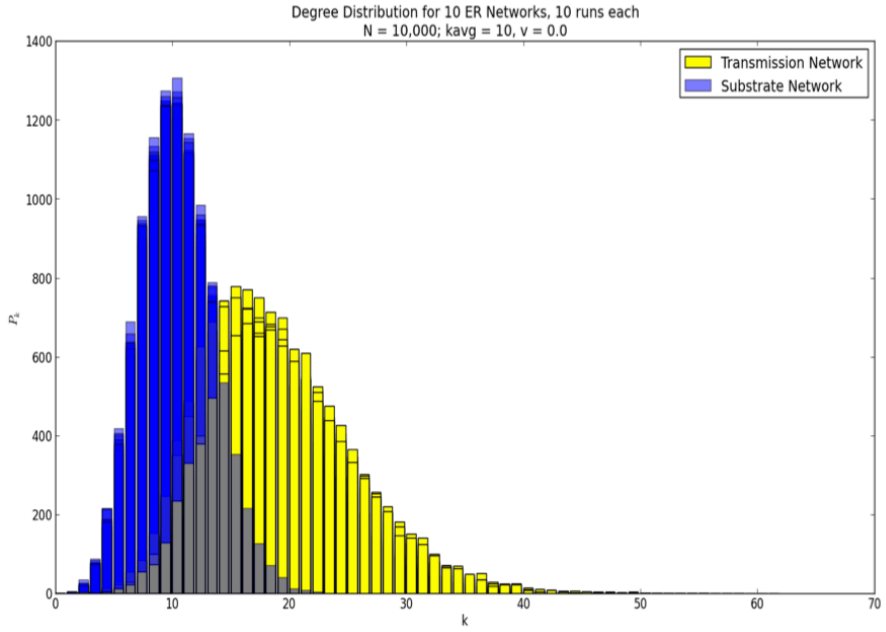

Figure 5. Histogram of degree distributions for many ER random networks constructed with $\langle k \rangle = 10$.

Relevant data: $\langle AR \rangle = 0.96371$.

These results can be qualitatively explained by noting the transmission network is a sub-network of the contact network with added directionality. The degree distribution of this transmission network should tend towards higher $\langle k \rangle$, because those individuals with higher degree are more likely to be infected. However, the structure of the degree distribution, a Poissonian distribution in the case of *Figure 5*, should be unaltered. These results confirm the

often tacit assumption that the transmission network created from a contact network resembles the contact network and has similar properties to its parent contact network.

2.2 NETWORKS OF TWO COMMUNITIES

Whereas in the last section several complex networks consisting of single communities were examined, in this section I will *wire* networks with varying properties together with a small number of cross connections. There is some precedent for using this method to build two-community models, such as in [16], which identified three epidemic regimes for weakly connected or coupled networks: one, where the epidemic does not spread across networks (the *Disease-Free* regime), a second where the epidemic spreads across networks, but does not cause a second epidemic (the *Mixed* regime), and a third where the epidemic spreads to both networks (the *Epidemic* regime). For all three regimes, an outbreak can occur on the network containing the index case–community A, but may not form on the second community–community B.

Additionally, [16] calculated the critical number of cross connections above which the networks will enter the Epidemic regime. This quantity was calculated to be

$$\langle k_{AB} \rangle_c = \frac{\sqrt{\langle k_A \rangle \langle k_B \rangle - \langle k_A \rangle^2} - \langle k_A \rangle}{2}$$

where $\langle k_A \rangle$ is the average degree of A, $\langle k_B \rangle$ is the average degree of B. The quantity $\langle k_{AB} \rangle_c$ was calculated assuming both A and B were random networks generated using the Molloy-Reed configuration model. However, for ER and BA networks the critical average cross-

connections that distinguish strongly coupled from weakly coupled networks have not been determined in the literature. Additionally, in this thesis the number of cross-connections will always be a constant parameter of the model and not an average. Therefore, in this thesis the symbol n will be used to represent the number of cross-connections between two networks and this notation will be expanded later in this chapter and in *Chapter 3*.

2.2.1 Methods

In this section, I will investigate the spread of epidemics across networks consisting of two communities A and B. In general, each community will be generated to have the same structure, i.e., ER for random networks and BA for scale-free networks. However, in general each network will contain a different number of individuals such that $N_A \neq N_B$ and a different average connectivity $\langle k_A \rangle \neq \langle k_B \rangle$; in other words, these will be *heterogeneous weakly connected communities*. The smaller communities are here called *small* or occasionally *minority* communities, while the larger networks are called *big* or occasionally *majority* communities. A visualization of one of these two-community networks is shown in *Figure 6* where the total $N = 100$ for ease of viewing. These types of networks have not been thoroughly researched in the literature in a computational and network-theoretic way.

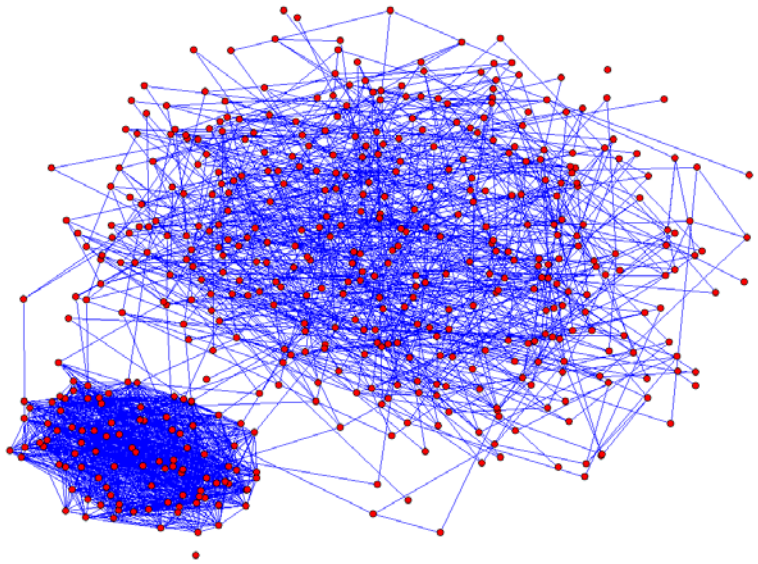

Figure 6. Representative heterogeneous two-community ER network with $N_{small} < N_{big}$ and

$$\langle k \rangle_{small} > \langle k \rangle_{big}.$$

For the purposes of this thesis, cross-connections were always formed randomly between the two communities, even in cases when the structure of each individual community was scale-free. This choice was justified by the small number of cross-connections relative to the number of individuals in each community ($n \ll N_A$ and $n \ll N_B$) and provided the benefit of easing analysis. However, future studies could formulate an algorithm for forming cross-connections analogous to the BA and ER algorithms (See *Discussion*).

Many of the simulations in this chapter were found on a two-community BA network consisting of one large, weakly connected community ($N_A = 8\,000$, $\langle k_A \rangle = 10$) and one small, densely connected community ($N_B = 2\,000$, $\langle k_B \rangle = 20$). Infections were started by choosing a

random index case, either in A or B, and progressing the infection forward. Unless otherwise stated, the following conditions held for all networks studied:

- Each time step was considered a single day in simulation time and the infectivity was set constant as $\beta = 0.1$.

  Most simulations were run on 10 different networks, with 10 runs on each network and with a total population $N = N_A + N_B = 10\,000$ for each run.

- The infection model was Susceptible Infected Recovered (SIR).

2.2.2 Results

As previously mentioned, the literature on the importance of cross-connections speaks to the importance of the coupling strength, or number of cross connections, between two communities. However, there has been little if any investigation into the role of cross-connections in scale-free networks, such as BA networks.

For a single cross connection $n = 1$, $Figure\ 7$ shows a slight but noticeable difference in infection duration. For all such simulations, the average outbreak duration t_b, or burnout time, was shown to increase. This result is two-fold: 1) an infection crossing between populations causes a secondary infection wave, manifested globally as an increased burnout time, and 2) crossing over occurs on average for as few as a single cross connection ($n = 1$). This result could be further studied by quantifying the amount by which each individual cross-connection

increases the average outbreak duration. However, a large increase in t_b and the severity of the second peak was seen with small increases in n.

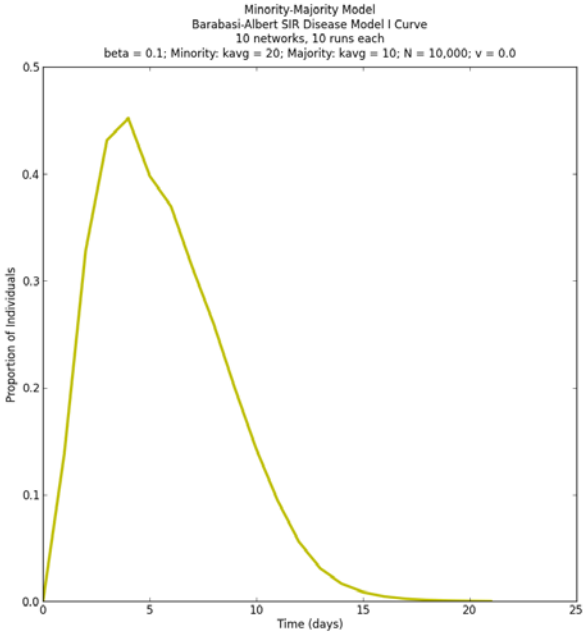

Figure 7. Average Infection Curve for BA network Ten runs on ten networks. $N_{majority} = 8\,000$; $\langle k_{majority} \rangle = 10$; $\langle k_{minority} \rangle = 20$; $n = 1$. Average $t_b = 20.07$ A small, but noticeable bump indicating the crossing of the infection between communities can be seen to the right of the infection curve. This is representative of all such trials. In this figure, the infection began on the large (majority) community.

The extent to which the number of cross-connections altered the attack rate was also investigated and found to be anomalous (*Figure 8*).

Cross Connections (n)	Average Attack Rate (AR)
1	0.7111
100	0.92393
500	0.83207
1 000	0.83533
1 500	0.92929

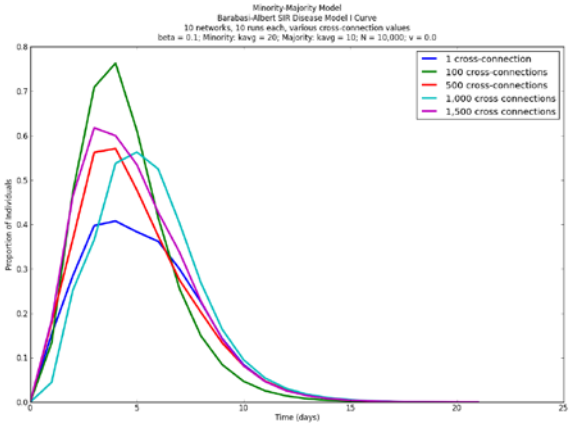

Figure 8. Attack Rate (AR) vs. Cross Connections for several values of n. The values of AR anomalously peak around $n = 100$ and $n = 1\ 500$.

The values of AR are expected to increase with increasing n, so the multiple peaks shown in the table in *Figure 8* are unexpected. This anomaly is likely the result of the random cross-connection algorithm breaking down for n large as the values of $\langle k \rangle$ change for both networks. The change in $\langle k \rangle$ for each community can be quantified by writing the exact expression

$$\langle k_i' \rangle = \langle k_i \rangle + \frac{2n}{N_i}$$

(2.2.1)

where we define a measure of the change in average degree by

$$\delta_i = \frac{2n}{N_i}.$$

(2.2.2)

The results of combining (2.2.1) with the data in *Figure 8* are shown in *Table 1*.

Table 2. Summary of how various cross connections change the average degree of communities in a two-network system.

Cross Connections (n)	Average Attack Rate (AR)	δ_{small}	% $\langle k_{small} \rangle$	δ_{large}	% $\langle k_{large} \rangle$
1	0.7111	10^{-3}	5×10^{-3}	2.5×10^{-4}	2.5×10^{-3}
100	0.92393	10^{-2}	5×10^{-2}	2.5×10^{-2}	0.25
500	0.83207	0.5	2.5	0.125	1.25
1 000	0.83533	1.0	5	0.25	2.5
1 500	0.92929	1.5	7.5	0.375	3.75

From these data it can be estimated that $\delta_c \sim 0.1$ for two community systems. Therefore, n values lower than $n = 100$ were used for all communities of $N = 10\,000$ for the remainder of this thesis.

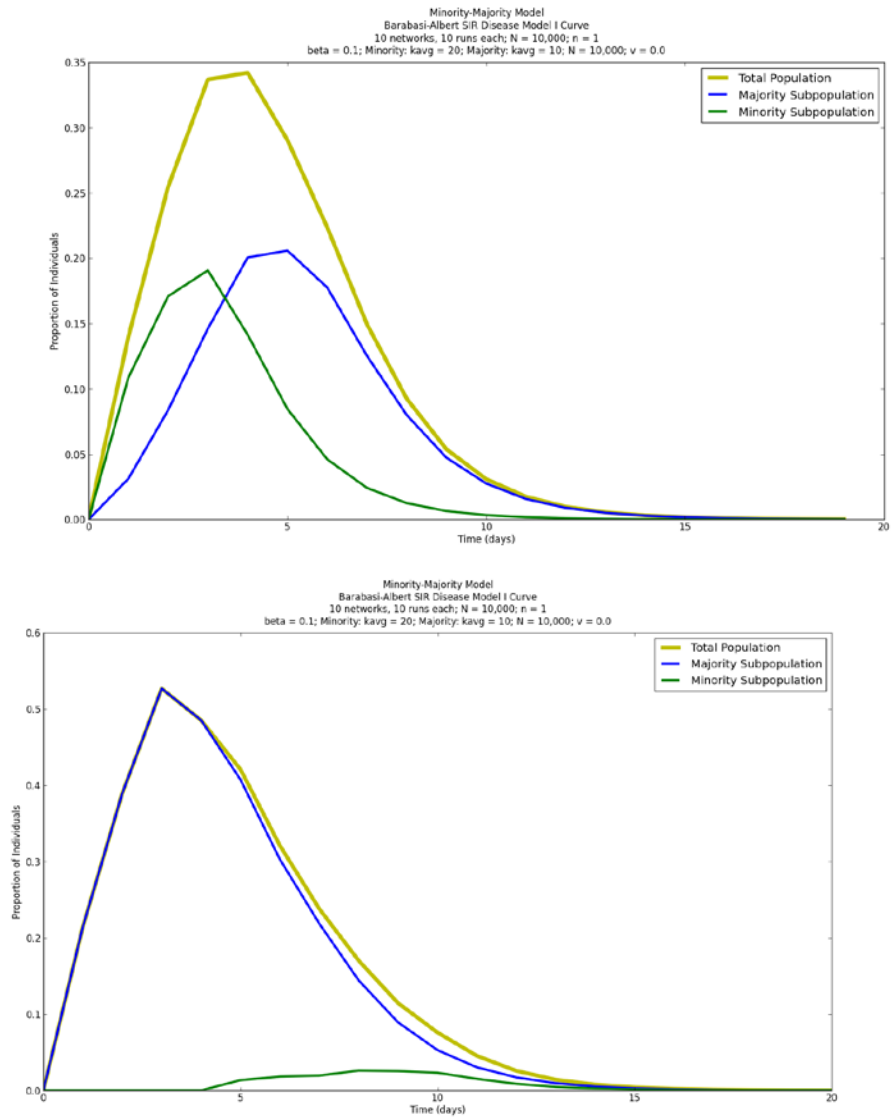

Figure 9. Infection curves for individual communities plotted alongside total network infection curve for infections originating in a) the denser population and b) the less dense population. Crossing over to the other population is more likely when the epidemic begins in the denser population.

The effects of starting in either the denser or less dense population, as well as starting in the larger or smaller population were studied. It was found that starting in the denser population resulted in a higher likelihood of crossing over to the other community. This trend is clearly

demonstrated in *Figure 9*, where the infection curves of each individual community are plotted along the infection curve for the total network.

Further investigation into the nature of two-community systems with asymptomaticity and differential stay-home probabilities, like the investigation in *Section 2.1*, is needed (See *Discussion*). Additionally, studies into multiple community systems with these extensions along with the extension of differential vaccination described in this section should be conducted.

Paragraph.

3.0 R_0 ON NETWORKS OF COMMUNITIES

The basic reproductive number plays a central role in the modern understanding of infectious diseases. It is often described as the most important parameter in epidemiology and the modeling of infectious diseases [8, 17]. The centrality of this measure is well reflected in the vastness of the literature on the subject. This literature can be broadly characterized as those works that study the value of $\mathcal{R}_0$ in a real-world epidemic and those that study the value of $\mathcal{R}_0$ on theoretical models. As in all theoretical and experimental research, these two classes of research critically depend on each other. Calculating $\mathcal{R}_0$ values for *continuous* compartmental disease models is well researched, especially in the procedure involving calculating the *next-generation matrices* found in [7]. However, little headway has been made in the literature towards calculating $\mathcal{R}_0$ values on networks.

In this chapter, I will attempt to expand on the work of *Chapter 2* by investigating $\mathcal{R}_0$ on networks, both in general (*Section 3.2*) and in the case of multiple community networks (*Section 3.1*). I will attempt to extend the notion of basic reproductive rate in two ways: first, by introducing a new technique for calculating $\mathcal{R}_0$ values on networks containing an arbitrary finite

number of communities, and second by drawing a novel correspondence between the spread of epidemics over contact networks and the Mean Squared Displacement (MSD) of the diffusion process.

3.1 BASIC AND MODIFIED REPRODUCTIVE RATES ON INHOMOGENEOUS CONNECTED COMMUNITIES

In this section I develop a generalized theoretical framework for describing essential properties of connected communities and use it to derive an expression for the basic reproductive number or basic reproductive rate $\mathcal{R}_0$ on networks consisting of an arbitrary number of communities. While continuous models consisting of two communities ([20]) and network models consisting of two communities ([7]) have been studied, analyses of networks containing an arbitrary number of communities are not found in the literature.

We begin by considering a single network (C) consisting of j communities $C_1, C_2, \ldots, C_j$ with populations $N_1, N_2, \ldots, N_j$ average degree $\langle k_1 \rangle$, $\langle k_2 \rangle$, …, $\langle k_j \rangle$ "weakly" connected with *n* edges. For the present analysis, we consider only static networks where n is a constant.

By definition the average connectivity of the ith community can be expressed in terms of the total connections of that community divided by the total population of that community, viz.

$$\langle k_i \rangle = \frac{k_i}{N_i} \leftrightarrow \langle k_i \rangle N_i = k_i$$

(3.1.1)

and for reasons that will become clear shortly, define the total network population as N_0, such that

$$N \equiv N_0 = \sum_{i=1}^{j} N_i.$$

(3.1.2)

Using these definitions, the network can be expressed as a j dimensional vector space formed with orthonormal basis vectors $\widehat{C_1}$, $\widehat{C_2}$, …, $\widehat{C_j}$, each of which corresponds to a particular community.

For any such network, we have the connectivity vector:

$$\boldsymbol{k} = \sum_i \langle k_i \rangle \widehat{C_i}$$

(3.1.3).

Connecting the communities with n cross-connections we have,

$$\langle k_i \rangle' = \langle k_i \rangle + \delta_i,$$

(3.1.4)

where δ_i is the contribution to the average degree of the *i*th community from adding n cross-connections. Thus,

$$\delta_i \equiv \delta_i(n) = \frac{n}{N_i}.$$

N.b., as before, $\langle k_i \rangle' \to \langle k_i \rangle$ as $\delta_i \to 0$.

The average degree of the total network is related to the average degree of each community by

$$\langle k \rangle \equiv \langle k_0 \rangle = \frac{1}{N} \sum_i k_i.$$

(3.1.5)

Introducing the weighted population vector $\boldsymbol{\omega}$, viz.

$$\boldsymbol{\omega} = \frac{1}{N} \sum_i N_i \vec{C}_i = \sum_i \left(\frac{N_i}{N} \right) \vec{C}_i$$

with $\omega_i = \frac{N_i}{N}$ we have

$$\boldsymbol{\omega} = \sum_i \omega_i \hat{C}_i$$

(3.1.3)

Connecting the communities the n cross connections yields,

$$\langle k_i \rangle' = \langle k_i \rangle + \delta_i$$

(3.1.4)

where δ_i is the contribution to the average degree of the ith community from adding n cross connections. Thus,

$$\delta_i \equiv \delta_i(n) = \frac{n}{N_i}$$

and, note, as before $\langle k_i \rangle' \to \langle k_i \rangle$ as $\delta_i \to 0$.

The average degree of the total network is related to the average degree of each community by

$$\langle k \rangle \equiv \langle k_0 \rangle = \frac{1}{N} \sum_i k_i.$$

(3.1.5)

Introducing the weighted population vector $\boldsymbol{\omega}$, viz.

$$\boldsymbol{\omega} = \frac{1}{N} \sum_i N_i \hat{C}_i = \sum_i \left(\frac{N_i}{N} \right) \hat{C}_i$$

recognizing $\omega_i = \frac{N_i}{N}$ this becomes

$$\boldsymbol{\omega} = \sum_i \omega_i \hat{C}_i$$

using (3.1.1) and (3.1.5) yields

$$\langle k \rangle = \frac{1}{N}\sum_i \langle k_i \rangle N_i = \boldsymbol{k} \cdot \boldsymbol{\omega}.$$

(3.1.6)

Defining the cross connection vector $\boldsymbol{\delta}$, viz.

$$\boldsymbol{\delta} = \sum_i \delta_i \hat{C}_i$$

(3.1.7)

or

$$\boldsymbol{\delta} = \frac{n/N}{\boldsymbol{\omega}}$$

Finally, we can recast the modified connectivity vector (3.1.4) as the sum of vectors, viz.

$$\boldsymbol{k}' = \boldsymbol{k} + \boldsymbol{\delta}.$$

(3.1.8)

Using this decomposition, we can calculate the Average Modified Reproductive Rate (AMRR) $\langle R_0' \rangle$ on Nonhomogeneous Connected Communities from the canonical *Basic Reproductive Rate* (BRR) $\mathcal{R}_0$ on a single community.

The *Basic Reproductive Rate* $\mathcal{R}_0$ of an SIR infection on a single community can be approximated using its definition as the product of the infectious period p, the infectivity β, and the average connectivity $\langle k \rangle$, i.e.,

$$\mathcal{R}_0 = p\beta\langle k \rangle.$$

(3.1.9)

However, for j connected communities where at least 1 of the j communities has a different average connectivity, $\mathcal{R}_0$ cannot be written in the canonical form. In fact, this *heterogeneity condition* can be seen as a condition that differentiates one community systems from many community systems. For this case, we introduce a *Modified Reproductive Rate* (MRR) $\mathcal{R}_0'$ for nonhomogeneous connected communities.

Following the convention of this section, we calculate the $\mathcal{R}_0$ value for each community sans cross connections and build an $\mathcal{R}_0$ vector

$$\boldsymbol{\mathcal{R}_0} = \sum_i \mathcal{R}_{0i}\hat{C}_i$$

(3.1.10)

where $\mathcal{R}_{0i}$ is the BRR for the ith community.

Using the assumption that for a basic SIR model p and β are invariant over different network topologies, the *Modified Reproductive Rate* (MRR) for the ith community can be expressed as

$$\mathcal{R}'_{0i} = (\mathcal{R}_{0i})' = (p\beta\langle k\rangle_i)' = p\beta\langle k\rangle'_i.$$

Using (3.1.4) yields,

$$\mathcal{R}'_{0i} = p\beta(\langle k_i\rangle + \delta_i)$$

(3.1.11)

and the MRR vector can be defined as

$$\boldsymbol{\mathcal{R}'_0} = \sum_i \mathcal{R}'_{0i}\hat{C}_i.$$

Recognizing that the MRR will be a weighted average of the BRRs the Average Basic Reproduction Rate (ABRR) can be calculated using the weighting vector, viz.

$$\langle R'_0\rangle = \boldsymbol{R'_0}\cdot\boldsymbol{\omega}.$$

(3.1.12)

Therefore,

$$\langle R'_0\rangle = \sum_i R'_{0i}\omega_i$$

and using (3.1.8),

$$\langle R_0' \rangle = p\beta \ \boldsymbol{\omega} \cdot \boldsymbol{k}'.$$

Finally, as in (3.1.6) we can define: $\langle k \rangle' = \boldsymbol{\omega} \cdot \boldsymbol{k}'$ and express the AMRR as

$$\langle R_0' \rangle = p\beta \langle k \rangle'$$

(3.1.13)

Where again $\langle \mathcal{R}_0' \rangle \to \mathcal{R}_0 \ as \ \boldsymbol{\delta} \to \mathbf{0}$.

It has been shown here that the expression for the basic reproduction rate for a network consisting of multiple communities (3.1.5) can be expressed in a format nearly identical to that of the basic reproduction rate of a network consisting of a single community (3.1.1).

3.2 R_0 AND THE MEAN SQUARED DISPLACEMENT ON TRANSMISSION NETWORKS

As far back as Einstein, the Mean Squared Displacement (MSD) has been examined as a central quantity in understanding stochastic diffusion processes [21, 22]. In the past decade [23] and [24] have examined and computationally characterized the diffusion of epidemics on several types of complex networks, including scale-free networks, as anomalous *superdiffusion*. However to my knowledge, there has not been a published effort to draw a correspondence between the MSD and the value of $\mathcal{R}_0$ on networks.

We seek a relationship between $\langle r^2 \rangle(p)$ and $\mathcal{R}_0$, where the first expression is the MSD of the infection on a network at a time p, where again p is the duration of the infectious state, and the second expression is the Basic Reproductive Number. Note that in this section alone, the $\langle \dots \rangle$ brackets indicate a distribution average.

For a diffusion process on a network we expect the one dimensional diffusion equation to hold and the MSD to be given exactly by

$$\langle r^2 \rangle(t) = 2Dt^{\gamma}, \tag{3.2.1}$$

where (3.2.1) is the Einstein diffusion relation for random diffusion along one dimension or along networks. For $\gamma = 1$ (3.2.1) describes pure diffusion, for $0 < \gamma < 1$ (3.2.1) describes sub-diffusion, for $1 < \gamma < 2$ (3.2.1) describes super diffusion, finally for $\gamma > 2$ (3.2.1) describes ballistic diffusion. For the case of normal diffusion, $\gamma = 1$, the density of the diffusive substance, or equivalently the probability of a single random walker occupying a given position, is given by a Gaussian distribution, viz.

$$P(r,t) = \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{r^2}{4Dt}} = \frac{1}{\sqrt{2\pi \langle r^2 \rangle(t)}} e^{-\frac{x^2}{2\langle r^2 \rangle(t)}}$$

where for a particle beginning at the origin $r(0) = 0$ which leads to $\langle r^2 \rangle(t) \equiv \sigma^2$ where σ^2 is the standard deviation whose dependence on time has been suppressed. This yields the diffusion distribution

$$P(r,t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}.$$

Plots of how $P(r,t)$ changes with time are shown in ***Figure 10***.

Again, in this section alone $\langle \ldots \rangle$ indicates the mean value of a distribution, viz.

$$\langle f \rangle(t) = \int_{-\infty}^{\infty} P(r) f(r,t) dr$$

such that the MSD is defined as

$$\langle r^2 \rangle(t) = \int_{-\infty}^{\infty} P(r) r^2(t) dr = 2Dt.$$

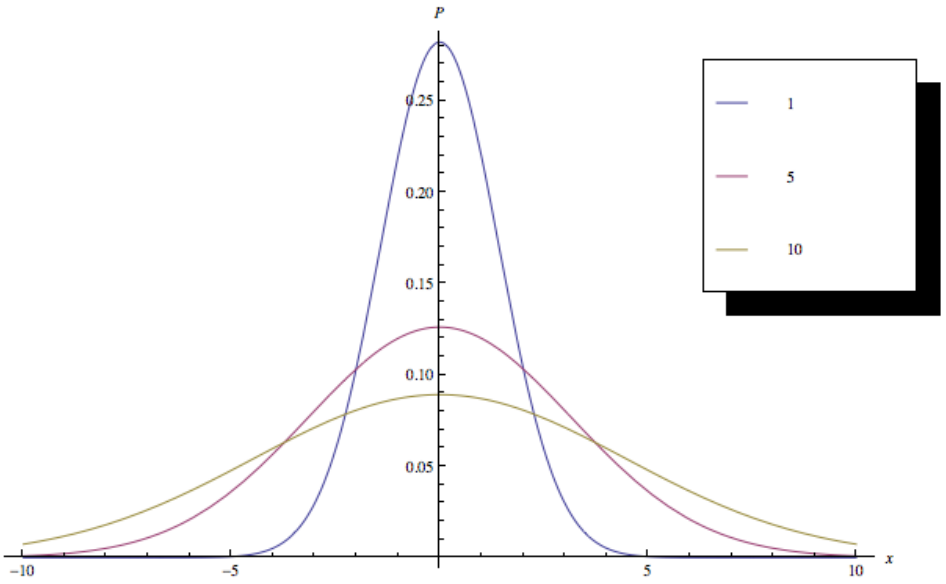

Figure 10. Plots of the Gaussian diffusion probability $P(r,t)$ at time steps, $t = 1, 5,$ and 10, $D = 1.0$ for all cases shown. The distribution $P(r,t)$ can be interpreted as representing many quantities, such as the density of a drop of diffusive material in a medium over time, the probability of a random walker being in a given position over time.

Note, the values of D and γ in (3.2.1) are expected to be dependent only on the structure of the network and independent of the properties of the disease, save for the duration of the infection p. Therefore, a relationship of the form

$$\langle r^2 \rangle(p) = f(\mathcal{R}_0),$$

(3.2.2)

should be likewise independent of the infection. This leads to a possible application of this equation to public health. If the structure of a network is determined, such that D and γ are known, then the $\mathcal{R}_0$ of an infection could be entirely determined given p. The network-dependent parameters D and γ could be found either from studying previous infection data, sociological data, or through some direct experiment. In reality, the network-dependent parameters would vary with dynamic network structures, for example, if individuals self-isolate or are quarantined. Likewise, there may be deviations from (3.2.2) when there is an exposed compartment in the model (e.g. SEIR). These extensions will not be treated here.

Now that some of the possible implications of (3.2.2) have been explored, we seek the function explicitly. To begin, we define distance on a network. On a non-geometric network, distance is described as the minimum number of edges between one node and another. Two nodes are said to be adjacent if there exists a path between them containing only a single edge, i.e., if there is *an* edge between them. Second, we define the origin of the network coordinate system to be the first infected, or index, case. In this *index coordinate system*, the distance from an infected node i to the origin O goes as the number of infected individuals between i and O inclusive, viz.

$$r_{iO}(t) \leq I^{(i)}(t) \tag{3.2.3}$$

where $I^{(i)}(t)$ is the number of infected individuals along the ith trajectory. The deviation from equality for these individual trajectories comes from when an individual infects someone *behind* them, i.e., closer than they are to the origin, the number of infected individuals will be undercounted. For a sufficiently infectious disease on a random network of sufficiently high average degree $\langle k \rangle$, the number of individuals infected *in front* of an individual more than compensates for the individuals infected *behind* an individual. This effect comes about quantitatively because individuals who are already infected or recovered cannot be infected again, so the diffusion is biased away from the origin. However, for power-law networks with hubs that may be located behind the infection wave, equation (3.2.3) likely deviates significantly from equality.

Introducing an undercounting diffusion term, use (3.2.3) to write the exact equation, viz.

$$r_{io}(t) + u^{(i)}(t) = I^{(i)}(t),$$

(3.2.4)

where $u^{(i)}(t)$ is the number of individuals infected horizontal to or behind the distal arm of the ith trajectory at time t. This undercounting term can be exactly written as

$$u^{(i)}(t) = h^{(i)}(t) + b^{(i)}(t)$$

(3.2.5)

where $h^{(i)}(t)$ is the number of horizontal transmissions along the ith trajectory and $b^{(i)}(t)$ is the number of backward transmissions along the ith trajectory. Each type of transmission should contribute equally to the undercounting, so each carries a factor of unity. However, we are not interested in the trajectories for infected individuals on the interior of the infection wave, as we might be interested in the particles of oil on the interior of an oil-droplet diffusing. In the case of an oil droplet, the amount of oil is conserved and the density of droplet particles on the interior consequently decreases with time. Contrastingly, an infection is a biological process that generates new infected individuals, without decreasing some "infection density" on the interior. Therefore, it is appropriate to treat the diffusion of a disease across a network as many forward biased random walkers with a common origin, rather than a diffusion of some fluid.

In the case of many random walkers, the relevant distances become those from the individuals on the perimeter of the infection wave to the origin. In the convention of (3.2.4), this can be written as

$$r_{i0}(t) = J^{(i)}(t) + u^{(i)}(t)$$

(3.2.6)

where $J^{(i)}(t)$ is the distance to the origin of an individual infected at time $t-1$, i.e., a *radius* of the transmission network. Going from (3.2.4) to (3.2.5), it is important to note that the *distance* from a newly infected individual to the origin will be equal to the *number* of infected individuals within the infection wave. Of course, this means

$$J^{(i)}(t) + u^{(i)}(t) = I^{(i)}(t).$$

(3.2.7)

However, (3.2.7) *does not* imply

$$J(t) + u(t) = I(t).$$

In fact, in general

$$J(t) + u(t) \geq I(t),$$

(3.2.8)

where omitting the index implies summation over all i and the equality in (3.2.8) holds only for a transmission network with no branching past the index case, which we will call a *ray network*. In words, the sum over all radii of the transmission network is less than or equal to the total size of the transmission network.

The deviation from equality comes when double-counting certain infected individuals. Introducing δ, the double-counting factor, (3.2.8) can be written exactly as

$$J(t) + u(t) - \delta(t) = I(t)$$

(3.2.9)

This double counting factor arises when two or more nodes are infected branching off of a single trajectory, the distance from one of the two nodes to the origin will accurately count the number of interior cases along the trajectory. However, including the distance of the second node will double-count some of these interior case. As time goes on and the forked trajectory grows along both sides of the fork, the double counting will not increase, but will remain equal to the distance from the branching point to the origin. This double-counting parameter is central to the network structure.

To the crudest, or zeroth order, approximation, set $\delta = u = 0$. This is tantamount to only treating tree-level networks without branching, which will be called *ray networks*. In this approximation

$$r(t) = J(t) = I(t).$$

(3.2.10)

Then, the average distance from the center can be written as

$$\bar{J}(t) = \frac{1}{\Delta_t I} \sum_{i \in \{\Delta_t I\}} J^{(i)} = \frac{J(t)}{\Delta_t I}$$

(3.2.11)

that is, as a sum over all such distances divided by the number of individuals infected at time $t-1$, $\Delta_t I = I_t - I_{t-1}$, which is the circumference of the transmission network, $C(t)$. Then the average radius of the transmission network can be written as

$$\bar{J}(t) = \frac{1}{\Delta_t I} \sum_{i \in \{\Delta_t I\}} J^{(i)}(t) = \frac{J(t)}{C(t)} = \frac{1}{C(t)} \sum_{i \in \{\Delta_t I\}} r_{iO} = \langle r \rangle(t)$$

where (3.2.10) was used for the third equality and the fourth equality is the definition of average distance of newly infected individuals from the origin. Then we can write

$$\bar{J}(t) = \langle r \rangle(t),$$

(3.2.12)

which equates the Mean Displacement to the average radius of the transmission network. Using (3.2.1), we can also write the MSD as

$$\langle r^2 \rangle(t) = \bar{J}^2(t) = 2Dt^{\gamma}.$$

(3.2.13)

Recall the definition of $\mathcal{R}_0$ as the total number of infected individuals at $t = p$, where p is the infection duration, we can define $\mathcal{R}_0$ in terms of I viz.,

$$\mathcal{R}_0 = I(p).$$

(3.2.14)

3.2.14 $\mathcal{R}_0$

Note ( ) is an exact definition of and is independent of any approximation. Substituting

3.2.10

( ) yields

$$\mathcal{R}_0^{(0)} = J(p),$$

(3.2.15)

where $\mathcal{R}_0^{(0)}$ is the zeroth order, or ray-level, approximation for the $\mathcal{R}_0$. Using the derivation for the Mean Squared Radius of the transmission network (3.2.11), we can write

$$\frac{J^2(p)}{C(p)} = \langle r^2 \rangle(p) = \frac{\left(\mathcal{R}_0^{(0)}\right)^2}{C(p)} = 2Dp^{\gamma},$$

therefore, recalling (3.2.2),

$$f_0(\mathcal{R}_0) = \frac{\left(\mathcal{R}_0^{(0)}\right)^2}{C(p)}$$

(3.2.16)

where $f_0(\mathcal{R}_0)$ is the zeroth order function for MSD in terms of $\mathcal{R}_0$, which yields

$$\mathcal{R}_0^{(0)} = C(p)\langle r^2\rangle(p) = \sqrt{2DC(p)p^{\gamma}},$$

(3.2.17)

as sought.

The derivation of (3.2.17), the zeroth order approximation of $\mathcal{R}_0$ only holds for transmission networks with ray structures. We must account for a nonzero δ term, the double counting factor, to extend this work to tree structures, random networks, and eventually complex networks. We expand the ray network structure first by allowing connections across rays, so long as they are between nodes at a further radius from the index case. When drawing a network including all such connections, the network will be a tree network with no horizontal connections, i.e., no connections between nodes at the same distance from the index case. We will also exclude backward diffusion from this approximation ($b(t) \equiv 0$).

The δ term can be easily factored into the previous derivation by combining (3.2.9) and (3.2.14), which leads to

$$\mathcal{R}_0^{(1)} = J(p) + \delta(p).$$

(3.2.18)

Note, we explicitly write δ as a function of time. Intuitively, this is appropriate because the number of branching points on the interior of a transmission network at a given time will depend on the random diffusion mediated by the infectivity β at each transmission.

Squaring (3.2.18) and dividing by the number of recently infected individuals, the circumference of the transmission network at time $t = p$, yields

$$\frac{\left(\mathcal{R}_0^{(1)}\right)^2}{C(p)} = \frac{\left(J(p) - \delta(p)\right)^2}{C(p)} = \frac{J^2(p) - 2J(p)\delta(p) + \delta^2(p)}{C(p)}$$

splitting the fraction gives

$$\frac{\left(\mathcal{R}_0^{(1)}\right)^2}{C(p)} = \frac{J^2(p)}{C(p)} - \frac{2J(p)\delta(p)}{C(p)} + \frac{\delta^2(p)}{C(p)} = \langle r^2 \rangle(p) - \frac{2J(p)\delta(p)}{C(p)} + \frac{\delta^2(p)}{C(p)}$$

where we have identified the mean squared radius of the transmission network at $t = p$, $\overline{J^2}(p)$.

Substituting (3.2.16) and multiplying through by $C(p)$ yields

$$\left(\mathcal{R}_0^{(1)}\right)^2 = \left(\mathcal{R}_0^{(0)}\right)^2 + B(p)$$

(3.2.19)

where we have defined the Branching Function, $B(p)$, such that

$$B(t) = \delta^2\left(1 - 2\frac{J}{\delta}\right).$$

(3.2.20)

where explicit time dependence has been omitted for the right side. A similar derivation will show including the complete expression for (3.2.9) will lead to an additional higher-order correction and, finally, to the complete $\mathcal{R}_0$.

We modify the basic ray network once more to allow for both backwards and horizontal diffusion with respect to the index case. An example of this is the infection beginning on one ray, traveling to a node on another ray at the same radius, and traveling back down that second ray towards the index case. Such hopping is the simplest possible backwards diffusion and would result in undercounting the number of infectious individuals by one for each step backwards and by one for each horizontal step. Therefore, the definition of the Undercounting Function $u(t)$ can be written as

$$u^{(i)}(t) = h^{(i)}(t) + b^{(i)}(t)$$

(3.2.21)

where $h^{(i)}(t)$ and $b^{(i)}(t)$ are the number of horizontal and backward steps on that trajectory at time t, respectively. Using a similar procedure as for the first-order case, we square (3.2.9) at $t = p$ to yield

$$\left(\mathcal{R}_0^{(2)}\right)^2 = \left(J(p) + u(p) - \delta(p)\right)^2 =$$

$$\left(J^2(p) - 2J(p)\delta(p) + \delta^2(p)\right) + u^2(p) + 2u(p)\left(J(p) - \delta(p)\right)$$

grouping terms and substituting $\left(\mathcal{R}_0^{(1)}\right)^2$ yields

$$\left(\mathcal{R}_0^{(2)}\right)^2 = \left(\mathcal{R}_0^{(1)}\right)^2 + U(p)$$

(3.2.22)

where the quantity $U(t)$, the Global Undercounting Function, has been defined as

$$U(t) = u^2\left(1 + 2\frac{J-\delta}{u}\right),$$

(3.2.23)

where we have omitted explicit time dependence for the right side.

Note that a ray network with forward, horizontal, and backwards branching and diffusion describes all possible network structures, thus $\mathcal{R}_0^{(2)}$ is the general expression for $\mathcal{R}_0$ on a network. This yields

$$\mathcal{R}_0 = \mathcal{R}_0^{(2)}$$

(3.2.24)

or by (3.2.17), (3.2.19) and (3.2.22)

$$(\mathcal{R}_0)^2 = \left(\mathcal{R}_0^{(1)}\right)^2 + U(p) = \left(\mathcal{R}_0^{(0)}\right)^2 + B(p) + U(p)$$

and substituting the expression for $\mathcal{R}_0^{(0)}$, we finally have

$$(\mathcal{R}_0)^2 = 2Dp^{\gamma}C(p) + B(p) + U(p).$$

(3.2.25)

This is the sought exact expression for the squared $\mathcal{R}_0$ in terms of only the MSD, the structural parameters of the network on which the disease is spreading, and the circumference of the transmission network, i.e., the number of newly infected individuals, at time p.

The relationship between the squared $\mathcal{R}_0$, the MSD and the network functions $B(p)$ and $U(p)$ is rich and open to futher analysis. The factors $u(t)$ and $\delta(t)$ should and can be calculated for a given network and the limiting cases of small $J(p)$, $u(p)$, and $\delta(p)$ should be explored. Furthermore, the dependence of the $\mathcal{R}_0$ on both $C(p)$, the change in the size of the infectious population from time $t = p - 1$ to $t = p$, and $J(p)$, the sum distance of the those individuals to the index case, should be examined (See *Discussion*). Each of these parameters should depend on both the properties of the pathogen and the properties of the contact network on which it spreads, so we have not succeeded in deriving a completely pathogen independent value of $\mathcal{R}_0$. Indeed, such a derivation is likely impossible. However, what has been accomplished is a closed form expression for the $\mathcal{R}_0$ value in terms of the physical properties of the network

($D, \gamma, \delta, \text{and } u$), properties of the pathogen (p), and properties of *both* the network and the pathogen (C and I). Estimates for the physical network parameters can be found using social network analysis, the literature is replete with estimates for the properties of various pathogens, and the mixed functions $C(t)$ and $I(t)$ can be found via simulations, or even by using a known value of $\mathcal{R}_0$. These are a few of the applications of this formula.

One change that should be noted between the current investigation and that undergone by [23] and [24] is the definition of displacement and, correspondingly, MSD. Whereas the previous papers defined the displacement as (adopting to my notation)

$$\langle r \rangle = \frac{1}{I(t)} \sum_{i=1}^{I(t)} r_i(t),$$

i.e., the mean displacement is defined as the displacement for every individual on the interior of the infected population. Contrastingly, in this investigation I have defined the mean displacement as (3.2.11) and the MSD as (3.2.13), which take the circumference of the transmission network C_t in the denominator and thus measure only the individuals infected in the last time step. As previously mentioned, this is analogous to taking the MSD of multiple forward biased random walkers, instead of the MSD of an oil drop or similar system. The advantage of this definition is that it relates the distance from the center of an individual on the edge of the infection wave to the number of infected individuals within a small sliver of the infection area. However, the change in convention means the conclusion of superdiffusion presented in [23] may not hold.

Thus, further investigation into the nature of the anomalous diffusion, i.e. the values of γ, for various network structures using these definitions is required.

4.0 SEIQR AND NETWORKS MODELS OF EBOLA

In the previous three chapters I described the development of both continuous and network models for modeling infectious disease spread, several results for two community systems, and some general mathematical properties of infections spreading on networks. In this chapter I will analyze and present new continuous differential equation models for the 2014 Ebola Virus Disease (EVD) outbreak. I will then extend this model to the case of a virtual two community systems represented by a dynamic network with nodal immigration. Finally, I will comment on the importance of this model for evaluating potential disease emigration from source communities to new communities, whether on the local, regional, or international level.

4.1 THE DOBSON AND MODIFIED DOBSON SEIQR MODELS

This work has heretofore described the synthesis of compartmental models with agent based network models. One relevant extension to the Susceptible Infectious Recovered or SIR model described in *Chapter 1* is the Susceptible Exposed Infectious Quarantined Recovered or SEIQR

Model. Ebola Virus Disease outbreaks have been modeled using SEIR models, SEIQR models without the Quarantined compartment, and SEIQR models in [13] and [8]. Indeed, the spread of EVD in the 2014 outbreak in West Africa can be well modeled by a simple SEIR model.

However, SEIQR models accommodate the most common and effective intervention for halting the spread of EVD and are widely used in developed nations where EVD has spread. The system proposed by Dobson [8], here called the Dobson Model (DM) (4.1.1), is such a model.

$$\left|\begin{array}{c} \dot{S} = \beta SI \\ \dot{E} = \beta SI - (\varepsilon + q)E \\ \dot{I} = \varepsilon E - (\delta + \alpha + q)I \\ \dot{Q} = q(E + I) - (\alpha + \delta)Q \\ \dot{R} = \alpha(I + Q) - \mu R \end{array}\right.$$

(4.1.1)

One of the primary motivations of the Dobson Model was to estimate values of $\mathcal{R}_0$ based on known parameters. Rearranging (4.1.1), [8] found

$$R_0 = \frac{\beta N \varepsilon}{(\varepsilon + q)(\delta + \alpha + q)}.$$

(4.1.2)

In the procedure of *Chapter* 1, we note (4.1.1) is *not* conservative. In fact, applying the conservation condition (1.1.2) for (4.1.1) gives

$$\sum_{\bar{i}} \dot{C}_{\bar{i}} = 2\beta SI - \delta(I + Q) - \mu R = 0,$$

(4.1.3)

which is not satisfied in general. Although the $-\mu R$ term found in (4.1.3) is common to nonconservative compartmental models, it can be incorporated into the $\dot{R}$ term in the 5th Dobson Equation. Negating the βSI term, which is ubiquitous in other models, can further modify the Dobson Model including the SIR model. Finally, adding a contribution proportional to δ to the 5th Dobson Equation and defining $\alpha' = \alpha + \delta$ as the net removal rate yields

$$\begin{cases} \dot{S} = -\beta SI \\ \dot{E} = \beta SI - (\varepsilon + q)E \\ \dot{I} = \varepsilon E - (\alpha' + q)I \\ \dot{Q} = q(E + I) - \alpha' Q \\ \dot{R} = \alpha'(I + Q) \end{cases}.$$

(4.1.4)

here referred to as the Modified Dobson Model (MDM) and schematically represented in *Figure* 11. For this model the new $\mathcal{R}_0$ can be easily computed as

$$\mathcal{R}_0 = \frac{\beta N \varepsilon}{(\varepsilon + q)(\alpha' + q)}.$$

(4.1.5)

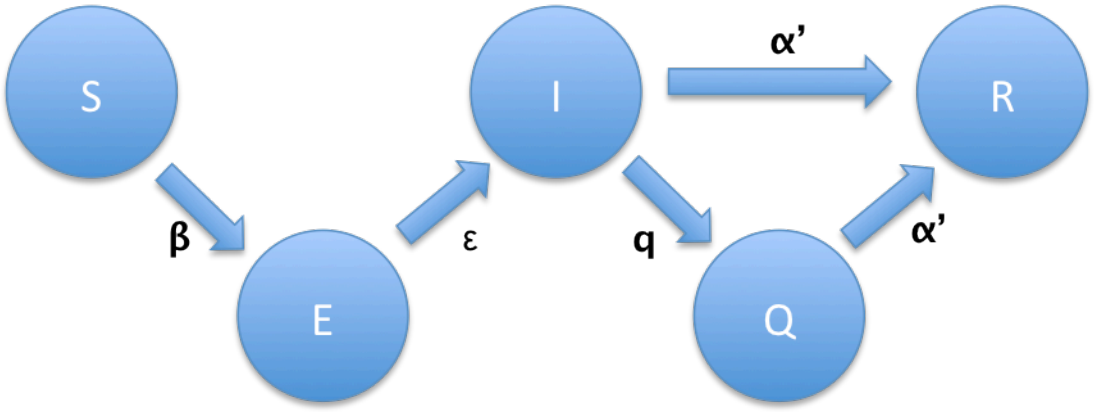

Figure 11. Schematic representation of the Modified Dobson Model (4.1.4). Circles indicate compartments and arrows indicate flow rates.

An additional extension to the Dobson Model was formulated to account for the case when a virgin population is subjected to a constant immigration of exposed individuals. This SEIQR Immigration Model or iSEIQR Model is novel in the literature and could be applied to myriad situations in addition to EVD outbreaks. In the iSEIQR Model an individual may immigrate into the Exposed population E with a constant probability η or into the Susceptible population S with a probability $1 - \eta$. Immigration occurs at a constant rate m. These conditions yield

$$\begin{cases} \dot{S} = (1-\eta)m - \beta SI \\ \dot{E} = \eta m + \beta SI - (\varepsilon + q)E \\ \dot{I} = \varepsilon E - (\alpha' + q)I \\ \dot{Q} = q(E + I) - \alpha' Q \\ \dot{R} = \alpha'(I + Q) \end{cases},$$

(4.1.6)

which has a nontrivial $\mathcal{R}_0$ expression and is schematically represented in *Figure 12*.

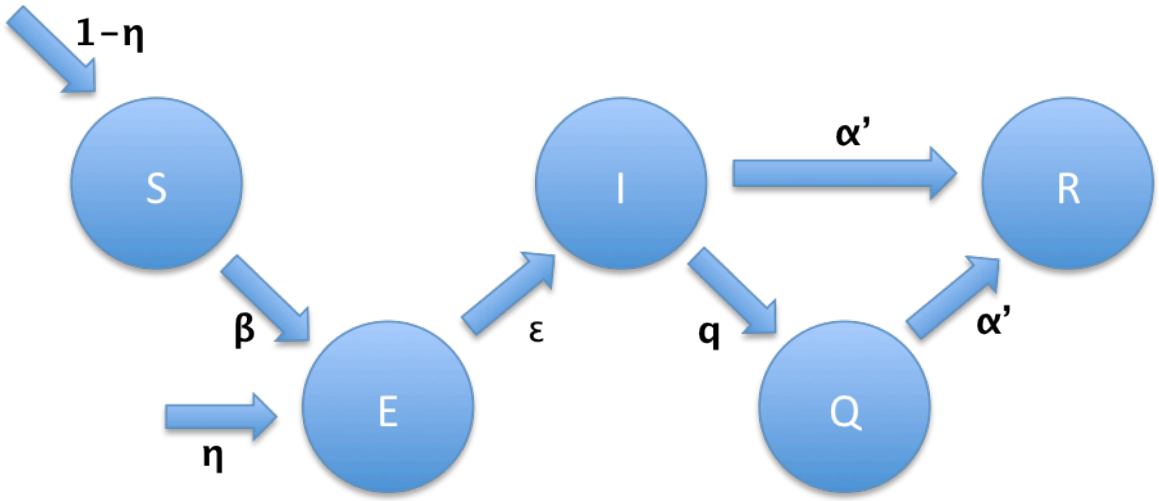

Figure 12. Schematic representation of the iSEIQR Model (4.1.6). The two arrows pointing towards the S and E compartments can be thought of as flows (migrations) from an external second community.

4.2 DISCRETIZATION OF COMPARTMENTAL DIFFERENTIAL EQUATION MODELS

The compartmental differential equation models described in this work now include the SIR, the SEIR, two SEIQR, and the iSEIQR models. These are just a few of the many models proposed in the literature, not to mention the infinite possible extensions to these models. Due to this vast array of differential equation models, I will here present a new notational convention for describing these models and their conservation equations. This notational convention will also

ease the discretization of compartmental differential equation models, which is tantamount to writing differential equations as discrete difference equations over networks.

The generalized compartmental recursion relation (1.1.2) can be written to describe a discrete difference equation, viz.

$$C_{t+1} = C_t + v_t$$

(4.1.7)

where v_t is the change in the compartment during a given time step t. Defining $\Delta C \equiv C_{t+1} - C_t$ leads to $\frac{\Delta C}{\Delta t} = v_t$. Without loss of generality we take $\Delta t \equiv 1$ then any compartmental model of ordinary differential can be written as

$$\begin{vmatrix} \frac{\Delta C_0}{\Delta t} = v_{t,0} \\ \vdots \\ \frac{\Delta C_n}{\Delta t} = v_{t,n} \end{vmatrix}$$

(4.1.8)

In vector notation, (4.1.8) can be written compactly as

$$\frac{\Delta \boldsymbol{C}}{\Delta t} = \boldsymbol{v}_t$$

(4.1.9)

where $\boldsymbol{C} = (C_0, \dots, C_n)^T$ and $\boldsymbol{v}_t = (v_{t,0}, \dots, v_{t,n})^T$. In this notation, the conservation condition can be written as a matrix equation, viz.

$$\mathbb{I}\boldsymbol{v}_t = 0, \forall t$$

(4.1.10)

where $\mathbb{I}$ is the identity matrix of order n. This is the generalized difference equation in vector form. Taking the limit $\Delta t \to 0$ yields the vector differential equation

$$\frac{d\boldsymbol{C}}{dt} = \dot{\boldsymbol{C}} = \boldsymbol{v},$$

(4.1.11)

which is the generalized system of compartmental differential equations in vector form. For this case, the conservation condition can thus be written

$$\mathbb{I}\boldsymbol{v} = \mathbb{I}\dot{\boldsymbol{C}} = 0.$$

(4.1.12)

This form of the conservation is compact and the derivation utilizes the correspondence between the continuous and discrete models.

In this form we may rewrite the conservation conditions for the models mentioned elsewhere in this thesis. For the SIR, SEIR, and Modified Dobson SEIQR Models (4.1.12) is the exact conservation expression. However, for the Dobson SEIQR Model (4.1.3) becomes

$$\mathbb{I}\dot{C}_{DM} = 2\beta SI - \delta(I+Q) - \mu R$$

and for the iSEIQR Model we can write the conservation equation as

$$\mathbb{I}\dot{C}_{IM} = m \neq 0,$$

(4.1.13)

where m is some nonzero positive constant. Of course, for $m < 0$ the Immigration SEIQR model becomes an Emmigration SEIQR model, which will not be further studied in this work.

One noteworthy application of the Immigration Model is to the potential spread of EVD through the United States originating from West Africans, or those who have recently spent time in West Africa, especially members of the media and healthcare workers. In this case, η would likely be extremely small, while m would be much higher and would approximate a non-uniform immigration rate as a constant rate. Another application of this model is between neighboring countries, such as Liberia and Sierra Leone. In fact, recent reports on the spread of EVD have focused on the spread of the disease within and across districts of West Africa [25]. This focus on the spread of infections between areas on a local, regional, and even national scale is common in the literature. One clear application for this work would be to model the spread of EVD into a particular region without the need to model the region from which the disease originated [20]. SEIQR model exhibits a latent stage and because quarantine is a primary source of disease containment. One study found ignoring the latent, here the Exposed, stage of a disease always leads to an underestimation of $\mathcal{R}_0$ values [26].

4.3 NETWORK ISEIQR AND TWO-COMMUNITY SEIQR MODELS

The transnational spread of Ebola Virus Disease (EVD) was simulated and studied using the discretized Immigration SEIQR (iSEIQR) Model on a dynamic complex network containing a single community fed by a constant rate of immigration and using the Modified Dobson (SEIQR) Model on a two community network, like that described in $Chapters\ 1-3$. Each added immigrant was randomly wired to $\langle k \rangle$ nodes. As in $Ch.\ 2$, while $N_{immigrants}$ is sufficiently low, that is $N_{immigrants} \ll N$, this random rewiring should not significantly effect the global network structure of non-random networks, e.g., scale-free networks.

The Immigration SEIQR Model was simulated on both Erdős-Rényi (ER) random networks and Barabási-Albert (BA) scale-free networks. Each simulation was run for 20 networks each with $N = 10\ 000$, and $\langle k \rangle = 10$. The parameters of the infection are displayed in $Table\ 2$ and were taken from literature values whenever possible.

Table 3. Infection parameters for iSEIQR Model on a network. Durations chosen as rounded average infection duration and recommended quarantine duration [25].

Infection Parameter	Value
Immigration Rate: m	$0.2/day$
Probability Exposed: η	$0.2/immigrant$
Probability of Quarantine: q	$0.2/(day * infectious)$
Quarantine Duration: p_q	$21\ days$
Duration of Infection: p	$10\ days$
Duration of Exposure: p_E	$10\ days$

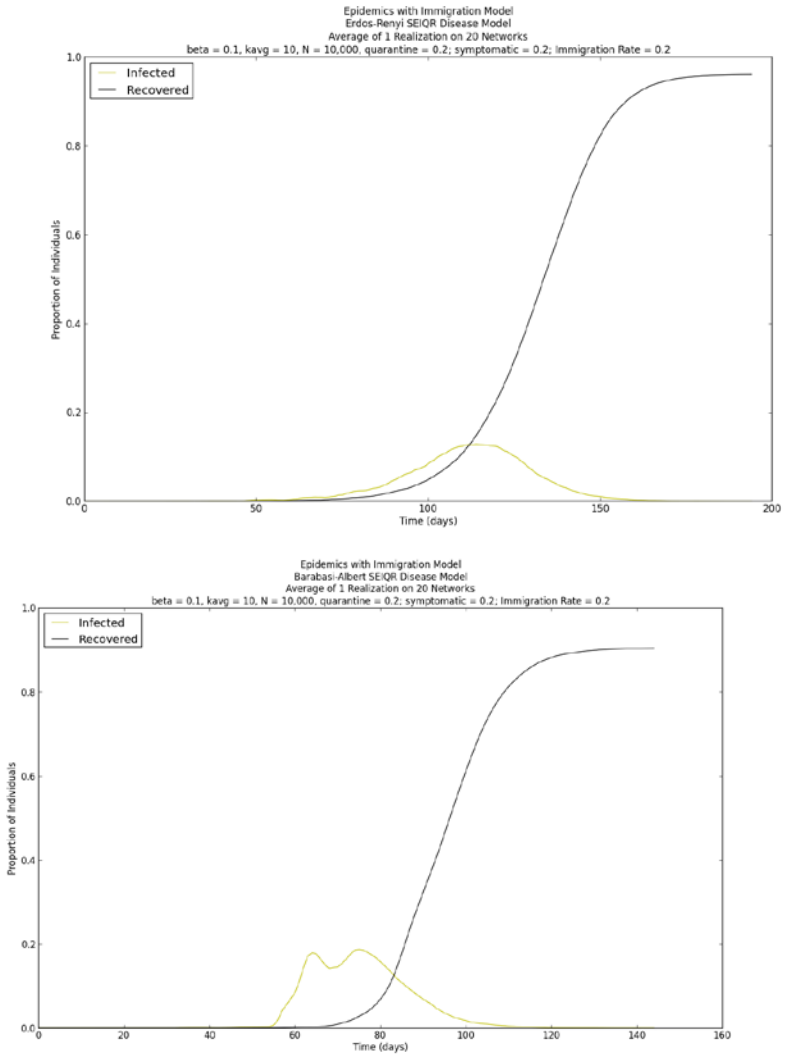

Figure 13. Infection and Recovered curves for iSEIQR network simulations with the parameter values listed in *Table 2*. a) ER networks, b) BA networks.

The results of the iSEIQR network experiments were remarkable and warrant further investigation. For the random (ER) networks tested, the infection duration was markedly increased and the average infection curve (I-curve) was dilated and exhibited noisy behavior (*Figure 13a*). In contrast, the scale-free (BA) networks tested demonstrated an epidemic with two peaks, one visible near $t = 60\ days$ and another for $t = 75\ days$, with a local minimum

near $t = 70\ days$ (*Figure 13b*). The second infection peak was likely the result of, on average, an Exposed immigrant entering the Infectious compartment and infecting a hub at some region of the community untouched by the epidemic.

Examining a single simulation of this model, the Recovered curve (*Fig. 14*) confirms the oscillatory behavior of constant immigration with a probability of exposure.

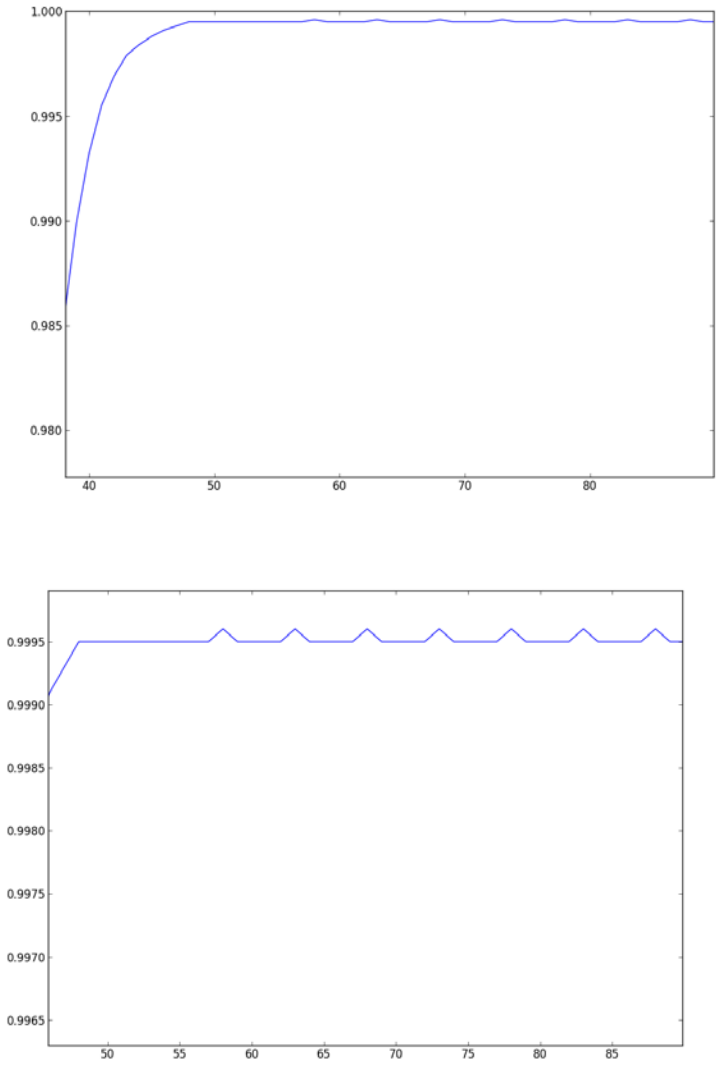

Figure 14. Recovered curves with respect to time for a single simulation. The periodic oscillation of infectious is clearly visible in both subfigures. a) zoomed, b) zoomed further.

The model presented here simulates a two community system, by creating a single community and dynamically importing individuals (“immigrants”) from some otherwise isolated community. Displacing these immigrants from their virtual *other* communities and wiring them into the simulated community is relevant to long-distance travel, where a contact network will not be preserved when an individual travels to a new region, country, or even continent.

DISCUSSION

The problem of modeling infectious diseases, or indeed the spread of any dynamic process across a medium, is complex and in general unsolvable exactly. Since its early history nearly 100 years ago with Kermack and McKendrick, modeling infectious diseases has been a field of applied mathematics and epidemiology that depends heavily on approximate solutions to the most nonlinear and complex problems. This thesis has expounded on and explored network analysis as one method to more exactly simulate this problem and yield solutions, which might lead to unexpected, but effective interventions and policies.

In *Chapter 1*, I have reviewed and discussed the spread of some compartmental disease models across networks with either random (ER) or scale-free (BA) structures. In *Chapter 2*, I have interrogated the special case of two weakly connected communities with varying properties between them, such as connectivity, population size, vaccination probability, and avoidance rates. This investigation yielded several conclusions:

1. Epidemic spread across two community systems is characterized by a secondary time delayed infection wave, which results in an increased outbreak duration (burnout time, t_b).
2. On average, as few as a single cross connection ($n = 1$) results in crossover and a secondary infection wave. In the convention of *Section 3.1*, this can be written as a

crossover occurs, on average, for a $\delta > 0$, where δ was a measure of the cross connections. The size of the secondary infection wave increases with increased connectivity for random networks, but is relatively independent of cross connections for scale-free networks. However, in both cases the time delay is diminished, i.e. the two epidemic waves occur at more similar times, with increased cross connections.

3. Increasing δ_i far above the critical values δ_c may alter the topology of the communities. An estimate of $\delta_c \sim 0.1$ was found for scale-free two community systems. Additionally, beginning the infection in a more densely connected community can increase the probability of the infection crossing over.
4. On average, vaccination and/or avoidance behavior reduces the severity of the infection wave on the vaccinated community. A consequence of this observation is that epidemics are less likely to spread across communities when the initially exposed community has a higher rate of vaccination and/or avoidance. These data on a single community support the conclusion that vaccination and public health education programs are valuable for containing epidemics, both in underserved domestic populations and in developing nations.

I have also expanded the SEIQR differential equation model of Dobson [8] to a similar conservative model and an immigration based SEIQR model. This Immigration SEIQR model (iSEIQR) was then simulated on a network with a constant immigration rate and several conclusions can be drawn.

5. Infection curve homology was shown between the dynamic networks and networks of two communities. Importantly, multiple infection waves are not seen in the model of [8] or other continuous compartmental differential equation models. This is one of the primary benefits of network models demonstrated in this work.

The general mathematical problem of characterizing network epidemic models on multiple communities was also studied. A novel and general vector form of the equations was formulated and used to study the magnitude of perturbation that weak cross connections had on

the properties of individual communities when the networks were joined. One important result of this work was:

6. It was shown that a random wiring process did not significantly perturb the connectivity of either network for weakly connected networks–i.e. small δ. This finding supports the validity of conclusions $1-4$.

   Finally:

7. A relationship between the Mean Squared Displacement (MSD) and the $\mathcal{R}_0$ of an epidemic on a network was derived via decomposition of a general complex network into a novel *ray network* and then adding in corrections for horizontal and backward diffusion. This result demonstrates several important physical concepts:
   a. Epidemic diffusion on a network is equivalent to multiple forward biased random walkers with a common origin.
   b. The MSD diffusion equation, and thus the diffusion coefficient D and the diffusion exponent γ are related to the $\mathcal{R}_0$ of epidemics on networks.

Further, it has been shown in [26] that the 2014 Ebola Virus Disease outbreak was clustered and can be studied using the methods of network analysis. Therefore, the correspondence between $\mathcal{R}_0$ and MSD may be useful in the future to more accurately understand both the contact network of epidemics and the $\mathcal{R}_0$ of a given outbreak.

This work raises many important questions for future research. There are several portions of this work that require further study. Some of the clearer examples of further inquiry are:

1. The differential equation models introduced here can be used to model real world epidemics, such as the 1917 Influenza, the 2004 SARS, and 2014 EBV outbreaks.
2. Basic reproductive rates for the iSEIQR model introduced should be found.
3. As in [8], the Modified Dobson model and iSEIQR model should be solved numerically to characterize solutions.
4. The relationship between in which network an epidemic originates and network heterogeneities should be further investigated.

5. The conclusions about strength of community coupling δ and the properties of each network should be investigated and threshold values found.
6. Values of the diffusion exponent γ for various network structures and epidemic models should be investigated, as in [23].
7. The calculation for basic reproductive rate in terms of MSD should be studied, including the limiting cases with low branching, undercounting, and network radius.
8. An extension to the *Barabási-Albert* (BA) algorithm should be formulated for adding cross connections between networks.
9. Finally, and perhaps obviously, many of these conclusions should be investigated on other network structures such as *Geometric Random Graphs* and *Small-World* networks.

APPENDIX A

TABLE OF COMMONLY USED SYMBOLS

Table 4. Table of commonly used symbols.

Symbol	Name
$\mathcal{R}_0$	Basic Reproductive Number/Rate
β	Infectious probability or inverse rate
λ	Recovery/Lethality probability or inverse rate
X_i	The *i*th compartment
AR	Attack Rate
t_b	Burnout time
k	Degree or connectivity
N	The total population of a network
N_i	The population of the *i*th community
n	The number of cross-connections

δ	Defined as a measure of the change in $\langle k \rangle$ caused by adding n cross-connections. Also, the double counting factor in *Chapter 3.2* and a rate constant in the Dobson Model (4.1.1)
$\vec{C}_i$	The *i*th orthonormal basis vector corresponding to the *i*th community in a network of multiple communities
p	The duration of the infection or the infection period
$\langle r^2 \rangle(t)$	The Mean Squared Displacement (MSD) at time *t*.
D	The Diffusion Coefficient
γ	The Diffusion Exponent ($\gamma = 1$ for pure diffusion)
σ^2	The Standard Deviation
$I^{(i)}(t)$	Infected individuals between the *i*th node and the origin *O*
$r_{iO}(t)$	Distance from the *i*th node to the origin *O* on the transmission network
$u^{(i)}(t)$	Undercounting on the *i*th trajectory on the transmission network
$h^{(i)}(t)$	Horizontal transmissions along the *i*th trajectory
$b^{(i)}(t)$	Backward transmission along the *i*th trajectory

$J^{(i)}(t)$	Distance to the origin of an individual infected at time $t-1$, i.e. a radius of the transmission network
$C(t)$	The circumference of the transmission network, defined as the number of individuals infected at time $t-1$
$B(t)$	The branching function defined in (3.2.20)
$U(t)$	The Global Undercounting function defined in (3.2.23)
η	The probability or inverse rate of immigration in the iSEIQR Model of *Chapter 4*
m	A constant scaling factor in the iSEIQR model related to the rate of immigration ($m > 0$)

APPENDIX B

SAMPLE SCRIPT IN PYTHON

```
'''
Isaac Freedman

v0.0.2

Community Immigration Disease SIR Model

Barabasi-Albert (BA) Network

Nodal States Key:
0 = [S]uceptible, 1 = [E]xposed, 2 = [I]nfected, 3 = [I]nfected [Q]uarantined, 4 = [R]ecovered

'''

from __future__ import division
from igraph import Graph, summary
from numpy import array, asarray, zeros, arange, append, delete
from numpy import ones, reshape, sqrt, floor, ceil, amax, amin
from numpy import average, argwhere, rint, random, log, exp, linspace
```

```
        from matplotlib import pyplot as plt
        from math import pi
        from random import shuffle
        import time

        def Welcome():
            print "Epidemic Immigration Model\n"
            print "Barabasi-Albert Network Epidemic Model:"
        ##    print "10 Realization(s) on 10 Different Networks"
            print
"_____"
            print "Total nodes: N =", N
        ##    print "\nActual average degree of Social Networks: <k> =",
Total_average_degree
            print "\nInfectious period: d = 10 (days)"
            print "Probability of infection: b =", beta
            print "Immigration Rate: =", r
            print "\nImmigrant Exposed Probability: I_e =", immigrant_exposed
            print "Probability of avoidance behavior if infected and symptomatic: sh =", avoid
            print
"_____"
            return None

        def startinggraphs(ngraphs, N, m):
            G = []

            for i in range(ngraphs):
                G.append(Graph.Erdos_Renyi(n = N, m = m))
            return G
```

```
def PlotAverages(Tm, avg_epi):
    # Plot averages
    plt.title("Epidemics on Connected Communities Model\nBarabasi-Albert SIR Disease Model with Asymptomatic Individuals\nAverage of 10 Realization on 10 Different Networks\nbeta = 0.1, kavg = 10, N = 10,000, v = 0.0; asymptomatic = 0.3; beta_asymp = 0.05", fontsize = 16)

    plt.plot(range(Tm), avg_epi[:Tm,2],"y", linewidth = 3)

    plt.xlabel("Time (days)")
    plt.ylabel("Proportion of Individuals")

    plt.legend(["Infected"])

    plt.show()
    return None

def Connectivities(G, nruns, ngraphs):
    # Extract degree information from the social networks

    social = []

    S_ks = []
    S_pks = []

    # Degree information
    for i in G:
        social += i.degree()

    S_data = {k:social.count(k) for k in set(social)}
```

```
    S_pk, S_k = S_data.keys(), S_data.values()

    return S_k, S_pk

##### INITIATE ######

# Infection Parameters
N = 100

kavg = 10       # average degree

beta = 0.1      # probability of infected infecting neighbor

quarantine = 0.2

immigrant_exposed = 0.2
exposed = 0.0
symptomatics = [1]*N
avoid = 0.0

r = 0.2

m = int(N*kavg/2)

immigrants = 0

# Run Parameters
nruns = 1
ngraphs = 1

total_run_count = 0
```

```
max_time = 500

# Histories
tbavgs = []
tbmaxs = []
start_times = []
crossings = []
all_trans = []
immigration_record = []

cputimes = []
#   avg_epi = ones((max_time, 10))

# Random starting graphs
G = startinggraphs(ngraphs, N+1, m) # Start the graph with N community members and 1 immigrant

G_S_k, G_S_pk = Connectivities(G, nruns, ngraphs)
Average_S_degree = float(sum(asarray(G_S_k) * asarray(G_S_pk)))/sum(G_S_k)

#####

Welcome()

##### BODY #####

# Run the infection over ngraphs graphs...

for graphs in range(ngraphs):
   # Start the CPU clock...
   start_times.append(time.time())
```

```
# Histories
tburn = [0]*nruns

S_hist = [[0 for x in range(nruns)] for y in range(max_time)]
E_hist = [[0 for x in range(nruns)] for y in range(max_time)]
I_hist = [[0 for x in range(nruns)] for y in range(max_time)]
IQ_hist = [[0 for x in range(nruns)] for y in range(max_time)]
R_hist = [[0 for x in range(nruns)] for y in range(max_time)]

# Run the infection nruns runs...
for runs in range(nruns):
    G[graphs].vs["State"] = 0               # 0 = Suceptible; 1 = Exposed;
                                # 2 = Infected; 3 = Infected Quarantined;
                                # 4 = Recovered

    # S = Suceptible; E = Exposed; I = Infected; IQ = Infected Quarantined;
    # R = Removed

    S = range(N)
    E = [N]
    I = []
    IQ = []
    R = []
    transmissions = []

    G[graphs].vs[N-1]["State"] = 1

    ### Infect
    d = [10]*(N+1)            # Duration of Infected stage
    epsilon = [10]*(N+1)      # Duration of Exposed stage
```

```
q = [21]*(N+1)            # Duration of Quarantine

# Run infection...
for tm in range(max_time):

   immigrants = floor((tm+1)*r)     # calculate the number of immmigrants added for each run by multiplying every burnout time by the rate and rounding down

   S_hist[tm][runs] = len(S)/(N+immigrants)
   E_hist[tm][runs] = len(E)/(N+immigrants)
   I_hist[tm][runs] = len(I)/(N+immigrants)
   IQ_hist[tm][runs] = len(IQ)/(N+immigrants)
   R_hist[tm][runs] = len(R)/(N+immigrants)

   if tm % (1/r) == 0: # if the current day is divisible by the period of immigration
      new_immigrant = int(N + floor(tm*r))
      E += [new_immigrant]  # record a new individual as exposed
      G[graphs].add_vertices(1)       # add an exposed individual (immigrant)
      ### For now, exposed with constant probability.
      ### Later, import immigrants from small population with infection running on it. The new infectious will be infected with a probability that depends on time.
      if random.random() < immigrant_exposed:
         G[graphs].vs[new_immigrant]["State"] = 1 # The new immigrant is Exposed
      else:
         G[graphs].vs[new_immigrant]["State"] = 0 # The new immigrant is Suceptible
      d += [10]
      epsilon += [10]
      measured_avg_deg = sum(G[graphs].degree())/len(G[graphs].degree())
```

```
    connections = []
    new_connections = 0

    print measured_avg_deg

    while new_connections < floor(measured_avg_deg):
        for i in G[graphs].degree():
            if random.random() > i/(2*max(G[0].degree())):
                connections.append((new_immigrant,i))
        connections = list(set(connections))
        new_connections = len(connections)

    print new_connections

    G[graphs].add_edges(connections)

for ego in I:
    # Symptomatic
    for alter in G[graphs].neighbors(ego):
        if G[graphs].vs[alter]["State"] == 0 and random.random() <= beta:
            G[graphs].vs[alter]["State"] = 1
            E.append(alter)
            S.remove(alter)
            transmissions.append((ego,alter))

    if d[ego] <= 0:
        G[graphs].vs[ego]["State"] = 4
        R.append(ego)
        I.remove(ego)

    if random.random() <= quarantine:
```

```
            G[graphs].vs[ego]["State"] = 3
            I.remove(ego)
            IQ.append(ego)

        d[ego] -= 1

    for ego in IQ:
        # Infected Quarantined
        q[ego] -= 1
        d[ego] -= 1

        if q[ego] <= 0 and d[ego] <= 0:
            G[graphs].vs[ego]["State"] = 4
            IQ.remove(ego)
            R.append(ego)
        elif q[ego] <= 0:
            G[graphs].vs[ego]["State"] = 2
            IQ.remove(ego)
            I.append(ego)

    for ego in E:
        # Exposed
        if epsilon[ego] <= 0:
            G[graphs].vs[ego]["State"] = 2
            E.remove(ego)
            I.append(ego)

        epsilon[ego] -= 1

    if len(I) == 0 and len(E) == 0 and len(IQ) == 0:
        break
```

```
            tburn[runs] = tm
            all_trans.append(transmissions)
            immigration_record.append(immigrants)

        tbavgs.append(sum(tburn)/nruns)
        tbmaxs.append(max(tburn))

        print "_____"
        print "Network #", graphs + 1
        print "\nAverage burnout time: t =", tbavgs[graphs], "days."
    ##  print "Crossovers =", len(crossings)
    ##  if len(tcross) != 0:
    ##      print "Average crossover time: t =", tcrossaverages[graphs-nocrossings],
"days."
        cputimes.append(time.time() - start_times[graphs])
        print "\nComputation time: t =", cputimes[graphs], "sec."

        S_hist = asarray(S_hist)
        E_hist = asarray(E_hist)
        I_hist = asarray(I_hist)
        IQ_hist = asarray(IQ_hist)
        R_hist = asarray(R_hist)

    plt.plot(range(len(R_hist)), R_hist)
    plt.plot(range(len(R_hist)), IQ_hist)

    plt.xlabel("Time (days)")
    plt.ylabel("Proportion of Individuals")
```

```
        plt.title("Recovered and Quarantined for ER Networks with Immigration\nQuarantine
= 0.2; Immigration Rate = 0.2; beta = 0.1\nN = 100")
        plt.show()
```

BIBLIOGRAPHY

[1] Mandelbrot, Benoit B. "Introduction." The Fractal Geometry of Nature. San Francisco: W.H. Freeman, 1982. Print.

[2] Kermack, W; McKendrick, A (1991). "Contributions to the mathematical theory of epidemics—I". *Bulletin of Mathematical Biology* **53** (1–2): 33–55.

[3] Grenfell, B.T., Bjørnstad, O.N., & Kappey, J. Travelling waves and spatial hierarchies in measles epidemics. *Nature* **414**, 716-723 (2001).

[4] Riley, Steven, et al. Transmission Dynamics of the Etiological Agent of SARS in HongKong: Impact of Public Health Interventions. *Science* **300**, 1961 (2003).

[5] The WHO Rapid Pandemic Assessment Collaboration. Pandemic Potential of a Strain of Influenza A (H1N1): Early Findings. *Science* **324**, 1557 (2009).

[6] "Ebola Virus Disease Fact Sheet." WHO. Web. 2 Nov. 2014.

[7] Diekmann, O., and Hans Heesterbeek. Mathematical Tools for Understanding InfectiousDiseases Dynamics. Princeton: Princeton UP, 2013. Print.

[8] Dobson, Andy. Mathematical Models for Emerging Diseases. *Science*, **364** 6215: 1294-95 (2014).

[9] Strogatz, Steven H. Nonlinear Dynamics and Chaos with Applications to Physics, Biology, Chemistry, and Engineering. Cambridge, MA: Westview, 2000. Print.

[10] The CDC and the World Health Organization. "Smallpox: Disease, Prevention, andIntervention". Slide 16-17. Presentation.

[11] Wallinga J, Teunis P (2004). "Different epidemic curves for severe acute respiratorysyndrome reveal similar impacts of control measures". Am. J. Epidemiol. 160 (6): 509–16.

[12] Mills, Christina E., Robins, James M., & Lipsitch, Marc. Transmissibility of 1918 pandemic influenza. *Nature* **432**, 904-906 (2004).

[13] Althaus, Christian L. (2014). "Estimating the Reproduction Number of Ebola Virus (EBOV) During the 2014 Outbreak in West Africa". PLoS Currents.

[14] Newman, M. E. J. Networks: An Introduction. Oxford: Oxford UP, 2010. Print.

[15] Albert R, Jeong, H, Barabási A. Diameter of the World-Wide Web. *Nature* **401** 130-131 (1999).

[16] Dickison, M., Havlin, S., and Stanley, H. E. Epidemics on Interconnected Networks. *Physical Review E* 85, 066109 (2012).

[17] Hebert, P. L., Frick, K. D., Kane, R. L., McBean, A. M. The Causes of Racial and Ethnic Differences in Influenza Vaccination Rates Among Elderly Medicare Beneficiaries. *HSR: Health Services Research* **40**:2 (April 2005).

[18] Schneider, Eric C., Cleary, Paul D., Zaslavsky, Alan M., Epstein, Arnold M. Racial Disparity in Influenza Vaccination. JAMA **286** 12, 1455-1460 (2001).

[19] O'Malley, A. S., Forrest, C. B. Immunization Disparities in Older Americans. Am J Prev Med 31(**2**): 150-158 (2006).

[20] Yoneyama, Teruhiko, Sanmay, Das, and Mukkai, Krishnamoorthy. A Hybrid Model for Disease Spread and an Application to the SARS Pandemic. ArXiv (2010).

[21] Einstein, A. (1905). "Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen". Annalen der Physik (in German) 322 (8): 549–560.

[22] Einstein, A. (1956). Investigations on the Theory of Brownian Movement.

[23] Wu Xiao-Yan, Liu Zong-Hua. Epidemic Diffusion on Complex Networks. Chin. Phys. Lett. **24** 4, 1118-1121 (2007).

[24] Jing Xing-Li, Ling Xiang, Hu Mao-Bin, Shi Qing. Random Walks on Deterministic Weighted Scale-Free Small-World Networks with a Perfect Trap. Chin. Phys. Lett. **31** 8, 080504-1-4 (2014).

[25] WHO Ebola Response Team. Ebola Virus Disease in West Africa–The First 9 Months of the Epidemic and Forward Projections. NEJM **371** 16, 1481-95 (2014).

[26] Wearing, Helen J., Rohani Pejman, Keeling, Matt J. Appropriate Models for the Management of Infectious Diseases. PLoS **2** 7, 0621-27 (2005).

[27] Faye, O., *et al.* Chains of transmission and control of Evola virus disease in Conakry, Guinea, in 2014: an observational study. Lancet Infect Dis. (2015).