Spring 5-2015

Using Community Structure Networks to Model Heterogeneous Mixing in Epidemics, and a Potential Application to HIV in Washington, D.C.

Katherine Ragland Paulson
Bates College, kpaulson@bates.edu

Using Community Structure Networks to Model Heterogeneous Mixing in Epidemics, and a Potential Application to HIV in Washington, D.C.

Katherine Paulson

Department of Mathematics, Bates College, Lewiston, ME 04240

Using Community Structure Networks to Model Heterogeneous Mixing in Epidemics, and a Potential Application to HIV in Washington, D.C.

An Honors Thesis
Presented to the Department of Mathematics
Bates College
in partial fulfillment of the requirements for the
Degree of Bachelor of Science
by
Katherine Paulson
Lewiston, Maine
March 30, 2015

Contents

Acknowledgments

I would like to begin by thanking my thesis adviser, Meredith Greer, for her continued support and guidance throughout this process. Additional thanks go to my thesis writing peers in the math department — Ty Daly, Alison Kimball, and Henry Schwab — for sharing thoughts, edits, and friendship, and to Pallavi Jayawant and Chip Ross for their support early on. Many thanks to other faculty who have expressed interest in my thesis or who have otherwise guided my academic and research development during my career at Bates College — Karen Palin, Adriana Salerno, and Larissa Williams. Finally, a huge thank you to my incredible friends and family who have shared their interest in my work, and even sat down to listen to me practice presenting. What makes me most excited about my work is the possibility of sharing my ideas with those that I care about, and I could not have made it through this year with as much continued enthusiasm without you by my side.

Introduction

Mathematical modeling is extremely useful in the field of epidemiology. Mathematicians use models to better understand and analyze the factors that influence the dynamic spread of infectious disease through a population. A model can help predict whether a disease will spread through a population, stabilize, or die out. In other cases, models are fit to data for a past epidemic in order to estimate parameters that governed that epidemic. Models can also help mathematicians to study the effectiveness of public health measures in limiting an infectious disease. No single model works for all epidemics — each model makes its own assumptions and accounts for unique complications appropriately. Therefore, although many existing epidemiological models are based upon a similar foundation, the creation of new models is a diverse and ongoing field of research with varied and important applications.

One of the most fundamental epidemiological math models is the SIR model, originally proposed by Kermack and McKendrick [**21**]. In this model individuals in a population are categorized into groups: Susceptible (S), Infected (I), and Removed (R). The interpretation of the removed compartment varies with the disease being modeled, but typically it implies that an infected individual either died or recovered with immunity. The result is that individuals in the removed compartment cannot spread disease and they also cannot be infected with the disease — hence they are removed from the system. In the case of many models for the spread of Human Immunodeficiency Virus (HIV), the removed category is assumed to be individuals with Acquired Immunodeficiency Syndrome (AIDS), under the assumption that once HIV infection has progressed to AIDS a person is severely sick and unlikely to be sexually active.

To analyze the flow of people from one compartment to another using the SIR model, let β represent the rate at which susceptible individuals have contact with others sufficient for the spread of disease,

and ν represent the rate of removal for a single infective. This flow between compartments is represented in Figure 0.1.

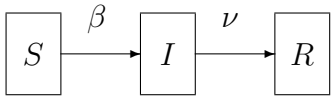

FIGURE 0.1. Compartmental diagram for SIR model.

We can formulate differential equations, as follows, to represent the change in the sizes of the S, I, and R populations:

$$\begin{aligned}\frac{dS}{dt} &= -\beta IS \\ \frac{dI}{dt} &= \beta IS - \nu I \\ \frac{dR}{dt} &= \nu I.\end{aligned}$$

From these equations, the predicted course of the disease through the population may be graphed over time. Furthermore, these differential equations can be used to solve for an expression for the basic reproduction number, $\mathcal{R}_0$. This value tells us, on average, how many secondary infections are caused by one infected individual introduced into an otherwise susceptible population. The basic reproduction number is an important figure in the discussion of how contagious an infection is, and also in the discussion of how quickly or effectively a disease might spread through a population. In particular, we may find that an infection is likely to die out if $\mathcal{R}_0 < 1$ and likely to stabilize or become epidemic if $\mathcal{R}_0 > 1$.

These SIR models are most straightforward when used to model diseases that are spread by air-born particles or by contact, such as the flu. The models become more complicated for sexually transmitted diseases because not all people in a population are at equal risk of infection. Instead, risk depends on many factors including sexual activity, partner selection, and tendency to use protection. When building models for the spread of HIV, one must also consider that sexual activity is not the only mode of transmission; HIV may also be spread by injection drug use or through breast feeding, for example. Hyman and Stanley review a range of methods useful in modeling HIV [**17**]. Many HIV models are an adaptation of the SIR model, often with

the addition of new compartments or with new parameters surrounding transmission. An HIV model by Boily et al., for example, has four compartments of infectives and four compartments of removed individuals (with AIDS) — accounting for those who are or are not coinfected with another sexually transmitted disease (STD) and those who are or are not being treated with Antiretroviral Therapy (ART) [**5**]. Compartmental separation based on STD coinfection and ART treatment allows the model to incorporate the impact of these two conditions on both susceptibility and time progression from HIV to AIDS. Many models are also simplified to a homosexual population where HIV is only spread through sexual activity. Punyacharoensin, et al. review a range of mathematical models for HIV in men who have sex with men (MSM) [**38**]. Each mathematician modeling HIV dynamics chooses his or her own perspective on the topic and offers different variations, and the diverse possibilities are part of what makes the field exciting and rapidly growing.

No one model can account for every intricacy of the spread of a disease. However, many models have been built to account for different nuances surrounding HIV transmission dynamics including varied sexual activity levels, STD coinfection, condom usage, education, partnership duration and concurrency, and ART treatment [**38**]. One such nuance is heterogeneous mixing. Many SIR-like models for HIV assume that each member of a population is equally likely to interact with any other member of the population. However, in reality, there are many factors — such as race, neighborhood, profession, socio-economic status, religion, education, and more — that impact how likely one person is to interact with another. One increasingly popular way to analyze the effects of heterogeneous mixing is by using networks. In this type of model, a vertex represents an individual, and an edge represents a relationship between individuals. Work has been done by Newman, Brauer, Rothenberg, and Miller, among others, to analyze concepts like final epidemic size and $\mathcal{R}_0$ in random networks [**32**] [**6**] [**40**] [**29**]. The first chapter of this paper is dedicated to understanding these models on random networks. Other researchers have used computer programming and agent-based analysis to simulate the spread of disease along a known network. Additionally, many recent papers have developed SIR-like models that can represent epidemic spread through a network in either discrete stochastic (probability-based) time steps or in a more traditional deterministic manner. The second and third chapters of this thesis will explore and develop an SIR-like model like

this, such that the change in infectives can be observed across time.

One framework through which to consider heterogeneous mixing is the hypothesis that people preferentially interact with other people of their race and in their neighborhood. In this paper I represent this hypothesis by building and analysing a network divided into sub-populations, where there is a higher density of edges within a sub-population than between sub-populations. This type of division in a network is called *community structure*. An interesting case where this type of analysis might be applicable is in the case of HIV in Washington, D.C.. The District of Columbia exists in one of the world's wealthiest nations, yet 2.7 percent of the city's population is living with HIV/AIDS as of 2012, a level comparable to parts of Sub-Saharan Africa, and defined as epidemic [**15**]. If you focus on sub-populations within the city (such as the MSM population or the black-male population) there is an even higher incidence. Figure 0.2 from the Kaiser Family Foundation [**15**] demonstrates the distribution by racial/ethnic groups. Data from the District of Columbia Department of Health shows the density of HIV positive individuals by ward in 2012, with Ward 6 and Ward 8 having highest prevalence at 2772.9 and 3057.5 living HIV cases per 100,000 adults, respectively, and Ward 3 having the lowest prevalence at 387.5 living HIV cases per 100,000 adults [**12**]. There are likely many reasons why disparity exists, but it is possible that sexual contact networks play a role, and exploring network models will allow us to examine this possibility further.

This research seeks to understand how we may use networks to model epidemics, more specifically HIV epidemics. It also seeks to take advantage of one of the large benefits of network models — the ability to model heterogeneous mixing — to analyze heavy prevalence of HIV in certain sub-populations in Washington, D.C.. While substantiated predictions cannot be made without detailed data regarding sexual activity in the city, it is possible to observe trends from the model's predictions. We can see how community structure in a network may impact initial and long term behavior of an epidemic, in particular how it relates to the perpetuation of disparity in prevalence between communities.

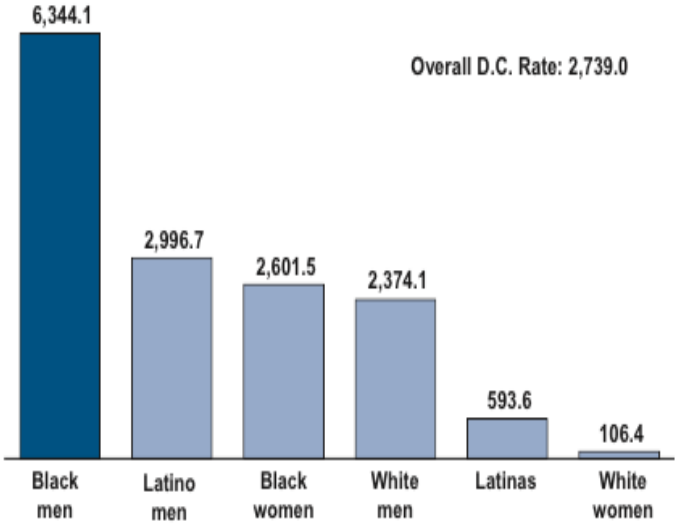

FIGURE 0.2. HIV cases per $100,000$ adults/adolescents in Washington, D.C. in 2010, by gender and race/ethnicity [**15**].

CHAPTER 1

Epidemiology in Random Networks

A network is a mathematical way of representing connections between elements in a set. A network can be used to visualize friendships between people, flights between cities, predator/prey relationships between animals, and much more. It is easy to see how networks might help us to understand epidemics, which consider the spread of disease between people.

In this chapter I will review the foundations of network theory, and then outline previous work that has been done by Brauer [**6**] and Miller [**29**], to analyze epidemics through random networks.

1. The fundamentals of networks and graph theory

A more formal definition of a network, also known as a graph, goes as follows.

Definition. A **network**, G, consists of two finite sets — the **vertex set** (denoted $V(G)$) and the **edge set** (denoted E(G)). We write, $G = (V(G), E(G))$. The vertex set contains elements called **vertices**, or **nodes**. The elements in the edge set are called **edges**, and an edge is written as a pair, $\{u, v\}$, where $u, v \in V(G)$.

For the purpose of this work, we will restrict our scope to simple graphs — graphs in which there may be no more than one edge between two distinct vertices and no edge may connect a vertex to itself in a loop. Note that when two nodes in a network are joined by an edge we say that they are *neighbors*, or alternatively that they are *adjacent*, and that when an edge has an endpoint at a particular node we say that the edge is *incident* with that node. Now, it helps to define a few more basic concepts of networks, the degree of a vertex and the degree sequence of a graph.

DEFINITION. The **degree** of a vertex, u, is the number edges incident with u. In the case of simple graphs, the degree is also equal to the number of vertices adjacent to u.

DEFINITION. The **degree sequence** of a network, G, with n vertices, is written as $d = (d_1, d_2, ..., d_n)$ where each d_i is the degree of a vertex in G, and the degrees are written in non-increasing order.

Another concept that provides us with information about the degrees represented in a graph, and that is particularly relevant to the analysis of epidemics, is called the degree distribution. We can define it as follows.

DEFINITION. Consider a network G. Let p_k represent the portion of vertices in G with degree k. Note that p_k is also equal to the probability that a randomly selected vertex in G will have degree k. Then the set of p_k values is the **degree distribution** of G. In many cases the degree distribution may be represented with a function, such as one of many popular probability distributions.

These definitions and network properties are illustrated with the sample network below in Figure 1.1.

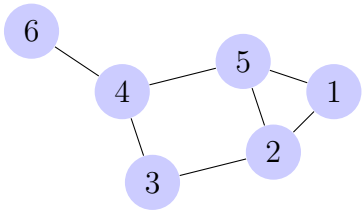

FIGURE 1.1. Graph, G, with
vertex set $V(G) = \{1, 2, 3, 4, 5, 6\}$;
edge set $E(G) = \{\{1, 2\}, \{1, 5\}, \{2, 3\}, \{2, 5\}, \{3, 4\}, \{4, 5\}, \{4, 6\}\}$;
degree sequence $d = (3, 3, 3, 2, 2, 1)$;
and degree distribution $p_1 = \frac{1}{6}$, $p_2 = \frac{1}{3}$, $p_3 = \frac{1}{2}$.

In networks representing the spread of disease, degree and degree distribution have several important interpretations. Most models allow all edges to be contacts through which disease has the potential to spread. In airborne infections like the flu, this would mean that an edge connects person a to person b if a and b interact — they are friends, family, co-workers, etc. Then, the degree of a vertex represents

the number of individuals with whom a person interacts. In the case of sexually transmitted diseases, this could mean any sexual interaction sufficient to spread disease. For simplicity, we could say that an edge exists between each pair of sexual partners in the network. Then, the degree of a vertex is equivalent to the number of sexual partners a person has. We can see that another reason why using networks to study epidemics is beneficial is that it allows us to restrict the number of contacts that any individual has, even as the population grows, and it also allows us to account for variety in the number of contacts each individual has. These points are particularly relevant to sexually transmitted disease, as the density or size of a population is unlikely to have a huge impact of the number of sexual partners any individual has, and not every person is equally sexually active.

2. The configuration model: generating a random network

Many mathematicians who study epidemics on networks, such as Brauer and Miller, start with a population represented by a network of n vertices with a known degree distribution. From this information, a random graph can be generated in two comparable ways. The first is Newman's configuration model [**33**]. Begin with a graph G, with n vertices and no edges. Next, assign each vertex degree k with probability p_k. If the sum of the degrees of the vertices is odd as a result of this randomization, the degrees must be re-chosen until the sum of the degrees is even, because each edge in a graph contributes two to the sum of the degrees and thus the sum of the degrees in a graph must always be even. The phenomenon that the sum of the degrees is even is commonly referred to as the *handshake theorem*. Now, add k spokes to each vertex with degree k, for every integer k. These spokes can be thought of as half-edges. After the spokes are added to the vertices, they are randomly connected to one another to build edges between vertices. The product is a randomly generated network on a given set of vertices with an assigned degree distribution.

We can also think of this random graph generation in another way, laid out by Miller [**29**]. In this method, vertices are again assigned degrees randomly according to the degree distribution. Then, each vertex is placed into a list k times, where k is the degree of the vertex. The order of the list is randomized and an edge is placed between each node in positions $2n$ and $2n + 1$ for $n = 0, 1...$ with the result being a randomly generated graph equivalent to the one built in Newman's

configuration model. This second way of thinking about the random generation is useful because it lends itself to computer implementation with an algorithm.

Each of these forms of random graph generation might create a graph with loops, or with multiple edges between a pair of vertices. However, in large populations where the degrees of vertices are relatively small, the probability of generating a loop or a duplicate edge is negligible. This is the case for networks used to model most epidemics, because typically we are considering networks of populations where an individual has few contacts relative to population size. Imagine the case of sexually transmitted diseases models for major cities — a person may have no, very few, or many sexual partners in a given period of time, but it is likely that any individual will not have a high number of sexual partners relative to the total population of the city. Therefore, in the context of our work, we can ignore the possibility of loop or duplicate edge generation and assume that we are working with a simple graph.

This network is generated randomly according to our degree distribution, and consequently it represents all possible outcomes from the randomization at once, and we can solve for properties of the network accordingly using *probability generating functions*.

3. Analysis of $\mathcal{R}_0$ using generating functions

Now, given a degree distribution for our population, and following analysis put forth by Brauer [**6**], we can come to an interesting interpretation of the basic reproduction number, $\mathcal{R}_0$, in networks.

To begin, define the probability generating function for this degree distribution as follows:

$$G_0(z) = p_0 + p_1 z + p_2 z^2 + p_3 z^3 + ... = \sum_{k=0}^{\infty} p_k z^k.$$

This generating function does not provide much useful information on its own, but it can be mathematically manipulated to produce new functions which may be valuable analytical tools. For example, if we evaluate the kth derivative of G_0 at 0 and divide by $k!$ then the output

is equivalent to p_k. Formally, we have:

$$p_k = \frac{1}{k!}\frac{d^k G_0}{dz^k}|_{z=0}.$$

Another piece of information accessible by manipulating the generating function is the mean degree of the network, denoted $<k>$. The mean degree is equal to the sum of degrees divided by n, the number of vertices in the network. If we let n_k denote the number of vertices with degree k, then we can see that the mean degree can be calculated as follows:

$$<k>= \frac{\sum_{v\in V} deg(v)}{n} = \frac{\sum_{k=0}^{\infty} kn_k}{n} = \sum_{k=0}^{\infty}\frac{kn_k}{n} = \sum_{k=0}^{\infty} kp_k.$$

Recall that we defined the generating function to be:

$$G_0(z) = \sum_{k=0}^{\infty} p_k z^k.$$

So consider the first derivative of the generating function:

$$G_0'(z) = \sum_{k=0}^{\infty} kp_k z^{k-1}.$$

Now evaluate this derivative at $z = 1$ to get:

$$G_0'(1) = \sum_{k=0}^{\infty} kp_k 1^{k-1} = \sum_{k=0}^{\infty} kp_k =<k> .$$

Thus, if we know the degree distribution of our network, we can solve for the probability generating function of the degree distribution, and from that we can find the mean degree in our network by evaluating $G_0'(1)$.

The average degree tells us how many contacts the average person in our population has, and can help us to interpret the idea of $\mathcal{R}_0$ in the context of networks as follows. For now we will assume that infected nodes infect all of their neighbors, but transmissibility can be incorporated into this analysis later to allow for the probability that disease does not spread along an edge. Suppose that an infection is introduced into a susceptible network at node x, and that individual x then infects one of his or her contacts, and call this person y. Let y have degree k. Now we would like to know how many people y will infect. Since we are in the early stages of the spread of this infection, and our network is

relatively large, assume that all of y's neighbors, with the exception of x, are susceptible. Now, y can spread the infection to $k-1$ people because he or she can infect any of her k neighbors except x (x is already infected). We will give this concept of $k-1$ the name *excess degree*, where the excess degree is the number of ways there are to leave a vertex, when tracing a path through a network, other than by going back along the way we came. Note that since we've allowed all edges in our network to represent the spread of disease, the mean excess degree is equal to the average number of people that one infected individual will infect when disease is introduced into a largely susceptible population. This is exactly our definition for the basic reproduction number $\mathcal{R}_0$. Next we will continue to follow Brauer's analysis to use our generating function to solve for an expression for the mean excess degree, $\mathcal{R}_0$, of our random network.

Following a random edge, the probability that we will come to a vertex of degree k (excess degree $k-1$) is proportional to k because the higher the degree of a vertex, the more edges are adjacent to it. Let q_{k-1} be the probability that a vertex reached by following a random edge has excess degree $k-1$. Now $q_{k-1} = ckp_k$ where c is a constant. We know that since q represents probabilities, if we sum q_{k-1} over all of the possible values of k it must equal 1. So we have,

$$1 = \sum_{k=0}^{\infty} ckp_k,$$

$$1 = c\sum_{k=0}^{\infty} kp_k,$$

$$c = \frac{1}{\sum_{k=0}^{\infty} kp_k},$$

$$c = \frac{1}{<k>},$$

and then we get that,

$$q_{k-1} = \frac{kp_k}{<k>}.$$

We can use this probability distribution to create the following probability generating function for the excess degree:

$$G_1(z) = \sum_{k=1}^{\infty} q_{k-1} z^{k-1} = \sum_{k=1}^{\infty} \frac{kp_k}{<k>} z^{k-1} = \frac{1}{<k>} \sum_{k=1}^{\infty} kp_k z^{k-1} = \frac{1}{<k>} G_0'(z).$$

In a parallel way to how we calculated the mean degree of the network, we can use $G_1(z)$ to calculate the mean excess degree, $<k_e>$.

So, $<k_e>$ is equal to the sum over the possible values for excess degree, of the value times the portion of the vertices with that excess degree. Written out this is

$$<k_e>= \sum_{k=1}^{\infty} (k-1) q_{k-1}.$$

Then by inserting our solution for q_{k-1} we get

$$<k_e>= \sum_{k=1}^{\infty} (k-1) \frac{kp_k}{<k>}.$$

Manipulating this more we arrive at

$$<k_e>= \sum_{k=1}^{\infty} (k-1) \frac{kp_k}{<k>} (1)^{k-2} = G_1'(1).$$

And thus, we can claim that $\mathcal{R}_0 = G_1'(1)$, the mean excess degree. Then, if we have chosen a probability distribution to represent the degree distribution of our network, we can solve for the generating functions, and therefore solve for $\mathcal{R}_0$.

Assume we're working in discrete time steps, or generations. In each generation, every susceptible node which has an infected neighbor becomes infected. Now we will discuss the way Brauer calculates the probability that an infection, initiated at a random vertex, will die out in n generations. We will call this probability z_n. Assume that we choose a random edge along which to introduce the infection. Now the probability that the vertex we reach has excess degree j is q_j. The probability that the infection will die out in n generations is equal to the probability that each of the j secondary infections dies out in $n-1$ generations, or $(z_{n-1})^j$. Now we have

$$z_n = \sum_{j=0}^{\infty} q_j (z_{n-1})^j = \sum_{k=1}^{\infty} q_{k-1} (z_{n-1})^{k-1} = G_1(z_{n-1}).$$

We can see how z_∞ is the probability that the infection will die out eventually. I will refer the reader to Brauer's text [**6**] for the remainder of this solution as well as a discussion of transmissibility. However, the result is that if $\mathcal{R}_0 < 1$ the infection will die out with probability of 1, and if $\mathcal{R}_0 > 1$ the infection will lead to an epidemic with probability $1 - G_0(z_\infty)$ but it will initially increase to a minor outbreak without causing a major epidemic with probability $G_0(z_\infty)$.

While this method of analyzing disease spread across our random network is interesting because of its conceptual connection to our idea of $\mathcal{R}_0$, it has limitations. It assumes that we are in the early stages of disease spread such that all of a node's neighbors are susceptible. In reality, many diseases we would like to model have already taken hold of a population such that we may have interactions between two already infected individuals. Another limitation is that it assumes that every contact spreads disease. Additionally, this model, while useful in predicting the final outcome of the disease spread, does not well handle questions surrounding what happens between patient zero and the end behavior.

4. An *SIR*-like random network model

To tackle the issue of transmission dynamics, Volz [**49**] offers a random network model where an infected node transmits disease to its neighbors independently with rate r and recovers at rate ν. Volz's model then more closely resembles an SIR model, as he works to describe the changes in the numbers of susceptibles, infectives, and removed over time. Using the configuration network and generating functions as previously described, Volz [**49**] derives a series of equations to analyze the dynamics of this system. Miller [**29**] then proposes a simplifications of Volz's equations to get the following:

$$\begin{aligned} \tfrac{dR}{dt} &= \nu I \\ S(t) &= G_0(\theta(t)) \\ I(t) &= 1 - R(t) - S(t). \end{aligned}$$

Note that $\theta(t)$ is equal to the probability that a random edge has not transmitted infection at time t.

It may also be shown that as we let the population size go to infinity and allow each vertex to be connected to every other vertex, these

equations simplify down to Kermack and McKendrick's original SIR equations [**29**]. This makes intuitive sense, since we know that the SIR model assumes any individual may infect any other individual in a fully mixed population.

While these models all begin to address the question of modeling epidemics with networks, they still treat every vertex in the network equally. There is a degree distribution to restrict the number of contacts held by the nodes such that the population is not fully mixed, but each individual has the same probability, p_k, to have degree k. Furthermore, individuals of the same degree have the same probability of having a connection with a third randomly chosen vertex. In reality, as has already been mentioned, there are many factors which influence how likely any two individuals are to interact with one another. One concept in graph theory that stems from these ideas is community structure, and this topic will be explored more deeply in the next chapter. Additionally, it is important to understand the diversity in the field of epidemic studies on networks, and so the foundations of this generating function analysis are worthy of attention. However, the remainder of this paper will largely diverge from the idea of generating functions, and instead take an approach that looks a lot more like a traditional SIR approach, only tailored to the language and properties of complex networks.

CHAPTER 2

Community Structure and Epidemics

Intuitively, we already have a good understanding that people tend to break up into groups in their environment. Our world is made up of divisions — by age, profession, gender, political affiliation, and much more. However, people may still interact with others outside of their own group, and the interactions between these groups are important because collectively the groups comprise one meta-population. In order to understand the network that represents an entire population, we must also look at the complex network structure, and one important feature of network structure is the presence or absence of these sub-populations, or *communities*. A network that has *community structure* has a vertex set which can be divided into disjoint subsets such that the density of edges within a subset is greater than the density of edges between subsets. The interpretation of community structure in the context of epidemics is that we may have collections of people within our larger population such that there are more interactions within the sub-population than between the sub-population and the remainder of the network. The result is that an infected individual is more likely to spread disease within his or her community than to a person outside of his or her community. As we will see in the remainder of this paper, the degree to which community structure is present in a network may have a profound effect on the transmission dynamics of any infectious disease across the network.

Before we address the mathematical definitions and implications behind community structure in networks, let's explore a few more examples of networks that have exhibited this type of organization. Much of the work that has been done with community structure in networks to date relates to the development of algorithms that can take network data and detect community structure. With these types of algorithms, it is possible to take a data set for nodes which may seem uniformly connected, and separate them out into groups appropriately such that visualization of community structure is possible. The results of one

such visualization are represented in the works of Moody and Newman [**30**] [**33**]. In the network, nodes represent middle school and high school students in one school system, and edges represent a friendship between two students. After a community detection algorithm was run, the nodes were colored by race, and it was determined that the communities highlighted social group segregation strongly related to race [**30**]. In another study, which also demonstrates a like-with-like phenomenon for multi-racial contact networks, a survey was taken of heterosexual couples in San Francisco, and it was found that there were many more same-race couples than mixed-race couples (Table 1) [**7**].

TABLE 1. A survey of heterosexual couples in San Francisco [**7**].

		Women			
		Black	Hispanic	White	Other
Men	Black	506	32	69	26
	Hispanic	23	308	114	38
	White	26	46	599	68
	Other	10	14	47	32

Both of these studies demonstrate how analysis of previous network data supports our hypothesis that there may be community structure by race in the sexual contact network in Washington, D.C.. Looking at these past findings then invigorates the purpose of this study — to look at the effects of community structure on epidemic trends — and connects back to the initial inspiration of HIV prevalence disparity by race and ward in Washington, D.C.. In the following sections, we will formalize our definition of community structure, and work towards the development of an SIR-like model for a random network with community structure that is adapted from the works of Kitchovitch et al. and Sattenspiel et al. [**24**] [**43**]. Then, we will continue in Chapter 3 with an analysis of the model and an application to and discussion of data from HIV in Washington, D.C..

1. Introduction to community structure in networks

Community structure in a network may be defined formally in both a weak and a strong sense. In these definitions, we will make use of the terms k_i^{in} and k_i^{out}, the intra-community and extra-community degree of node i respectively. In other words, k_i^{in} is the number of neighbors node i has in its own community and k_i^{out} is the number of neighbors

node i has outside of its own community. With these ideas in mind, let's now define strong and weak community structure.

DEFINITION. A subset, X, of a vertex set is a **community in the weak sense** if

$$\sum_{i \in X} k_i^{in}(X) > \sum_{i \in X} k_i^{out}(X)[\mathbf{24}].$$

Now we have that weak community structure tells us there is higher edge density within a community, X, than between X and the remainder of the network, because the quantity of edges joining nodes in X to other nodes in X is greater than the quantity of edges joining nodes in X to nodes outside of X. Weak community structure is important and telling on its own, but an even more powerful statement may be made if a network is found to have strong community structure.

DEFINITION. A subset, X, of a vertex set is a **community in the strong sense** if $k_i^{in}(X) > k_i^{out}(X)$ for every node i in X.

Strong community structure, then, tells us that if we select any random vertex in X it will have more neighbors in X than outside of X. For the purpose of this model, we will assume we are generating a random network with weak community structure. When we generate the network, as we will outline in Section 3 of this chapter, edges will be drawn between nodes with assigned probabilities. Therefore, we cannot be certain that every single node has more connections inside its community than outside its community once we add in this element of chance. However, we do have a way to quantify community structure, and allow for varying degrees of strength, and this quantification will be described in the next section.

2. Quantifying assortive mixing with mixing matrices

When we take a collection of nodes and assign each node to a particular type or category, such as we do when we divide a network into communities, we say that selective connectivity by type is called *assortive mixing* [**33**]. In order to quantify the magnitude of the community structure present in our network, we can use a matrix that stores information about the assortive mixing. There are a number of ways to design a mixing matrix, but we will go through a few.

Newman reviews a mixing matrix commonly used [**33**]. In this matrix, E, we let E_{ij} be the number of edges in the network which connect nodes of type i to nodes of type j. From here, we can create a normalized mixing matrix by dividing E by the number of nodes in the network. In the normalized mixing matrix, entry e_{ij} tells us the percent of edges in the network which fall between nodes of type i and nodes of type j [**33**]. While it is important to understand this mixing matrix in the context of network theory and assortive mixing, for the purposes of this study we will design a slightly different mixing matrix.

An interpretation of a mixing matrix that I will use for this paper is presented by Sattenspiel et al. [**43**]. Although Sattenspiel's work does not explicitly use the language of networks, it uses the idea of mixing to discuss the relationship between groups in a population. We will call this matrix M. In M, entry M_{ij} is the probability that a randomly selected neighbor of node u is of type j, given that u is of type i. In our model of a random network with community structure, we can create community structure by reserving a portion of a node's neighbors for inner-community connections, and then assigning the remaining neighbors randomly in proportion to community size. Let ρ be the portion of reserved inner-community links. Now, we can use ρ to define M_{ii} and M_{ij} for $i \neq j$ where n_i is the number of nodes in community i and N is the total number of nodes in the network, as follows:

$$M_{ii} = \rho + (1 - \rho)\frac{n_i}{N},$$

$$M_{ij} = (1 - \rho)\frac{n_j}{N}.$$

Using matrix M we will compose another matrix that represents slightly different probabilities about intra and extra community links. This additional matrix will be built because it provides us with information necessary to randomly generate a network with community structure through the method outlined by Kitchovitch et al. [**24**]. Let this new matrix be P, where P_{XY} is the probability that a randomly chosen node u in community X is adjacent to a randomly chosen node v in community Y. To see how matrices M and P provide us with related, but different, information, let's consider an example. If we divide Bates College campus by graduating class, and $M_{\text{junior,senior}} = 0.5$, then if we take a junior and select one of this junior's friends at random, the probability that this friend is a senior is 50 percent. If instead we are given that $P_{\text{junior,senior}} = 0.5$, this means that if we select any junior

at random and any senior at random there is a 50 percent chance that they are friends. We have already given the formula for finding M in terms of ρ, and now we can solve for P_{XY} in terms of M. We achieve this expression for P_{XY} by multiplying M_{XY} by the average degree, k, to find the expected value for the number of neighbors u will have in Y, and then dividing by the total number of nodes in Y, n_Y. Then we have the following:

$$P_{XY} = \frac{kM_{XY}}{n_Y}.$$

The resulting mixing matrices, M and P, will become hugely important in the model used in this paper. By varying ρ between 0 and 1 we can vary the degree to which assortive mixing occurs — from a completely mixed population to almost completely isolated communities — and analyze the effects of community structure on transmission dynamics.

3. Generating a network with community structure

To begin to create a model to help us understand community structure in epidemics, let's make a few definitions for a network on N nodes:

TABLE 2. Several definitions for variables used in a model for community structure.

n_X	Number of nodes in community X
k	Average degree in our network
ρ	Portion of neighbors reserved for inner-community links
M	Mixing matrix (outlined in Chapter 2.2)
P	Second mixing matrix (outlined in Chapter 2.2)
ς	Set of all communities in our network

After assigning k, ρ, and n_X for each X in ς, we can randomly construct a network in a way not too different from the configuration model described in the previous chapter. In this random graph generation we consider each pair of distinct vertices in our network and add an edge between them with an appropriate probability. Suppose we have vertices a and b in a network G. If a and b are in the same community, V, then an edge is added between them with probability P_{VV}. If a and b are in different communities, V and W respectively, then an edge is added between them with probability P_{VW}. With this

randomly generated network in mind, we can begin to analyze transmission dynamics in such a network.

4. Boundary nodes and transmission between communities

Now let's consider how a disease may spread in a network with community structure, where the network is generated by the method we have just outlined. An infection may pass between members of a shared community, or it may be transmitted along an edge that connects one community to another. However, it is not a given that all nodes in a community have the potential to spread disease to nodes in other communities. For a node to have this potential it must be what we call a *boundary node*. This concept of a boundary node was presented by Kitchovitch et al. [**24**].

DEFINITION. A node v in a community X is a **boundary node** if v has at least one neighbor, u, such that u is not a node in X.

To visualize what a boundary node is more clearly, we can observe the highlighted nodes in Figure 2.1. Note that not all the boundary nodes are circled there, only a few to portray the idea.

In order to be able to discuss transmission dynamics in the entire population, we must be able to quantify the number of boundary nodes in a community, since it is only through these boundary nodes that an infection may pass between two distinct communities. Using our defined probabilities, we may calculate the expected value for B_X, the number of boundary nodes in a community, X, as follows.

CLAIM 2.1. *In a network G where ς is the partition of G into communities, and for community X with defined values for n_X and matrix P, the expected value for the number of boundary nodes in X is*

$$B_X = n_X\Big[1 - \prod_{Y\in\varsigma, Y\neq X} (1 - P_{XY})^{n_Y}\Big] [\mathbf{24}].$$

PROOF. Let x and y be nodes in communities X and Y, respectively. Then the probability that x is adjacent to y is P_{XY}. Thus, the probability that x and y are not adjacent to one another is $1 - P_{XY}$. Next consider the probability that x is not only not adjacent to y, but it is not adjacent to any node in Y. This will equal the probability that x is not adjacent to an arbitrary node in Y raised to the number of nodes in Y. Thus, the probability that x is not adjacent to any nodes

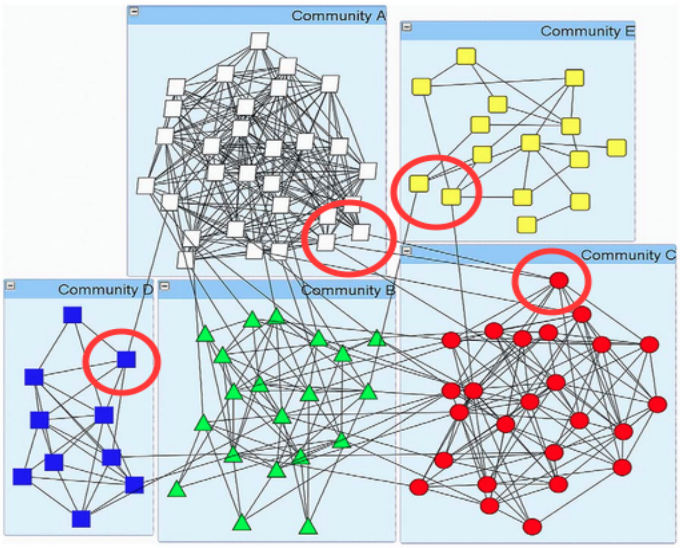

FIGURE 2.1. A few examples of boundary nodes, circled in red, in a network with community structure. The original network image was generated by Kitchovitch et al. and the red circles were added for the purposes of this thesis [**24**].

in Y is $(1 - P_{XY})^{n_Y}$.

Now we have found the probability that x is not adjacent to any node in Y, but in fact, we want to discuss the probability that x is not adjacent to any node in any community other than X. To address this we may take the product of the probabilities that x is not adjacent to a node in Y for all $Y \in \varsigma$ such that $Y \neq X$. Subtracting this value from one we arrive at the following probability that x is adjacent to at least one node outside of its community, X:

$$1 - \prod_{Y \in \varsigma, Y \neq X} (1 - P_{XY})^{n_Y}.$$

Next, by multiplying by n_X, the number of nodes in X, we get the expected number of these n_X nodes that will be adjacent to at least one node outside of X. In other words, we find the expected number of the n_X nodes in X that are boundary nodes. This number is equal to

$$n_X \left[1 - \prod_{Y \in \varsigma, Y \neq X} (1 - P_{XY})^{n_Y}\right]$$

as desired. □

This value for the number of boundary nodes will be incorporated into our SIR equations, and we will see that it decreases as we change P by strengthening community structure through increasing ρ. This trend and its implications for transmission dynamics will be explored in the next chapter. For now, we will continue to develop the foundations of our community structure SIR model by exploring the idea of force of infection.

5. Force of Infection

In order for our model to work, we need to be able to talk about how likely certain susceptible individuals are to become infected in a given period of time. We can derive an equation to aid in this discussion, and we say that this function describes the *force of infection* experienced by a susceptible individual. The force of infection will be related to both the number of infectious contacts a person has and the rate at which disease will spread along a contact between an infectious and a susceptible individual. In Kermack and McKendrick's model we have that $dS/dt = -\beta IS$. In this model, βI represents the force of infection experienced by one individual. Here, β is the rate at which individuals make contacts sufficient to spread disease. We multiply β by I — the portion of the population which is infected, and thus, the portion of contacts expected to be with infected individuals. Therefore, βI is the rate at which individuals make sufficient contact with infected individuals such that disease transmission will occur.

We will now outline what is called the *force of infection function* in Kitchovitch's SIR-like model, and then propose modifications. In this section we will use the parameters and variables as represented in Table 3.

Now, we will derive the force of infection function that comes from Kitchovitch et al. as follows.

Claim 2.2. *The force of infection, $f(k,i)$ experienced by a susceptible individual with degree k, where i is the probability that a randomly selected node is infected and τ is transmissibility, can be expressed as:*

$$f(k,i) = \sum_{s=0}^{k} \binom{k}{s} (1-[1-\tau]^s) i^s (1-i)^{k-s}.$$

Proof. Consider a susceptible node, u, with degree k. Ultimately, we are trying to solve for $P(T)$, the probability that transmission occurs between u and one of its neighbors in a unit of time. Note that

TABLE 3. Parameters and variables for use in the development of the force of infection function.

γ	recovery rate – probability that an infected node recovers in one unit of time
s_u	number of infected neighbors of vertex u
k_u	degree of vertex u
$\lambda(s)$	probability that a susceptible individual with s infected neighbors becomes infected during one unit of time
τ	transmissibility of the disease – the probability that infection will spread during one unit of time along an edge joining an infected and a susceptible individual
k_X	mean degree in community X
k_X^{in}	mean number of neighbors in X for vertices in X
k_X^{out}	mean number of neighbors outside of X for vertices in X
S_X	number of susceptible individuals in community X
I_X	number of infected individuals in community X
R_X	number of removed individuals in community X

$P(T)$ is conceptually equivalent to the force of infection, $f(k, i)$, on our node u with degree k. Let $P(s)$ represent the probability that s of u's k neighbors are infected, and let i be the probability that when we select a neighbor for u from our network the neighbor will be infected. Think of this scenario as an experiment with k trials and two possible outcomes — "success" and "failure." The probability of a "success" is i and a success is defined as a selected neighbor being infected. It is easy to see now that $P(s)$ follows a binomial probability distribution. So the probability that s of u's k neighbors are infected is $P(s) = \binom{k}{s} i^s (1-i)^{k-s}$.

Next consider the probability that u will become infected when s of its k neighbors are infected. We will denote this $P(T|s)$, the conditional probability that transmission occurs given a value for s. Let τ be the transmissibility, or probability that infection will spread along an edge between an infected and a susceptible individual. Then $1 - \tau$ is the probability that infection will not spread along one of u's s edges that are incident with infected nodes. Consequently, $(1-\tau)^s$ is the probability that infection will not spread along any of u's s edges incident to infected nodes. Thus, $P(T|s) = 1 - (1-\tau)^s$ is the probability that u will be infected by at least one of its neighbors if s of its neighbors are

infected.

From here, to find $P(T)$, the probability that u will become infected during one unit of time when u has k neighbors and i is the probability that a randomly selected node from the network is infected, we will use the law of total probability. The law of total probability allows us to say that $P(T) = \sum_{s=0}^{k} P(s)P(T|s)$. Substituting in what we have already calculated, and using the function notation of $f(k,i)$ instead of our probability notation, $P(T)$, we get that $P(T)$ is equal to

$$f(k,i) = \sum_{s=0}^{k} \binom{k}{s} (1-[1-\tau]^s) i^s (1-i)^{k-s}$$

as desired.

□

When applying this force of infection function in the community structure model, Kitchovitch et al. propose using the average k values k_X^{in} and k_X^{out} depending on whether we are accounting for the force of infection on a node from within its community or from outside its community. However, in this force of infection function k must be an integer, and so we must round the average values. While using rounded average degrees might work when modeling flu-like epidemics where degrees are large, this application becomes impractical at low k values like the ones we might expect for sexually transmitted diseases. To account for this, we will propose the following modification.

Claim 2.3. *The internal force of infection f_X^{in} and external force of infection f_X^{out} experienced by a susceptible individual in community X, where i is the probability that a randomly selected node is infected, τ is transmissibility, and p_{kXX} and p_{kXY} are the internal and external degree distributions, can be expressed as:*

$$f_X^{in} = \sum_{s=0}^{\infty} \left[(1-[1-\tau]^s) \left(\sum_{k=0}^{\infty} \binom{k}{s} i^s (1-i)^{k-s} p_{kXX} \right) \right] \text{and}$$

$$f_X^{out} = \sum_{s=0}^{\infty} \left[(1-[1-\tau]^s) \sum_{Y \in \varsigma, Y \neq X} \left(\sum_{k=0}^{\infty} \binom{k}{s} {i_Y}^s (1-i_Y)^{k-s} p_{kXY} \right) \right]$$

where the internal degree distribution, p_{kXX}, is

$$p_{kXX} = \binom{n_X - 1}{k} {P_{XX}}^k (1-P_{XX})^{n_X-1-k}$$

and the external degree distribution between communities X and Y, p_{kXY}, is

$$p_{kXY} = \binom{n_Y}{k} {P_{XY}}^k (1 - P_{XY})^{nY-k}.$$

PROOF. From the original force of infection function from Claim 2.2, we have that

$$P(T) = \sum_{s=0}^{\infty} P(T|s)P(s).$$

In this function, k is an input and so it is fixed. Now consider that we allow k to have a degree distribution such that a node has degree k with probability p_k. Now we can apply the law of total probability here a second time and substitute in for $P(s)$ appropriately to obtain,

$$P(T) = \sum_{s=0}^{\infty} \left(P(T|s) \sum_{k=0}^{\infty} P(s|k)p_k \right).$$

Next, the values which have been previously calculated may be substituted in to afford

$$P(T) = f(i) = \sum_{s=0}^{\infty} \left((1 - [1 - \tau]^s) \sum_{k=0}^{\infty} \binom{k}{s} i^s (1 - i)^{k-s} p_k \right).$$

What is missing now is the degree distribution, p_k, for this model. Let's consider the force of infection within and between communities separately. First we will find the internal degree distribution. Let X be an arbitrary community in our network and x an arbitrary node in X. When we generated our random network, we considered every node x' in X such that $x' \neq x$ and added an edge between x and x' with probability P_{XX}. We can think of this as an experiment with $n_X - 1$ trials where a "success" is the addition of an edge and the probability of success is P_{XX}. Then, the probability of k successes is the probability of x having degree k, and we have shown that it fits a binomial distribution. Thus,

$$p_{kXX} = \binom{n_X - 1}{k} {P_{XX}}^k (1 - P_{XX})^{n_X - 1 - k}.$$

Now this degree distribution can be inserted into our force of infection function, and the entire function has been defined. By inputting all of the variables it will tell us the force of infection that a single node in

community X experiences from within its community.

When we discuss the external force of infection it becomes more complicated. First consider the force of infection a node in X experiences from another arbitrary community Y. So, we will find the expected value for the number of neighbors our arbitrary node x in X will have in community Y. This follows a similar binomial distribution to the internal degree distribution of X, but now we have n_Y trials instead of $n_X - 1$ trials and our probability of success has changed to P_{XY}. Thus, we have the degree distribution for connections between communities X and Y as

$$p_{kXY} = \binom{n_Y}{k} {P_{XY}}^k (1 - P_{XY})^{n_Y - k}.$$

However, we want to know about the force of infection on x from all communities outside of its own, not just from community Y. So, when we substitute in for $P(s)$, the probability of having s infected neighbors, we will sum the expected value of infected neighbors in each community over all of the communities which are not X. The result is

$$f_X^{out} = \sum_{s=0}^{\infty} \left[(1 - [1 - \tau]^s) \sum_{Y \in \varsigma, Y \neq X} \left(\sum_{k=0}^{\infty} \binom{k}{s} i_Y^s (1 - i_Y)^{k-s} p_{kXY} \right) \right]$$

as desired.

□

One additional benefit of this proposed force of infection function is that it incorporates a degree distribution. One of the great things about using networks to model epidemics is that you can say that each person does not have an equal number of interactions. As previously outlined, this may be particularly applicable in the case of sexually transmitted diseases. However, while this force of infection function makes theoretical sense, high computational power is required to solve and analyze differential equations with this function, as there are many sums involved. Thus, I will propose the following simplification to be used for the implementation of our model.

Claim 2.4. *The external and internal force of infection in our community structure network model may be given by the following equations:*

$$f_{in}(X) = \tau P_{XX} I_X,$$

$$f_{out}(X) = \tau \sum_{Y \in \varsigma, Y \neq X} P_{XY} I_Y.$$

PROOF. First I will derive the internal force of infection function $f_{in}(X)$ for an arbitrary community X.

Again, consider an arbitrary susceptible node x in our arbitrary community X. We will find the expected value for the number of infected neighbors of x. The number of infective nodes in X is I_X and the probability that x is connected to any randomly chosen one of those I_X nodes is P_{XX}. Therefore, the expected value for the number of infective neighbors of x in X is $P_{XX} I_X$. Next if we multiply by τ, the rate at which infection spreads along an edge between a susceptible and an infected individual, we will have a force of infection function $f_{in}(X) = \tau P_{XX} I_X$, which describes the rate at which disease is transmitted to x.

Now let's consider the force of infection that our arbitrary susceptible node x experiences from outside its community. Again, we will calculate the expected value for the number of infective neighbors x will have. In this case, we will calculate the number of infective neighbors of x in an arbitrary community Y and then sum over all of the communities Y in our network such that Y is not X. The probability that x is adjacent to a randomly selected node in Y is P_{XY}. Then, by multiplying by I_Y, the number of infective nodes in Y, we obtain the expected number of these I_Y infectives which are neighbors of x. Summing over our set of communities gives the expected number of external infective neighbors of x as

$$\sum_{Y \in \varsigma, Y \neq X} P_{XY} I_Y.$$

Then, as we did before, we multiply by τ to obtain the rate at which infection spreads to x. This is our external force of infection function

$$f_{out}(X) = \tau \sum_{Y \in \varsigma, Y \neq X} P_{XY} I_Y$$

as desired.

□

Note that in these functions we are using the average degrees instead of building in a degree distribution. We decided this was a problem

before because the original force of infection function required k to be an integer, and rounding the average would skew the results. However, in these new force of infection functions we are not rounding to an integer, and so using an average should be valid. Unfortunately this method does not use a degree distribution to account for the range in connectivity, but its simplicity makes for much greater computational power, and we will build in complexities in other ways.

Also observe that we have just outlined several force of infection functions. While each one should be justified and may take into account different factors, it is important to realize that there is no one right answer, and that part of building a good model is choosing a force of infection function that seems logical and makes realistic simplifications. With this in mind, we will take the final proposed internal and external force of infection functions forward into the next section, to build our SIR-like model.

6. Building the SIR-like model

Now, using these pieces that we have laid out, we will derive equations to represent the change in susceptibles, infectives, and removed in a single community, X. We could create a set of differential equations like this for each community in the network. The SIR model, as adapted from the paper by Kitchovitch et al. [**24**], is given by these differential equations:

$$\frac{dS_X}{dt} = -\Big[S_X f_{in}(X) + \frac{B_X}{n_X} S_X[f_{in}(X) - f_{in}(X) f_{out}(X)]\Big] + \Lambda n_X - \mu S_X,$$
$$\frac{dI_X}{dt} = \Big[S_X f_{in}(X) + \frac{B_X}{n_X} S_X[f_{in}(X) - f_{in}(X) f_{out}(X)]\Big] - \gamma I_X - \mu I_X,$$
$$\frac{dR_X}{dt} = \gamma I_X - \mu R_X.$$

Our force of infection function tells us about a single susceptible node becoming infected. However, we are concerned with the dynamics of entire populations of susceptible individuals. Going back again to Kermack and McKendrick's model, we see that they gave $dS/dt = -\beta IS$. Including S, the number of susceptible individuals, along with βI, the force of infection on an individual, we get the force

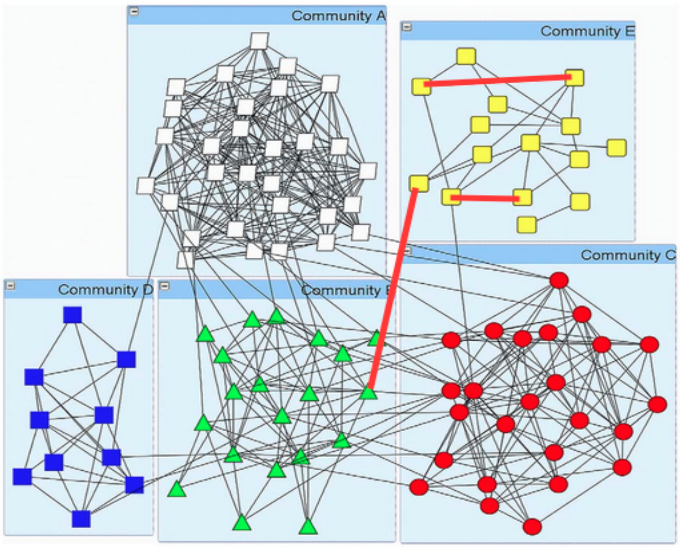

FIGURE 2.2. The three types of edges along which disease may spread: non-boundary node to non-boundary node, boundary node to intra-community node, boundary node to extra-community node. The original network image was generated by Kitchovitch et al. and the red highlights were added for the purposes of this thesis [**24**].

of infection on the whole population of susceptibles. For our network model then, Sf may represent the force of infection experienced by a whole population of susceptibles. However, to make sense of our community structure we want to think of our susceptible population in categories.

A population of susceptibles in X may lose members in several ways: a node in X may be infected by another individual in X or a boundary node may be infected by an individual not in X. Note that boundary nodes may be infected either by nodes in X or by nodes outside of X. Each of these types of links through which infection may spread is highlighted in Figure 2.2.

First consider nodes in X which are infected by other nodes in X. The number of susceptible nodes in X is S_X and the internal force of infection each of these nodes experiences is $f_{in}(X)$. This is where the $S_X f_{in}(X)$ term in our differential equations comes from.

Now consider the boundary nodes. The expression $\frac{B_X}{n_x} S_X$ is the expected value for the number of susceptible boundary nodes in X. Each of these boundary nodes experiences both an external and an internal force of infection. However, we have already accounted for boundary nodes in X being infected by other nodes in X with our $S_X f_{in}(X)$ term. Therefore, when we develop an expression for the boundary nodes leaving the susceptible group for the infective group we will subtract for the potential that boundary nodes will be infected by an internal and an external node simultaneously. In summary, our boundary node term appears as $\frac{B_X}{n_X} S_X[f_{in}(X) - f_{in}(X) f_{out}(X)]$.

Putting these two parts together we have that the rate of change of susceptibles in our community X is

$$-\Big[S_X f_{in}(X) + \frac{B_X}{n_X} S_X[f_{in}(X) - f_{in}(X) f_{out}(X)]\Big].$$

Then, just as in Kermack and McKendrick's model, the rate of change in infectives over time is the opposite of the change in susceptibles, with an added term to account for the infected individuals who become removed — γI_X.

The final component of these differential equations accounts for individuals who enter and leave the population through immigration or birth/death. In models looking at the flu, or epidemics that cover short time spans, these population dynamics may be less important because they may not change significantly over the course of an outbreak. However, when modeling HIV we may consider how prevalence changes over decades, and in this case, there is likely a significant inflow and outflow of people from a population. We will define the birth and immigration rate as Λ and assume that all incoming individuals are susceptible. To represent this inflow, we add the term Λn_x to our differential equation for S_X. Next define the death and emigration rate as μ. We will assume that susceptible, infected, and removed individuals all leave the population at the same rate, and so subtract μS_X, μI_X, and μR_X from the differential equations for S_X, I_X, and R_X respectively. The parameters Λ and μ may be chosen such that there is no net change in the size of the overall population, because the terms representing the removal of individuals may cancel with the terms representing the addition of individuals. In practice, trends in population dynamics should be taken into consideration, as some populations are increasing or decreasing in size, but this will be an acceptable assumption for the theoretical

discussion of our model.

Now that the model is set up, we can turn to a discussion of analysis and potential application. The use of computer tools to solve math models is ubiquitous in the realm of epidemiology on networks. In the remainder of this paper, computer technology will serve as a platform through which we can solve challenging equations and address complex questions about the affect of community structure on transmission dynamics and epidemic trends.

CHAPTER 3

Model Analysis and Application to HIV in Washington, D.C.

One function of modeling epidemics is to predict the impact that certain parameters, public health measures, or population characteristics may have on the spread of disease. In this chapter, I will return to my original inspiration — the HIV/AIDS epidemic in Washington, D.C. — to look at the impact of community structure in a population on the spread of disease. For this application of the model that was outlined in the previous chapter, I divided the population of D.C. into three communities by race — black, white, and Latino — and used the resulting network as an example of how community structure may impact long term HIV prevalence in the population, as well as the disparity in HIV prevalence between the racial/ethnic groups. First, however, I will demonstrate the model on a simple sample population to analyze some of its properties and discuss a few potential effects of community structure on transmission dynamics in a general epidemic.

1. Analysis of a generic community network

Consider a network made up of three communities — X, Y, and Z — each containing 100 nodes. Start with two infectives in X and one infective in Y and allow the remaining nodes to be susceptible. Let the average degree be 1.5, suppose $\gamma = 0.1$, $\tau = 0.3$, and that there is no birth, death, or immigration. By implementing our model in *Mathematica*® (see Appendix A) with these values, we obtain the progression in number of infectives over time represented by Figure 3.1.

Qualitatively, we can begin to make observations from the plots. When there is little or no community structure, each of the communities experiences a roughly equivalent rise and fall in number of infectives. As we increase ρ to increase community structure the communities begin to diverge, where each one has its maximum number of infectives at a different time point (Figure 3.1A). The result of this is that if we dramatically increase community structure by setting ρ

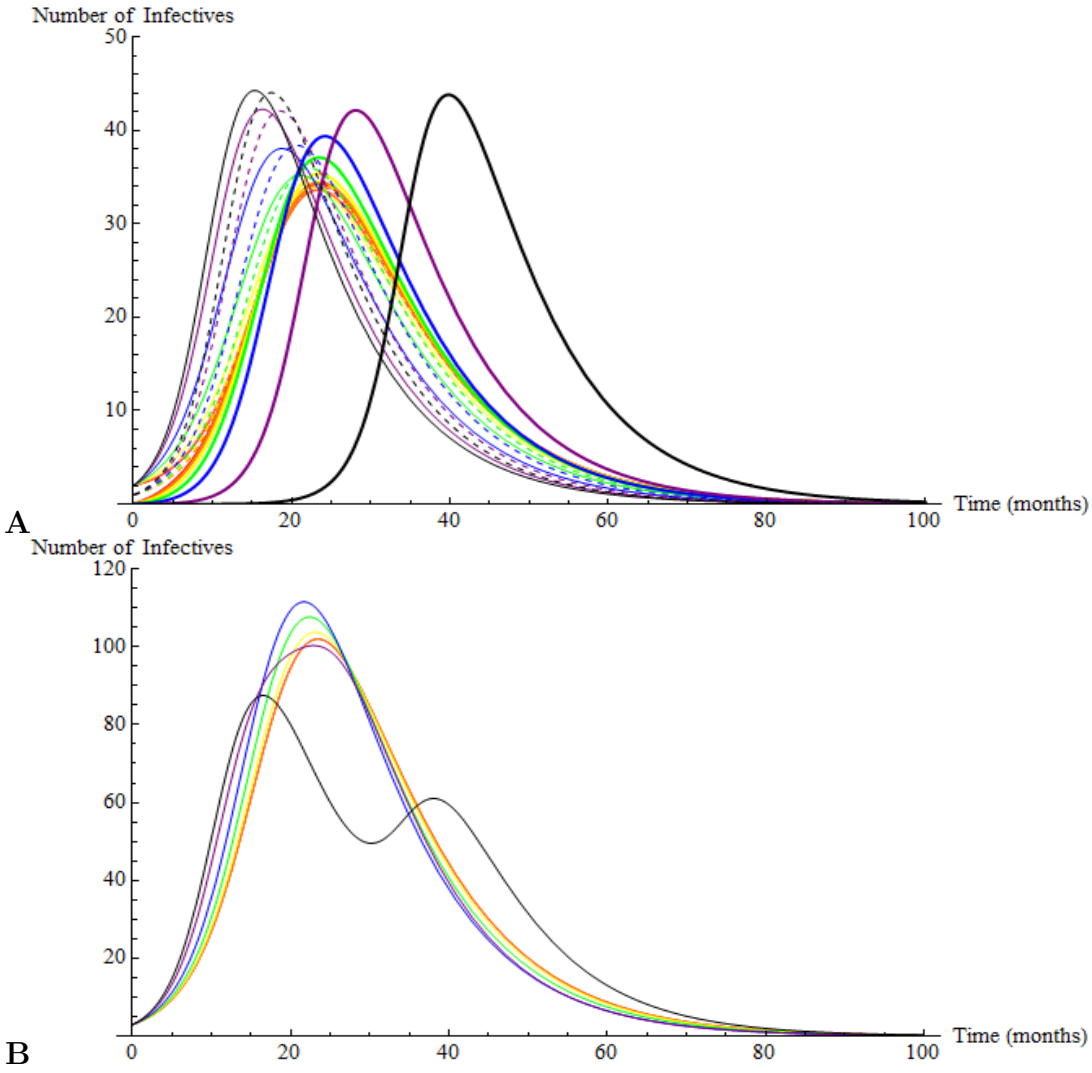

FIGURE 3.1. Number of infectives as a function of time in (A) communities X (solid), Y (dashed), and Z (thick) and (B) the entire population. The mixing parameter ρ is also varied, with red representing a fully mixed population ($\rho = 0$) and mixing progressing in chromatic order to black ($\rho = 0.99$).

to 0.99, and look at the number of infectives in the total population, we see two spikes (Figure 3.1B). One is from the two communities (X and Y) that began with infectives and the other is from community Z, which began with no infectives. It takes some time for the disease to penetrate community Z, but after this delay Z experiences a pattern of

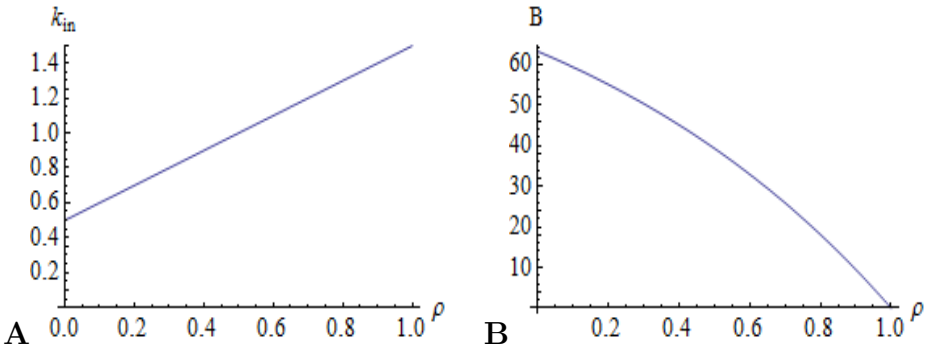

FIGURE 3.2. Average (A) internal degree, k_{in}, and (B) number of boundary nodes, B, for communities as a function of the mixing parameter ρ. Calculations are for a sample network with average degree $k = 1.5$ in three communities of 100 nodes each.

infective increase and decrease similar to that of communities X and Y.

We can calculate the average internal degree ($\bar{k}_{in}$) and the number of boundary nodes (B) in each community as a function of ρ to help us to understand these trends; these values are represented in Figure 3.2 (for *Mathematica*® notebook see Appendix B). Since communities X, Y, and Z each have the same number of nodes and the same average degree, $\bar{k}_{in}$ and B will be the same in each community.

Notice that as ρ increases (community structure strengthens) the average internal degree increases and the expected number of boundary nodes decreases. Since the average degree is the same no matter what ρ is but increasing ρ decreases the likelihood of having external connections, it makes sense that increasing ρ increases internal connections. To make this more clear, we can consider an analogous scenario. If I set out to make five friends at Bates College, and I could make a friend in any graduating class with equal probability, I would likely have fewer friends in my class than if the majority of my five new friends had to be in my class year. Since the percentage of infectives is initially higher in community X than in the entire population, increasing community structure and therefore increasing the average internal degree effectively increases the likelihood that a given interaction for a susceptible in X will be with an infective. This will cause the comparatively rapid increase in infectives for X.

The number of boundary nodes decreases with increasing ρ. This means that as we strengthen community structure there are fewer nodes in each community with external neighbors. As a result, there are fewer avenues through which the infection can spread to community Z, which started with zero infectives.

One previous study, by Sattenspiel et al., modeled HIV in a hypothetical population that was divided by levels of sexual activity [**43**]. The population consisted of 50,000 individuals with one contact per month, 33,000 with two contacts per month, 9000 with four contacts per month, 2000 with eight contacts per month, and 1000 with sixteen contacts per month. Although Sattenspiel's study uses different populations and parameters from this paper, it obtained similar results, where multiple peaks in infectives occurred in cases with extreme community structure [**43**]. It is interesting to note, however, that the model used by Sattenspiel et al. uses assortive mixing without using the language of networks.

Looking to previous work provides validation that our model is producing realistic trends. Even though the differential equations behind the Sattenspiel plots are different from our differential equations, the two models produce comparable results.

While the analysis of this simple network helps us to understand a few basic ways community structure might affect transmission dynamics, it may be more interesting to consider the application of the model to a real world epidemic. Going forward, we will look at HIV in Washington, D.C., first by considering how we might choose the parameters that organize our network and transmission through it, and second by running the model and discussing the results and their context.

2. Choosing the parameters

Much of the information that can be pulled from the analysis of a model depends on how parameters are chosen. In a model that is meant to produce predictions for a current epidemic, these values can be crucial to the integrity of the results. For this application, we will try to make educated choices for the parameter values, while recognizing the limitations to their accuracy. However, even if the parameter values do not reflect reality as closely as we may like, we can still draw

trends, if not precise numerical predictions, from the model analysis.

The parameters and values that needed to be chosen for this study were the average degree k, the birth/immigration rate Λ, the death/ emigration rate μ, the removal rate γ, the transmissibility τ, the population size of each community, and the initial distribution of susceptible, infected, and removed individuals. Each rate will go under the assumption that our time scale is by months.

The interpretation of **average degree** k is the average number of sexual partners an individual has in a given month. In most sexual networks, we can expect that there will be a few individuals who are highly sexually active, but the majority of individuals will have relatively few sexual partners. In one study, data from a 1996 survey of sexual behavior in Sweden was analyzed and it was found that the number of partners held by individuals in the population fit a power law distribution — one that behaves as I have said, with the majority of individuals having few contacts and a small number of individuals having many contacts [**28**]. In the model developed by Sattenspiel et al., as mentioned in the previous section, the population of homosexual males consisted of 50,000 individuals with one contact per month, 33,000 with two contacts per month, 9000 with four contacts per month, 2000 with eight contacts per month, and 1000 with sixteen contacts per month [**43**]. We can see that in this study the overwhelming majority of individuals in the population were assigned to one or two contacts per month, but that there is a smaller group of individuals with a relatively high number of contacts. In my analysis, I will look at average k values of 1, 1.5, and 2, to test a range and analyze the effect of k.

The **birth/immigration rate** Λ that I will use is 0.003. The birth rate in the District of Columbia as of 2010 is 15.2 births per 1000, meaning that for every 1000 people in the city, 15.2 people are born in a year [**15**]. Then, dividing 15.2 by 12,000 gives us the birth rate per month per single individual of 0.0013. Similarly, the immigration rate is available from data and can be calculated to 0.0014 [**11**]. Summing and rounding the birth and immigration rates gives a Λ value of 0.003.

This study will assume for simplification that the size of the total population is unchanging. This will allow us to assess the effects of community structure on transmission dynamics more clearly, because changes in susceptible, infective, and removed populations will be due to disease spread, not demographics. To fit this assumption, allow the

death/emigration rate μ to equal Λ, which we have said will be 0.003. Then

$$\begin{aligned}&\Lambda n_X + (-\mu S_X) + (-\mu I_X) + (-\mu R_X)\\ &= \Lambda n_X - \mu(S_X + I_X + R_X)\\ &= 0,\end{aligned}$$

so $\frac{dS_X}{dt} + \frac{dS_X}{dt} + \frac{dS_X}{dt} = 0$ and the total population size, $S_X + I_X + R_X$, is unchanging.

It is worth taking a minute, however, to discuss the population changes in Washington, D.C.. According to the United States Census Bureau, the population of the city was 601,723 in 2010, and rose to 658,893 in 2014 [**47**]. This constitutes a 9.5 percent change in population size. Furthermore, the composition of the population is changing over time. Figure 3.3 shows the racial make up of the District of Columbia over the last 200 years [**45**]. It is clear that demographics are dynamic, and strongly affected by current events, changing culture, politics, and many other factors. For example, in the 1950s the white population of the city dropped off dramatically as white people began to move to the suburbs. The black population experienced a similarly sharp decline after Martin Luther King Jr.'s assassination in 1968, when riots destroyed predominantly black neighborhoods and prompted a black migration to the suburbs [**45**]. More recently there has been a rise in Latino and Asian populations in the city. While there are current trends, and the population dynamics should not be ignored, it would be difficult to account for these changes in a long term model, since it is clear that demographics are dynamic and unpredictable.

The next parameter to be selected was the **removal rate** γ. As previously mentioned, for HIV models the removal rate translates to the rate at which HIV positive individuals progress to AIDS. In people who are HIV positive, the virus attacks the CD4 cells of the immune system, leaving the person vulnerable to infection. When significant damage has been done to the immune system such that opportunistic infections are acquired, or CD4 counts drop below 200 cells per cubic millimeter of blood (200 cells/mm^3), a person is diagnosed with Acquired Immune Deficiency Syndrome (AIDS). For reference, CD4 counts in a healthy individual are between 500 and 1600 cells/mm^3. Once a person meets these qualifications for AIDS, we can say that

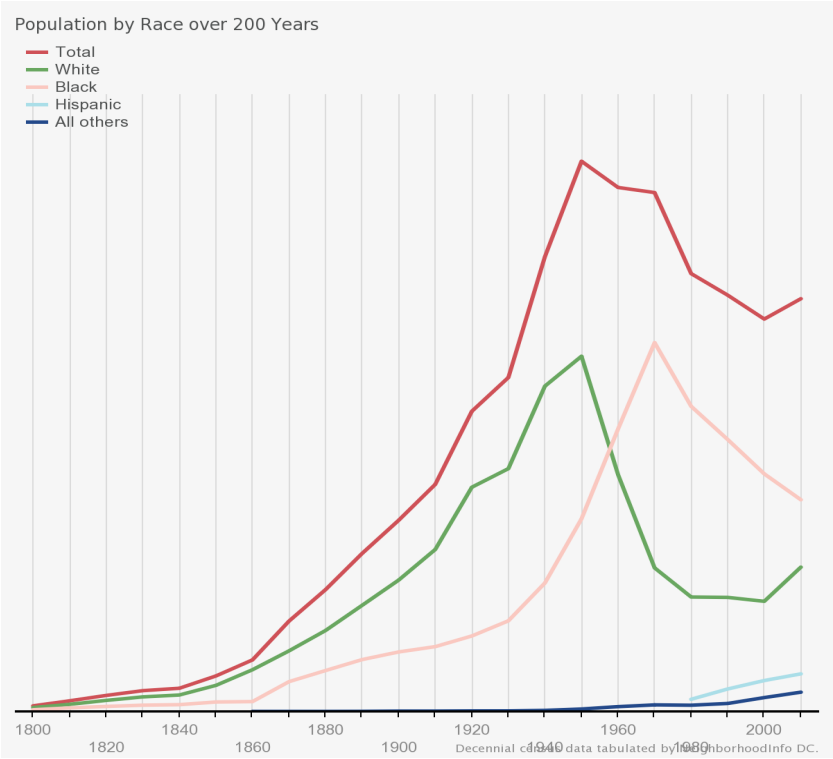

FIGURE 3.3. Dynamics of racial composition of the Washington D.C. population between 1800 and 2010 [**45**].

the person will be sick enough that they are no longer likely to be sexually active, and so will not participate in the spread of the disease. Therefore, we can place these individuals with AIDS in the removed compartment.

Now we will discuss the time between HIV diagnosis and progression to AIDS. In untreated individuals, this time is typically 10 -15 years, but can be shorter or longer [**50**] [**48**]. There are many factors that affect the progression to AIDS. Among them are genetics, diet, exercise, nutrition, age, stress, and co-infection with other viruses. Additionally, treatment with Antiretroviral Therapy (ART) will slow the progression of the disease. Therefore, early diagnosis of HIV and adherence to treatment are critical for health and life expectancy [**50**] [**48**]. For the purpose of this study the removal rate, γ, will be 0.008. This means that in a given month, 8 people per thousand infectives will develop AIDS, or 96 per year. A rate like this translates to an

average removal time per individual of just over 10 years.

Transmissibility, τ, is similarly difficult to determine, since it depends on more than just whether or not a person is HIV positive. For example, being infected with another virus makes a person more vulnerable to contracting HIV [**48**]. Treatment with ART also lowers an HIV positive individual's risk of spreading the virus to a sexual partner because it decreases the amount of virus in the body. Additionally, transmissibility through sexual contact depends on what type of sexual contact takes place. Transmission risks are estimated at 0.5-3.38 percent for anal intercourse and 0.05-0.19 percent for vaginal intercourse [**37**]. Assuming that most HIV transmission occurs from anal intercourse between men having sex with men, that a person may likely have intercourse more than once with a given sexual partner, but that condoms may be used for some interactions, we will estimate that the probability of an infective transmitting HIV to a susceptible partner in a month is one percent. Therefore, we will say τ is 0.01.

Finally, the **initial population sizes** and **distribution between susceptible, infective, and removed compartments** needs to be determined. From the United States Census Bureau, we can obtain values for the total population of Washington, D.C., and the approximate percentage of the population by race [**47**]. From the Kaiser Family Foundation we can obtain the percentage of each demographic population which is HIV positive [**15**]. Calculations of community sizes and initial S and I populations follow from there and are presented in table 1. We will assume that the initial removed population is zero. Although this is not true, what we will track is the change of invectives over time, and the number of individuals in the removed category has no impact on the rate of change of infectives ($\frac{dI}{dt}$), so we will not focus our attention on the accurate portrayal of the removed compartment.

TABLE 1. For blacks, Latinos, and whites in Washington, D.C., the estimated percent of overall population, population, percent living with HIV, size of HIV population, and size of susceptible population.

	%Population	Population	% HIV$^+$	HIV$^+$ Population	S Population
Black	50	323,000	4.3	13,890	309,110
Latino	10	64,600	1.8	1,160	63,440
White	40	258,400	1.2	3,100	255,300

With these parameters and initial conditions for HIV spread in our black, white, and Latino network communities, we can proceed to use *Mathematica*® for computer simulation of disease progression within and between the communities over time.

3. Analysis of the Washington, D.C. HIV model

Often when we discuss HIV epidemics we discuss prevalence, or the portion of a population that is infected. HIV tends to persist at a low level in populations, but there may be trends in increasing or decreasing prevalence. The goal of any public health measure would be to decrease prevalence of HIV or to improve the lives or prognosis of HIV positive individuals. Taking this into consideration, what we will focus on is the total number of infectives as a function of time. Implementation of our model and parameters in *Mathematica*® (see Appendix A) affords the following figures for the number of infectives. From these figures we can draw that both the degree of community structure, ρ, and the average degree of the nodes, k, have an important impact on transmission dynamics. Notably, community structure appears to affect the initial behavior of the spread as well as the long term behavior in number of infectives.

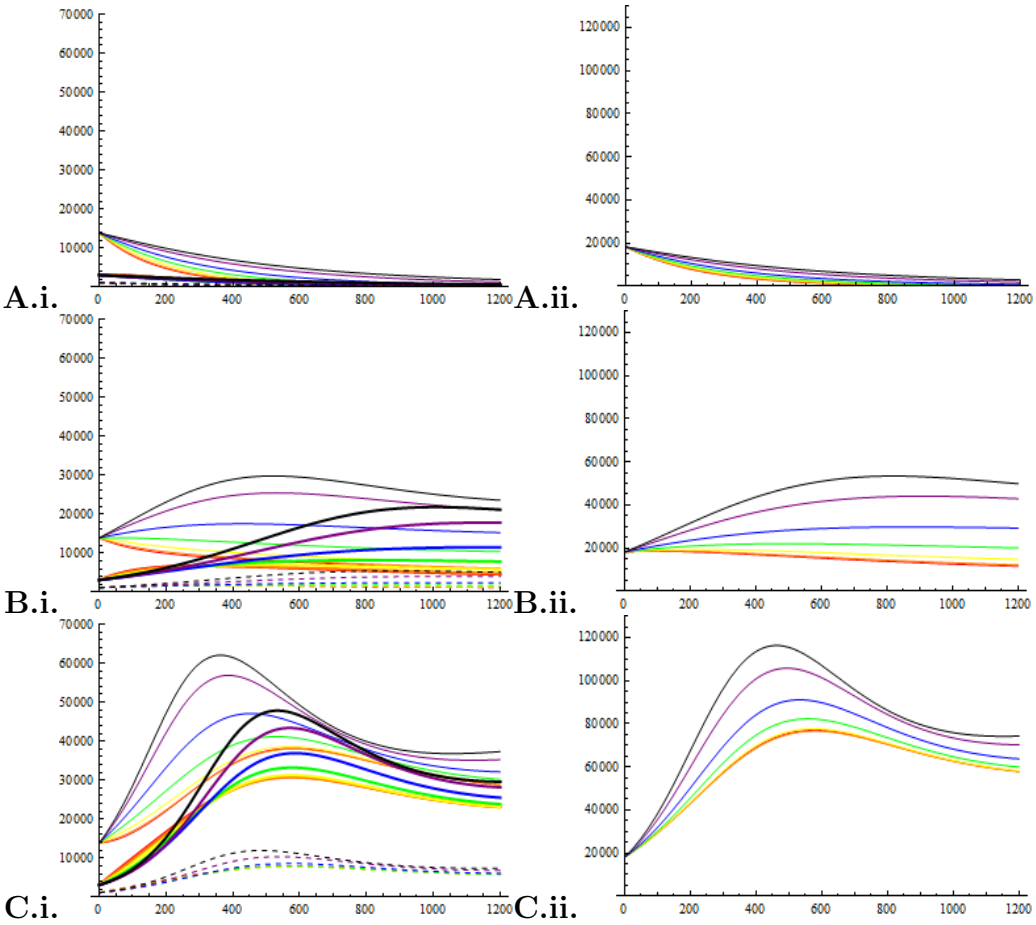

FIGURE 3.4. For (A) $k = 1$, (B) $k = 1.5$, and (C) $k = 2$, and varying values of ρ, the total number of infectives in (i) black, white, and Latino communities and (ii) the total population, as a function of months. The mixing parameter ρ takes on the values 0, 0.1, 0.3, 0.5, 0.7, 0.9, and 0.99 with 0 (red) being a fully mixed population and 0.99 (black) demonstrating the most segregated population.

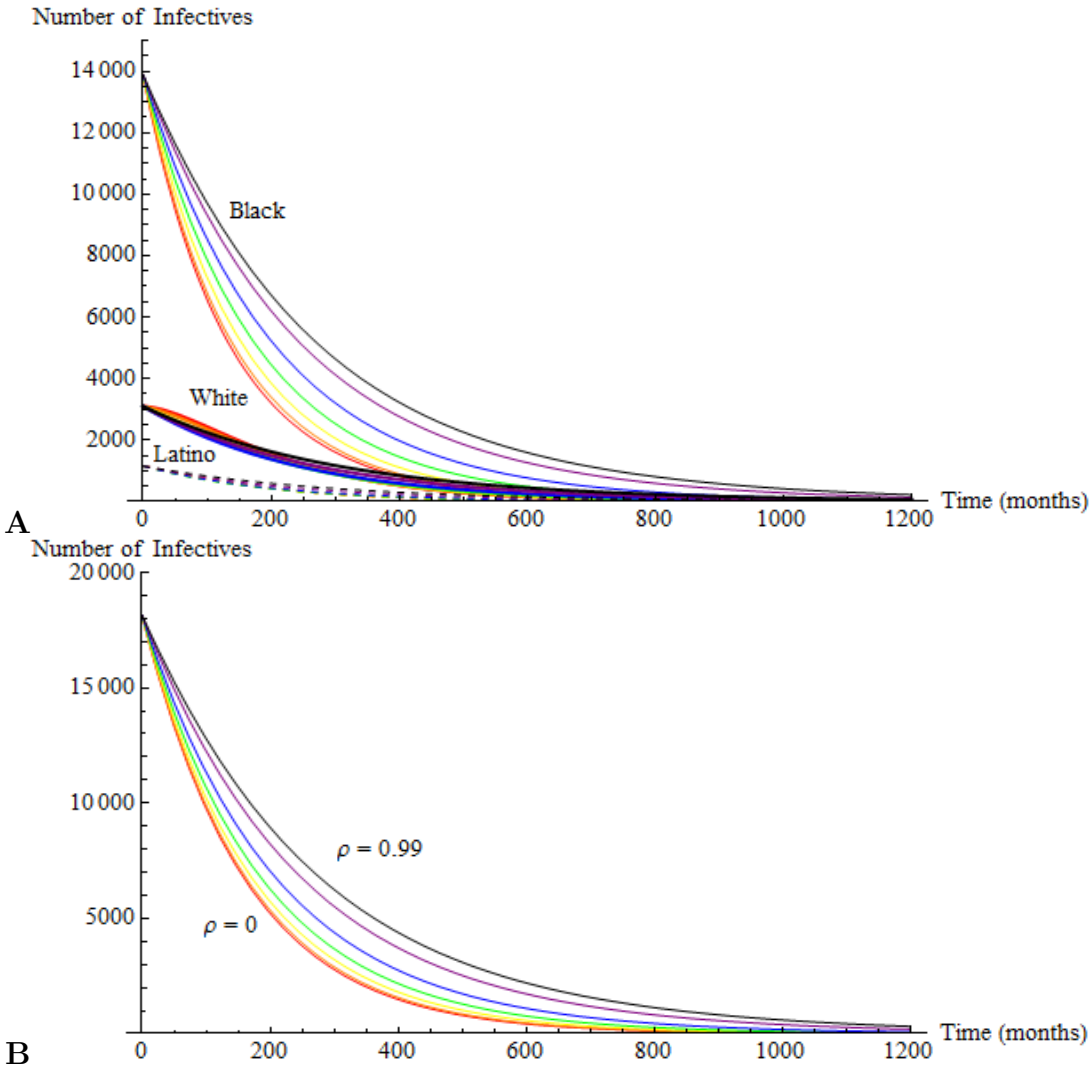

FIGURE 3.5. For $k = 1$ and varying values of ρ, the total number of infectives in (A) black, white, and Latino communities and (B) the total population, as a function of months. The mixing parameter ρ takes on the values 0, 0.1, 0.3, 0.5, 0.7, 0.9, and 0.99 with 0 (red) being a fully mixed population and 0.99 (black) demonstrating the most segregated population.

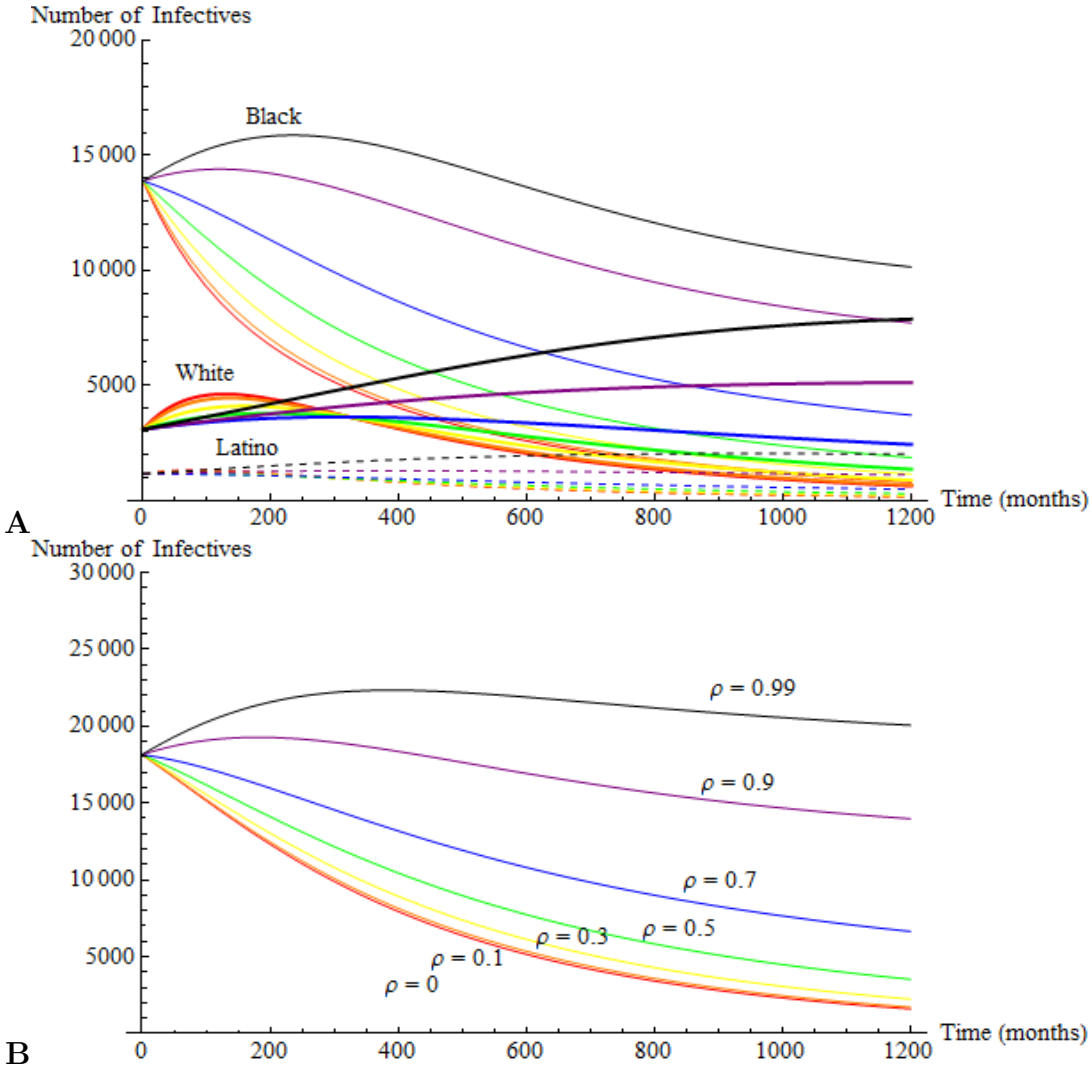

FIGURE 3.6. For $k = 1.5$ and varying values of ρ, the total number of infectives in (A) black, white, and Latino communities and (B) the total population, as a function of months. The mixing parameter ρ takes on the values 0, 0.1, 0.3, 0.5, 0.7, 0.9, and 0.99 with 0 (red) being a fully mixed population and 0.99 (black) demonstrating the most segregated population.

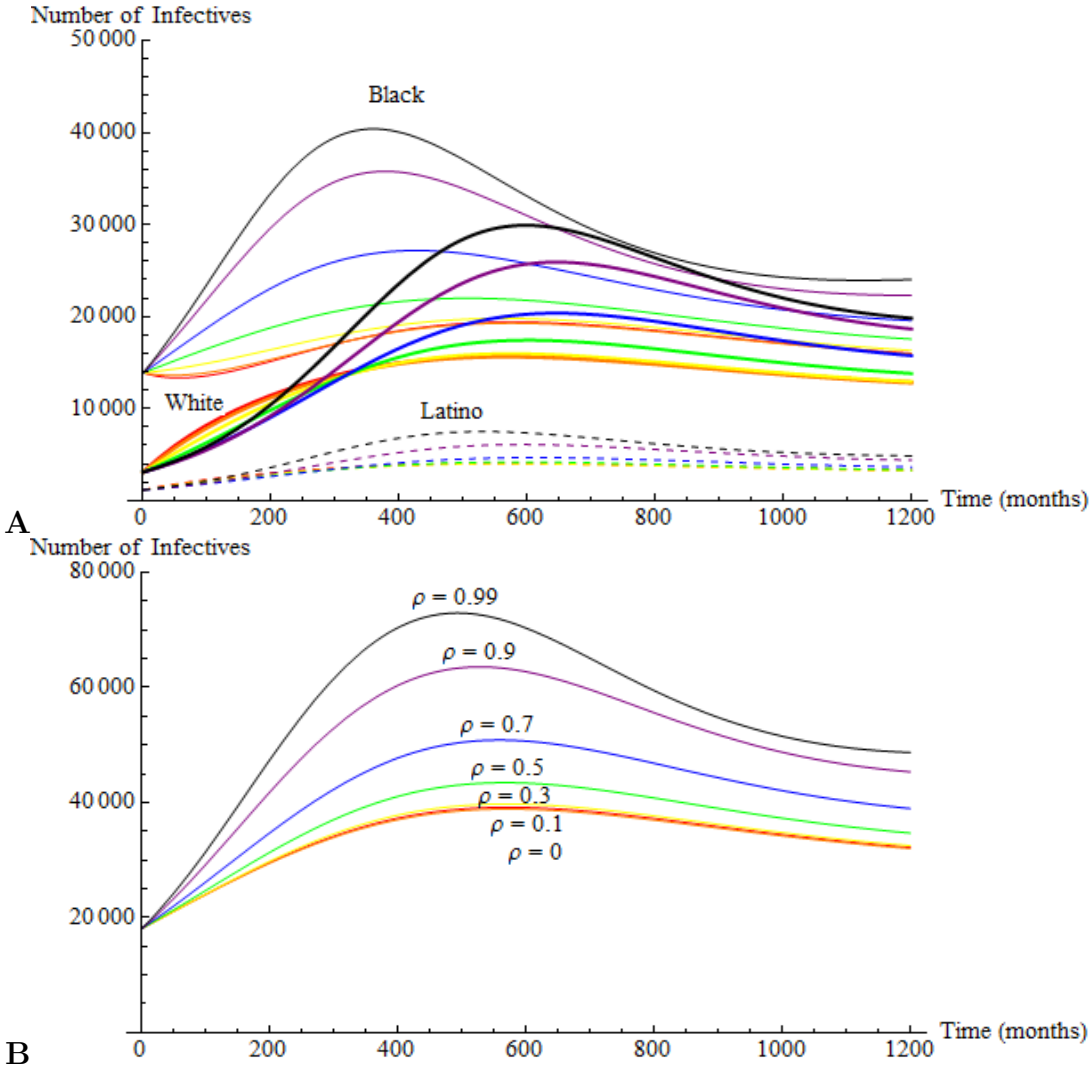

FIGURE 3.7. For $k = 2$ and varying values of ρ, the total number of infectives in (A) black, white, and Latino communities and (B) the total population, as a function of months. The mixing parameter ρ takes on the values 0, 0.1, 0.3, 0.5, 0.7, 0.9, and 0.99 with 0 (red) being a fully mixed population and 0.99 (black) demonstrating the most segregated population.

Looking at the initial behavior of each curve, one noticeable trend is that the fully mixed populations represented by the red lines show a narrowing in the gap between black and white HIV prevalence. In contrast, as community structure is strengthened we see that the trends

instead work to maintain the differences between racial groups. We can interpret these results as demonstrating that community structure may work to perpetuate prevalence disparities between communities when one community begins with a significantly higher initial prevalence than the others.

The second trend that becomes apparent is that as the curves stabilize, the number of infectives in the total population is greater when community structure is stronger. Note that in Figures 3.6B and 3.7B the number of infectives at time 1200 is strongly dependent on the mixing factor, ρ. However, we should be careful in this analysis because our end behavior is being measured at 1200 months, or 100 years. A lot can change in 100 years that could impact the dynamics of HIV transmission. New treatments could be discovered, there could be a huge shift in demographic composition, new public health measures or laws could be put in place, etc. Additionally, forecasting out 100 years may seem illogical for a disease that has only been officially recognized since the early 1980s. Still, the time scale on this model may be affected by more accurate calculations of the model parameters. Additionally, the purpose of this model is not to make accurate numerical predictions, but instead to study the trends that may arise as a consequence of community structure in disease propagation on networks.

It becomes immediately apparent from the curves that the model predicts dramatic changes when only the average degree k is varied. As we increase the average degree in our network, we allow each person to have more contacts, and this allows for more potential for disease spread. The result is that when $k = 1$ each population experiences an immediate decrease in HIV prevalence, when $k = 1.5$ the curves are less dramatic, with gradual increases or decreases in prevalence depending on time and ρ, and when $k = 2$ the overarching trend is a great increase in prevalence followed by a decrease and stabilization in prevalence (Figure 3.4). Since we expect that the prevalence in Washington, D.C., is not experiencing dramatic immediate changes, but rather a more gradual change, we are supported in our choices for k. Another thing to note is that when $k = 1.5$ (as compared to $k = 1$ and $k = 2$) we see the greatest difference in long term behavior of HIV prevalence as ρ is varied. Therefore, the model suggests that community structure may have the strongest impact on prevalence at a particular, and relatively small, value of k. It may be that as k is increased, the number of boundary nodes in our communities is large enough that spread

between communities is not inhibited. Even if there are still more connections inside a network than between networks, transmission between communities could occur freely and our curves would begin to behave most like a fully mixed population even as ρ is increased. In light of this observation, community structure analysis becomes even more relevant to sexual contact networks, since these types of networks tend to have smaller average degrees than friendship or acquaintance networks.

4. Further investigation of racial/ethnic HIV disparity

While I have posited that community structure may be present in the D.C. sexual contact network, and that this may be related to the disparity in HIV prevalence by race, there are other important social factors that may be behind why certain populations are more heavily affected than others.

Some differences between racial populations are associated with social network properties. Adimora and Schoenbach review several differences between sexual contact networks in white populations and similar networks in black populations in the United States, and how that might affect the spread of sexually transmitted diseases [**2**]. In black populations, sexual contact occurs more often between people with many partners and people with fewer partners [**26**]. Additionally, black men and women have been found to be more likely to have concurrent sexual partnerships than white men and women [**1**]. Having overlapping partnerships allows for a more rapid spread of disease, even if the total number of partners is the same [**31**]. Each of these nuances is related to network structure — who is connected to who. Patterns of this kind vary between populations and are important, as this paper demonstrates, to any analysis of disease spread, but in particular the spread of sexually transmitted diseases like HIV since sexual contact networks tend to be sparse (have less connectivity or fewer edges) and strongly affected by social factors.

A further investigation of racial disparity in HIV prevalence requires a deeper look into the underlying social factors that affect the network patterns we see. A history of institutionalized racism has left us with a population where socioeconomic status and race are linked by trends (Figure 3.8). In 2008, the median household income in Washington, D.C., was $107,600 for non-Hispanic white households, $39,200 for

black households, and $43,500 for Hispanic households [**22**]. Income inequality manifests itself in a number of other inequalities. Poorer families on average have lower quality education, experience more stress, have less access to and knowledge of healthy food options, and have less leisure time for activities and exercise. Poor sex-education and reproductive health services lead to patterns of sexual behavior that favor the spread of sexually transmitted diseases [**2**]. In addition to decreasing the overall health of an individual, chronic stress may make individuals more susceptible to HIV and may also speed up HIV disease progression after infection [**27**]. Lower socioeconomic status may be a cause of instability in relationships, and promote the behavior of taking on concurrent sexual partners [**39**]. The ability to form stable relationships is also important because sexually transmitted diseases are less likely to propagate in contact networks with long-term monogamous relationships [**25**]. For example, a network where everyone has one partner a month will spread disease differently if we assume that each month the individuals may take a different partner than if we assume that each month most individuals are remaining with the same partner as the previous month.

Racial division in social networks also reinforces cultural norms of sexual behavior that may influence the spread of sexually transmitted diseases. Geographical and relational racial segregation concentrates issues like poverty and poor education in disadvantaged groups in a parallel to the way segregation concentrates illness, in particular HIV. Geographical separation into racially homogeneous neighborhoods may also enhance community structure in sexual networks, because people are more likely to choose partners from their own neighborhood [**51**]. Our residence also decides what school district our children will go to, so it influences who adolescents interact with and what type of education they are receiving.

Additionally, racial segregation may serve to propagate stigma that can increase risky behavior and inhibit screening and treatment. Earnshaw et al. evaluate the impact of stigma on racial divides associated with HIV [**14**]. Disadvantaged communities may have less access to services for prevention and treatment, but they may also take less advantage of the services available to them. For example, medical mistrust and conspiracy theories in black populations may contribute to negative attitudes towards condoms [**4**]. As compared to whites, minorities are more likely to be tested and diagnosed at a later stage in the HIV disease progression [**8**]. In many cases, this may be because

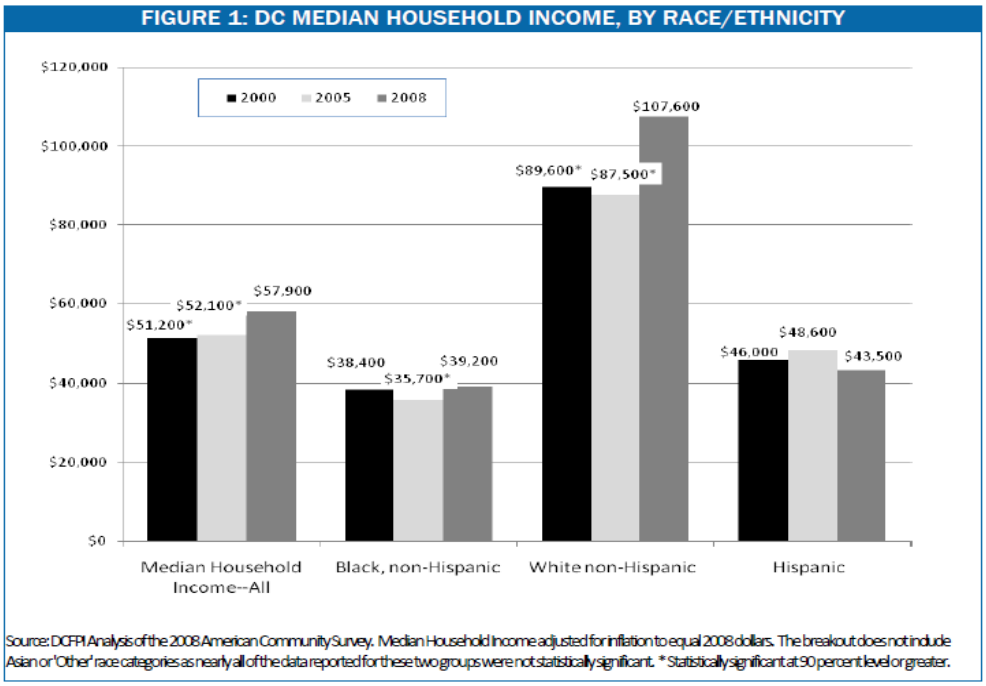

FIGURE 3.8. Median household income in Washington, D.C., in 2000, 2005, and 2008 by race/ethnicity [**22**].

stigma works as a barrier to HIV screening [**35**]. People may fear that others will judge them for being tested because of the stigma attached to the behaviors associated with HIV [**14**]. After diagnosis, we find that stigma may also be a barrier to treatment. Latinos and blacks have been found to delay treatment longer than whites [**46**]. Additionally, once care has been sought, minorities are less consistently adherent to the antiretroviral therapy they may receive [**19**]. Each of these factors combined may play a role in the trends observed in Washington, D.C. for newly diagnosed AIDS cases. While most of the cases of HIV in the district are in black individuals, the number of AIDS diagnoses for the black population is disproportionately high. This vast difference is displayed in Figure 3.9.

While differences in socioeconomic status, education, perception of stigma, access to services and treatments, and myriad other social factors may be enhanced by community structure, they could also be considered in the construction of a community structure model. Each of the upstream factors discussed could be built into the model by conferring different properties to each population, or allowing each

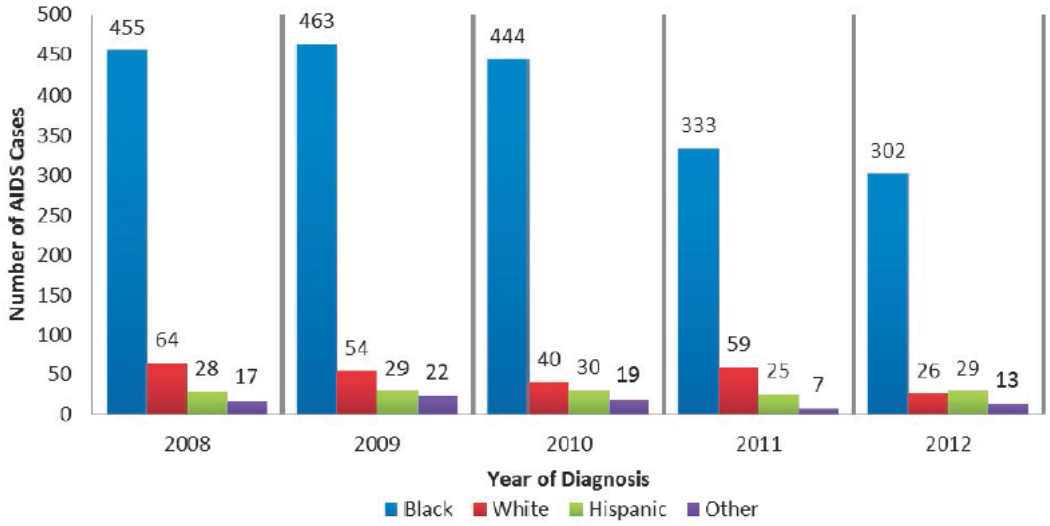

FIGURE 3.9. Newly diagnosed AIDS cases in Washington, D.C. by race [**12**].

community to take on different parameter values. For example, poor education and stigma could reduce condom usage in a community and increase risky sexual activity. The effect of these factors is essentially an increase in average degree, k. Later testing in one group may also effectively increase k or transmissibility τ, because if an infected individual knows of their status as HIV positive, they may take steps to reduce the likelihood of transmission to their partners. Additionally, with diagnosis and proper treatment, ART can reduce an individual's viral load, therefore reducing the transmissibility for any sexual interaction the individual may have. So, we may expect that τ could be varied between communities that are more or less likely to seek and adhere to treatment. The removal rate, or the rate at which HIV positive individuals progress to an AIDS diagnosis, may also vary between communities in relation to overall health and tendency to seek and adhere to treatment as discussed previously. In this paper, the parameters were kept consistent between communities, but it is conceivable that the trends predicted in my model would be different or exaggerated if different properties were conferred to each population. Varying the properties of the communities could be an interesting extension to this study, and I would suggest that it should be addressed if any serious effort were made to find and fit comprehensive real world data to this model. Some previous works, including the studies of Kitchovitch and Lió [**24**] and Sattenspiel et al. [**43**] which we have already looked at,

allow for differences between communities or even base the community division lines on these property differences, and we will look at these and select other papers as we conclude this thesis in the next chapter.

CHAPTER 4

Conclusions and Lessons from the Literature

While addressing the implications of my work, it is important to understand the context in which it is produced. Using network theory to address epidemics is a relatively new field, and incorporating ideas of community structure into the dynamics is an even newer area of study. Most of the related works available have been produced in the last decade. This means that the discussion of community structure in epidemic analysis is promising, and diverse and versatile in its applications. In this section, I will review a selection of works which are related to my model, and discuss how they may inform my process or potential future studies in the field.

Although a large part of my model is based off of the work by Kitchovitch and Lio, what I have not previously discussed is that their model introduces into the force of infection function a factor for awareness, or perceived risk, and in fact this is where most of the emphasis is placed in the analysis of the model [**24**]. This awareness factor may be related to access to information, health care, preventative measures, and more. As a consequence of awareness, an individual may engage in less risky behavior or take steps to reduce the likelihood of contracting a particular infection. The argument is that each community may take on a different value for this awareness factor, and that through conferring different properties to each community in this way we may change the way community structure impacts transmission in the population. This awareness factor could be related to each of the parameter affecting sociological concepts discussed in the previous section in relation to differences between each racial/ethnic community.

Sattenspiel et al., in the paper discussed in section 3.1, divides individuals into separate communities based on sexual activity, and investigates the effect of sexual activity on risk of infection as well as the effects of mixing between groups on individual risk and population trends [**43**]. In this way the model addresses something my model does not — the possibility of a degree distribution — and allows for the

variance of a particular property between communities. Conclusions drawn in this paper are that the initial spread of the disease and the steady-state proportion of infected individuals are both influenced by the degree of mixing [**43**].

Another study worth highlighting is the work by Kenah and Robins to use probability generating functions, in a method not unrelated to that explored in Chapter 1, to study mean outbreak size and final epidemic size in percolation networks with random and proportionate mixing [**20**]. In the discussion of proportionate mixing, the population is divided into sub-populations, but the number of neighbors a node will have in any given sub-population is proportionate to the size of the sub-population. This is unlike our model with assortive mixing, where there is a like-attracted-to-like factor. What makes the division into categories useful, then, is the ability to allow parameters such as transmissibility and recovery rates to vary between sub-populations [**20**].

While many works I have discussed generate random networks guided by particular patterns or parameters, some studies look at networks for which connection data already exists. In Cuba, a contact-tracing detection system has been in place since 1986, meaning that data for a large network of HIV transmission exists [**9**]. This study used a network of 5389 vertices and 4073 edges to study the spread of HIV on a known network for Cubans diagnosed as HIV positive between 1986 and 2006. However, instead of trying to model disease progression over time, the main goal of this paper was to examine the network properties of the existing sexual contact network. This network is interesting in contrast to my HIV network because in this case, edges only represent sexual partnerships through which HIV has been transmitted. Degree distributions, clustering/assortivity, and path lengths were studied using statistical analysis. Of note, it was determined that the network demonstrated assortive mixing with preferential connections between nodes that share geographical region, age, or mode of detection. Reverse analysis like this justifies and informs the work that we have done which presumes community structure and assortive mixing exist. Additionally, with increasing data collection and computational possibilities, this type of analysis may be promising for expansion of community structure studies in the future.

In an example which even more clearly demonstrates the power of big data and computer power, Coelho, Cruz, and Codeco present a software called Epigrass, designed to simulate epidemics on networks

[**10**]. In this software, data is stored for the state (susceptible, exposed, infected, removed) and neighbors of each node. Then, in discrete time steps, the software determines whether or not infection will spread through any given edge that connects a susceptible and an infected node. In the paper, Epigrass is demonstrated to analyze the spread of respiratory disease between 76 Brazilian cities, which we can think of as communities, by travelers along bus transportation routes. An interesting addition in this model is the incorporation of travel time, as it relates to the possibility of recovery en route [**10**]. This work is a good example of using high computational power and programming to store large data sets on known networks and simulate the spread of disease. Additionally, this application demonstrates the range of interpretations of what we think of as a community and a connection between communities. The idea of having travelers that move between geographically separated communities is a common one.

Another study that incorporates the idea of travelers is the work by Apolloni, which develops a model with community structure and non-homogeneous mixing and applies it to the case of H1N1 influenza in Europe and Mexico in 2009 [**3**]. The populations are divided by country and by age (adults and children), with adults having a higher likelihood of travelling between countries. From there, an expression is solved for the global invasion parameter $\mathcal{R}_*$, which tells about the number of communities a single infected community may infect. In general, an epidemic may fail if spatial mobility is small enough such that no travelers will make it to other sub-populations before a local outbreak ends [**3**].

Sometimes, community structure and SIR models are used to study applications where something other than disease is being spread. Huang, Park, and Lai study the spread of information through a network with modularity, or community structure, using concepts from SIR theory [**16**]. The major conclusion drawn is that as the number of modules, or communities, is increased, the lifespan of the information spread increases to a point, hits a maximum, and then begins to decrease as the number of communities is still increased. When translating this to our conception of epidemics, we could say that initially an epidemic will persist for longer in a population if the number of communities is increased, but that after a point to duration of the epidemic would decrease with increasing quantity of communities. The idea of increasing the number of communities might be an interesting addition to my model, where the number of communities represented is relatively

small.

Often, one of the goals of building a model is to be able to suggest the most efficient methods of allocating limited resources. For infectious diseases, one common resource is vaccines, and in the case of networks we can think of planned vaccination as the strategic removal of selected nodes. Salathe and Jones conclude that immunization interventions in populations with community structure may be most effective when targeted at individuals with the potential to spread infection from one community to another – individuals who fit into our understanding of boundary nodes [**41**]. This conclusion is interesting, particularly as they claim that this concentration of resources may be even more effective than a concentration on highly connected nodes. Note that this proposed intervention is not a step to increase community structure in the network, since we can still allow individuals to have the same number of internal and external links, it would be an effort to limit the potential for disease spread along those external links. The model generated in this paper is a SIR model on networks with community structure where the number of communities is large relative to the population of each community. Average epidemic size, epidemic duration, and peak prevalence are investigated. It is argued that community structure allows for persistent and low prevalence as infection spreads from one community to another and we get a series of small outbreaks spread across time [**41**]. Again, this model with a large number of smaller communities provides an intriguing contrast to my model with a small number of larger communities.

Moving in the opposite direction, some studies look at populations which are divided into only two communities. Xiao-Long Peng et al. demonstrate with a stochastic Susceptible-Infected-Susceptible (SIS) model on a network of two communities, how recurrent outbreaks and extinctions can happen in one population as a result of infection from a separate population with endemic prevalence (Figure 4.1) [**36**]. The paper gives the example of zoonotic infections, where an infectious disease may remain endemic in animal populations while causing periodic outbreaks in human populations by occasional animal to human transmission. In this case, the two communities are the animals and the humans, where disease is more likely to spread animal to animal or human to human than animal to human. The expected value for the length and frequency of outbreaks in the community where disease is emerging and re-emerging is evaluated, and it is found to be dependent on factors such as average node degree in each community,

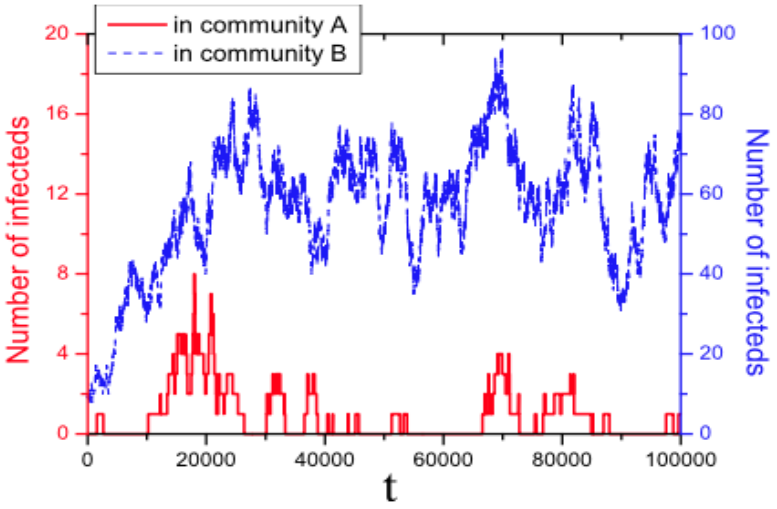

FIGURE 4.1. Stochastic Susceptible-Infected-Susceptible (SIS) model model for number of infectives as a function of time in a network of two communities [**36**].

the number of inter-community links, and other parameters such as transmission rate and recovery rate. Peng et al. take the approach of specifying the number of inter-community links in order to adjust community structure, instead of adjusting the probability that a node will take a neighbor in a community outside of its own as we did. One observation from the paper is that as the number of inter-community links is increased (or community structure is weakened), the periods of extinction get shorter as more outbreaks overlap with one another [**36**].

To underscore the exciting recent growth and usage of community structure models in epidemic analysis, we can see how one paper uses community structure to look at one of the largest and most recent global epidemics — the Ebola outbreak of 2014. Kiskowski sought to model the early dynamics of the outbreak in Guinea, Sierra Leone, and Liberia using a model with multiple dimensions of community structure [**23**]. For this model, the population was divided by households, communities of households, and countries. Households were highly connected and connections within a household were more likely to spread infection than connections between two separate households. A discrete, stochastic SEIR analysis of this network sought to explain why each of the three countries exhibited qualitatively different early growth dynamics. In particular, Ebola spread most rapidly in Liberia, then

Sierra Leone, then Guinea. The conclusion was that the variations in regional trends could be a product of mixing. Two individuals infected by the same person are more likely to share the same neighbors in a less mixed community, so local saturation effects may occur such that contacts of infected individuals are more likely to have been infected already, thus limiting the spread and causing a linear trend in infections. Exponential growth, conversely, may result from exposure to new communities. Kiskowski was able to fit World Health Organization Ebola data for each country only by varying community mixing parameters for each country, while keeping the number daily interactions, transmission rates, and R_0 the same for all three countries [**23**]. This model emphasized the importance of reducing contacts between exposed and unexposed groups, and is an exciting application of creative community structure to relevant and modern epidemics.

Just this short collection of studies demonstrates the range of methods used to analyze community structure in epidemics. Some works use stochastic models, some use deterministic. Some use existing networks, and some use randomly generated networks. Computer programs have been produced to simulate disease spread in an agent-based way that considers each node separately, while other researchers design SIR equations or even ignore compartmental dynamics and instead use probability generating functions to analyze long term behavior and other epidemic properties. As much variety as there is in methods for community structure analysis, there is in interpretation of community structure. Each model chooses a different way to segregate a population into sub-populations and the rationale behind the divisions guides the questions being addressed in the study as well as the mathematics behind the dynamic analysis of the model.

In this paper, a broad approach was taken to understand the foundation and context of any of these community structure models. Methods of network epidemic analysis with generating functions and an interpretation of $\mathcal{R}_0$ were outlined. Next, differential equations for SIR dynamics on a community structured network were proposed, with foundations in work by Kitchovitch et al. [**24**] and Sattenspiel et al. [**43**], and each component of these equations was outlined to highlight the process behind the development of a model. The model was run in a sample population, for which it was demonstrated that community structure changes the time point of maximum number of infectives in each community, and in the case of extreme community structure, peaks from different communities can be so offset such that multiple

peaks in total population prevalence are observed. It was discussed that the observed trends may be a result of the effect of the mixing parameter, ρ, on mean internal degree and number of boundary nodes. Finally, parameters were chosen to represent HIV in black, white, and Latino populations in Washington, D.C.. The curves from this trial suggest that the degree of community structure may impact initial and long term behavior of prevalence, particularly in the case where average degree is small, but just large enough for disease spread. Additionally, it may be that a high degree of community structure works to perpetuate HIV prevalence disparity between racial/ethnic groups in our population. The model put forth in this paper is then an exploration of network epidemics and heterogeneous mixing, and a part of an exciting and ongoing investigation of the possibilities that community structure analysis may provide.

Appendix A. *Mathematica*® notebook: prevalence plots

```
(* This notebook takes the model put forth in this paper, and evaluates
it for parameters and data chosen to represent HIV in Washington, D.C..
The graphs that are generated represent the change in the number of
infectives in the BLACK, WHITE, and LATINO communities, and the overall
population, over a 100 year time span for varying levels of community
structure. A similar notebook, with different parameters and population
sizes, was used to run the model on the sample network of 300 nodes
discussed in Chapter 3. *)

Clear[k, kX, kY, kZ, Lambda, Gamma, Tau, BX, BY, BZ,
  nX, nY, nZ, Ntot, P, M, Rho, tstart, tfinish, SXstart, SYstart,
  SZstart, IXstart, IYstart, IZstart, ix, iy, iz, SX, SY, SZ, InfX,
  InfY, InfZ, solSX, solSY, solSZ, solIXA, solIXB, solIXC, solIXD,
  solIXE, solIXF, solIXG, solIYA, solIYB, solIYC, solIYD, solIYE,
  solIYF, solIYG, solIZA, solIZB, solIZC, solIZD, solIZE, solIZF,
  solIZG, solRX, solRY, solYZ, forceXin, forceXout, forceYin,
  forceYout, forceZin, forceZout, solIALLA, solIALLB, solIALLC,
  solIALLD, solIALLE, solIALLF, solIALLG];

k = 1; (*Average degree -- It is written here as 1, but this notebook
was also run with k values of 1.5 and 2*)
Lambda = 0.003; (*Birth/Immigration rate*)
mu = 0.003; (*Death/Emigration rate*)
Gamma = 0.01; (*Removal rate*)
Tau = 0.01 ; (*Transmissibility*)
nX = 323000; (* # of nodes in population X --> BLACK*)
nY = 64600; (* # of nodes in population Y --> LATINO*)
nZ = 258400; (* # of nodes in population Z --> WHITE*)
Ntot = nX + nY + nZ; (* total population size of the network *)

M = {{(Rho + (1-Rho)*nX/Ntot), (1-Rho)*nY/Ntot, (1-Rho)*nZ/Ntot},
    {(1-Rho)*nX/Ntot, (Rho + (1-Rho)*nY/Ntot), (1-Rho)*nZ/Ntot},
    {(1-Rho)*nX/Ntot, (1-Rho)*nY/Ntot, (Rho + (1-Rho)*nZ/Ntot)}};
(*Mixing matrix, Mij tells us the probability that a neighbor of
mine is from community j given that I am in community i*)
```

```
P = {{k*M[[1, 1]]/nX, k*M[[1, 2]]/nY, k*M[[1, 3]]/nZ},
    {k*M[[2, 1]]/nX, k*M[[2, 2]]/nY, k*M[[2, 3]]/nZ},
    {k*M[[3, 1]]/nX, k*M[[3, 2]]/nY, k*M[[3, 3]]/nZ}};
(*Mixing matrix, Pij tells us the probability that a node in the
ith communitiy is adjacent to a node in the jth community*)

tstart = 0; (*Time will start at 0 months*)
tfinish = 1200; (*Count time from 0 to 1200 months - 100 years*)
SXstart = 309110; (*Initial BLACK susceptible population - from 2012 data*)
SYstart = 63440; (*Initial LATINO susceptible population - from 2012 data*)
SZstart = 255300; (*Initial WHITE susceptible population - from 2012 data*)
IXstart = 13890; (*Initial BLACK HIV positive population - from 2012 data*)
IYstart = 1160; (*Initial LATINO HIV positive population - from 2012 data*)
IZstart = 3100; (*Initial WHITE HIV positive population - from 2012 data*)

BX = nX (1 - ((1 - P[[2, 1]])^nY*(1 - P[[3, 1]])^nZ));
    (*Expected value for # boundary nodes in BLACK population*)
BY = nY (1 - ((1 - P[[1, 2]])^nX*(1 - P[[3, 2]])^nZ));
    (*Expected value for # boundary nodes in LATINO population*)
BZ = nZ (1 - ((1 - P[[1, 3]])^nX*(1 - P[[2, 3]])^nY));
    (*Expected value for # boundary nodes in WHITE population*)
forceXin[ix_] := Tau *P[[1, 1]]*ix;
    (*Internal force of infection for BLACK community*)
forceYin[iy_] := Tau *P[[2, 2]]*iy;
    (*Internal force of infection for LATINO community*)
forceZin[iz_] := Tau *P[[3, 3]]*iz;
    (*Internal force of infection for WHITE community*)
forceXout[iy_, iz_] := Tau*(P[[1, 2]]*iy + P[[1, 3]]*iz);
    (*External force of infection for BLACK community *)
forceYout[ix_, iz_] := Tau*(P[[2, 1]]*ix + P[[2, 3]]*iz);
    (*External force of infection for LATINO community *)
forceZout[ix_, iy_] := Tau*(P[[3, 1]]*ix + P[[3, 2]]*iy);
    (*External force of infection for WHITE community *)

Rho = .0;
(* Portion of ones neighbors reserved for inner-community
connections. This mixing parameter is changed repeatedly and plots
are generated for each value. Later the plots are overlaid on top
of one another. *)

(* This section solves our system of SIR differential equations and
generates plots for the number of infectives over time for community
X (BLACK), community Y (LATINO), and community Z (WHITE), as well as
the overall network. The following blocks do the same evaluation,
only for different values of Rho. Note that we color and pattern
the curves such that curves representing the BLACK community are
normal lines, curves representing the LATINO community are dashed,
and curves representing the WHITE community are thick. Additionally,
```

```
the curves progress in chromatic order from red to black as we
increase Rho.*)

solution = NDSolve[{
    SX'[t] == -(SX[t]*forceXin[InfX[t]] + BX/nX *SX[t]*
            (forceXout[InfY[t], InfZ[t]] - forceXin[InfX[t]]
            *forceXout[InfY[t], InfZ[t]])) + Lambda*nX - mu*SX[t],
    SY'[t] == -(SY[t]*forceYin[InfY[t]] + BY/nY *SY[t]*
            (forceYout[InfX[t], InfZ[t]] - forceYin[InfY[t]]
            *forceYout[InfX[t], InfZ[t]])) + Lambda*nY - mu*SY[t],
    SZ'[t] == -(SZ[t]*forceZin[InfZ[t]] + BZ/nZ *SZ[t]*
            (forceZout[InfX[t], InfY[t]] - forceZin[InfZ[t]]
            *forceZout[InfX[t], InfY[t]])) + Lambda*nZ - mu*SZ[t],
    InfX'[t] == (SX[t]*forceXin[InfX[t]] + BX/nX *SX[t]*
            (forceXout[InfY[t], InfZ[t]] - forceXin[InfX[t]]
            *forceXout[InfY[t], InfZ[t]])) - Gamma*InfX[t] - mu*InfX[t],
    InfY'[t] == (SY[t]*forceYin[InfY[t]] + BY/nY *SY[t]*
            (forceYout[InfX[t], InfZ[t]] - forceYin[InfY[t]]
            *forceYout[InfX[t], InfZ[t]])) - Gamma*InfY[t] - mu*InfY[t],
    InfZ'[t] == (SZ[t]*forceZin[InfZ[t]] + BZ/nZ *SZ[t]*
            (forceZout[InfX[t], InfY[t]] - forceZin[InfZ[t]]
            *forceZout[InfX[t], InfY[t]])) - Gamma*InfZ[t] - mu*InfZ[t],
    SX[0] == SXstart,
    SY[0] == SYstart,
    SZ[0] == SZstart,
    InfX[0] == IXstart,
    InfY[0] == IYstart,
    InfZ[0] == IZstart},
   {SX, SY, SZ, InfX, InfY, InfZ}, {t, tstart, tfinish},
   MaxSteps -> \infty];
solIXA = Plot[Evaluate[{InfX[t]} /. solution], {t, tstart, tfinish},
   PlotRange -> {0, 15000},
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14], PlotStyle -> Red];
solIYA = Plot[Evaluate[{InfY[t]} /. solution], {t, tstart, tfinish},
   PlotRange -> {0, 15000},
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14],
   PlotStyle -> {Red, Dashed}];
solIZA = Plot[Evaluate[{InfZ[t]} /. solution], {t, tstart, tfinish},
   PlotRange -> {0, 15000},
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14], PlotStyle -> {Red, Thick}];
solIALLA =
  Plot[Evaluate[{InfX[t] + InfY[t] + InfZ[t]} /. solution], {t,
    tstart, tfinish}, PlotRange -> {0, 20000},
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14], PlotStyle -> Red];
```

```
Rho = .1;
solution = NDSolve[{
    SX'[t] == -(SX[t]*forceXin[InfX[t]] + BX/nX *SX[t]*
            (forceXout[InfY[t], InfZ[t]] - forceXin[InfX[t]]
            *forceXout[InfY[t], InfZ[t]])) + Lambda*nX - mu*SX[t],
    SY'[t] == -(SY[t]*forceYin[InfY[t]] + BY/nY *SY[t]*
            (forceYout[InfX[t], InfZ[t]] - forceYin[InfY[t]]
            *forceYout[InfX[t], InfZ[t]])) + Lambda*nY - mu*SY[t],
    SZ'[t] == -(SZ[t]*forceZin[InfZ[t]] + BZ/nZ *SZ[t]*
            (forceZout[InfX[t], InfY[t]] - forceZin[InfZ[t]]
            *forceZout[InfX[t], InfY[t]])) + Lambda*nZ - mu*SZ[t],
    InfX'[t] == (SX[t]*forceXin[InfX[t]] + BX/nX *SX[t]*
            (forceXout[InfY[t], InfZ[t]] - forceXin[InfX[t]]
            *forceXout[InfY[t], InfZ[t]])) - Gamma*InfX[t] - mu*InfX[t],
    InfY'[t] == (SY[t]*forceYin[InfY[t]] + BY/nY *SY[t]*
            (forceYout[InfX[t], InfZ[t]] - forceYin[InfY[t]]
            *forceYout[InfX[t], InfZ[t]])) - Gamma*InfY[t] - mu*InfY[t],
    InfZ'[t] == (SZ[t]*forceZin[InfZ[t]] + BZ/nZ *SZ[t]*
            (forceZout[InfX[t], InfY[t]] - forceZin[InfZ[t]]
            *forceZout[InfX[t], InfY[t]])) - Gamma*InfZ[t] - mu*InfZ[t],
    SX[0] == SXstart,
    SY[0] == SYstart,
    SZ[0] == SZstart,
    InfX[0] == IXstart,
    InfY[0] == IYstart,
    InfZ[0] == IZstart},
   {SX, SY, SZ, InfX, InfY, InfZ}, {t, tstart, tfinish},
   MaxSteps -> \infty];
solIXB = Plot[Evaluate[{InfX[t]} /. solution], {t, tstart, tfinish},
   PlotRange -> {0, 15000},
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14], PlotStyle -> Orange];
solIYB = Plot[Evaluate[{InfY[t]} /. solution], {t, tstart, tfinish},
   PlotRange -> {0, 15000},
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14],
   PlotStyle -> {Orange, Dashed}];
solIZB = Plot[Evaluate[{InfZ[t]} /. solution], {t, tstart, tfinish},
   PlotRange -> {0, 15000},
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14],
   PlotStyle -> {Orange, Thick}];
solIALLB =
  Plot[Evaluate[{InfX[t] + InfY[t] + InfZ[t]} /. solution], {t,
    tstart, tfinish}, PlotRange -> {0, 20000},
```

```
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14], PlotStyle -> Orange];

Rho = .3;
solution = NDSolve[{
    SX'[t] == -(SX[t]*forceXin[InfX[t]] + BX/nX *SX[t]*
            (forceXout[InfY[t], InfZ[t]] - forceXin[InfX[t]]
            *forceXout[InfY[t], InfZ[t]])) + Lambda*nX - mu*SX[t],
    SY'[t] == -(SY[t]*forceYin[InfY[t]] + BY/nY *SY[t]*
            (forceYout[InfX[t], InfZ[t]] - forceYin[InfY[t]]
            *forceYout[InfX[t], InfZ[t]])) + Lambda*nY - mu*SY[t],
    SZ'[t] == -(SZ[t]*forceZin[InfZ[t]] + BZ/nZ *SZ[t]*
            (forceZout[InfX[t], InfY[t]] - forceZin[InfZ[t]]
            *forceZout[InfX[t], InfY[t]])) + Lambda*nZ - mu*SZ[t],
    InfX'[t] == (SX[t]*forceXin[InfX[t]] + BX/nX *SX[t]*
            (forceXout[InfY[t], InfZ[t]] - forceXin[InfX[t]]
            *forceXout[InfY[t], InfZ[t]])) - Gamma*InfX[t] - mu*InfX[t],
    InfY'[t] == (SY[t]*forceYin[InfY[t]] + BY/nY *SY[t]*
            (forceYout[InfX[t], InfZ[t]] - forceYin[InfY[t]]
            *forceYout[InfX[t], InfZ[t]])) - Gamma*InfY[t] - mu*InfY[t],
    InfZ'[t] == (SZ[t]*forceZin[InfZ[t]] + BZ/nZ *SZ[t]*
            (forceZout[InfX[t], InfY[t]] - forceZin[InfZ[t]]
            *forceZout[InfX[t], InfY[t]])) - Gamma*InfZ[t] - mu*InfZ[t],
    SX[0] == SXstart,
    SY[0] == SYstart,
    SZ[0] == SZstart,
    InfX[0] == IXstart,
    InfY[0] == IYstart,
    InfZ[0] == IZstart},
   {SX, SY, SZ, InfX, InfY, InfZ}, {t, tstart, tfinish},
   MaxSteps -> \infty];
solIXC = Plot[Evaluate[{InfX[t]} /. solution], {t, tstart, tfinish},
   PlotRange -> {0, 15000},
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14], PlotStyle -> Yellow];
solIYC = Plot[Evaluate[{InfY[t]} /. solution], {t, tstart, tfinish},
   PlotRange -> {0, 15000},
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14],
   PlotStyle -> {Yellow, Dashed}];
solIZC = Plot[Evaluate[{InfZ[t]} /. solution], {t, tstart, tfinish},
   PlotRange -> {0, 15000},
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14],
   PlotStyle -> {Yellow, Thick}];
solIALLC =
  Plot[Evaluate[{InfX[t] + InfY[t] + InfZ[t]} /. solution], {t,
    tstart, tfinish}, PlotRange -> {0, 20000},
```

```
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14], PlotStyle -> Yellow];


Rho = .5;
solution = NDSolve[{
    SX'[t] == -(SX[t]*forceXin[InfX[t]] + BX/nX *SX[t]*
            (forceXout[InfY[t], InfZ[t]] - forceXin[InfX[t]]
            *forceXout[InfY[t], InfZ[t]])) + Lambda*nX - mu*SX[t],
    SY'[t] == -(SY[t]*forceYin[InfY[t]] + BY/nY *SY[t]*
            (forceYout[InfX[t], InfZ[t]] - forceYin[InfY[t]]
            *forceYout[InfX[t], InfZ[t]])) + Lambda*nY - mu*SY[t],
    SZ'[t] == -(SZ[t]*forceZin[InfZ[t]] + BZ/nZ *SZ[t]*
            (forceZout[InfX[t], InfY[t]] - forceZin[InfZ[t]]
            *forceZout[InfX[t], InfY[t]])) + Lambda*nZ - mu*SZ[t],
    InfX'[t] == (SX[t]*forceXin[InfX[t]] + BX/nX *SX[t]*
            (forceXout[InfY[t], InfZ[t]] - forceXin[InfX[t]]
            *forceXout[InfY[t], InfZ[t]])) - Gamma*InfX[t] - mu*InfX[t],
    InfY'[t] == (SY[t]*forceYin[InfY[t]] + BY/nY *SY[t]*
            (forceYout[InfX[t], InfZ[t]] - forceYin[InfY[t]]
            *forceYout[InfX[t], InfZ[t]])) - Gamma*InfY[t] - mu*InfY[t],
    InfZ'[t] == (SZ[t]*forceZin[InfZ[t]] + BZ/nZ *SZ[t]*
            (forceZout[InfX[t], InfY[t]] - forceZin[InfZ[t]]
            *forceZout[InfX[t], InfY[t]])) - Gamma*InfZ[t] - mu*InfZ[t],
    SX[0] == SXstart,
    SY[0] == SYstart,
    SZ[0] == SZstart,
    InfX[0] == IXstart,
    InfY[0] == IYstart,
    InfZ[0] == IZstart},
   {SX, SY, SZ, InfX, InfY, InfZ}, {t, tstart, tfinish},
   MaxSteps -> \infty];
solIXD = Plot[Evaluate[{InfX[t]} /. solution], {t, tstart, tfinish},
   PlotRange -> {0, 15000},
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14], PlotStyle -> Green];
solIYD = Plot[Evaluate[{InfY[t]} /. solution], {t, tstart, tfinish},
   PlotRange -> {0, 15000},
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14],
   PlotStyle -> {Green, Dashed}];
solIZD = Plot[Evaluate[{InfZ[t]} /. solution], {t, tstart, tfinish},
   PlotRange -> {0, 15000},
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14],
   PlotStyle -> {Green, Thick}];
solIALLD =
  Plot[Evaluate[{InfX[t] + InfY[t] + InfZ[t]} /. solution], {t,
```

```
    tstart, tfinish}, PlotRange -> {0, 20000},
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14], PlotStyle -> Green];

Rho = .7;
solution = NDSolve[{
    SX'[t] == -(SX[t]*forceXin[InfX[t]] + BX/nX *SX[t]*
            (forceXout[InfY[t], InfZ[t]] - forceXin[InfX[t]]
            *forceXout[InfY[t], InfZ[t]])) + Lambda*nX - mu*SX[t],
    SY'[t] == -(SY[t]*forceYin[InfY[t]] + BY/nY *SY[t]*
            (forceYout[InfX[t], InfZ[t]] - forceYin[InfY[t]]
            *forceYout[InfX[t], InfZ[t]])) + Lambda*nY - mu*SY[t],
    SZ'[t] == -(SZ[t]*forceZin[InfZ[t]] + BZ/nZ *SZ[t]*
            (forceZout[InfX[t], InfY[t]] - forceZin[InfZ[t]]
            *forceZout[InfX[t], InfY[t]])) + Lambda*nZ - mu*SZ[t],
    InfX'[t] == (SX[t]*forceXin[InfX[t]] + BX/nX *SX[t]*
            (forceXout[InfY[t], InfZ[t]] - forceXin[InfX[t]]
            *forceXout[InfY[t], InfZ[t]])) - Gamma*InfX[t] - mu*InfX[t],
    InfY'[t] == (SY[t]*forceYin[InfY[t]] + BY/nY *SY[t]*
            (forceYout[InfX[t], InfZ[t]] - forceYin[InfY[t]]
            *forceYout[InfX[t], InfZ[t]])) - Gamma*InfY[t] - mu*InfY[t],
    InfZ'[t] == (SZ[t]*forceZin[InfZ[t]] + BZ/nZ *SZ[t]*
            (forceZout[InfX[t], InfY[t]] - forceZin[InfZ[t]]
            *forceZout[InfX[t], InfY[t]])) - Gamma*InfZ[t] - mu*InfZ[t],
    SX[0] == SXstart,
    SY[0] == SYstart,
    SZ[0] == SZstart,
    InfX[0] == IXstart,
    InfY[0] == IYstart,
    InfZ[0] == IZstart},
   {SX, SY, SZ, InfX, InfY, InfZ}, {t, tstart, tfinish},
   MaxSteps -> \infty];
solIXE = Plot[Evaluate[{InfX[t]} /. solution], {t, tstart, tfinish},
   PlotRange -> {0, 15000},
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14], PlotStyle -> Blue];
solIYE = Plot[Evaluate[{InfY[t]} /. solution], {t, tstart, tfinish},
   PlotRange -> {0, 15000},
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14],
   PlotStyle -> {Blue, Dashed}];
solIZE = Plot[Evaluate[{InfZ[t]} /. solution], {t, tstart, tfinish},
   PlotRange -> {0, 15000},
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14],
   PlotStyle -> {Blue, Thick}];
solIALLE =
  Plot[Evaluate[{InfX[t] + InfY[t] + InfZ[t]} /. solution], {t,
```

```
    tstart, tfinish}, PlotRange -> {0, 20000},
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14], PlotStyle -> Blue];

Rho = .9;
solution = NDSolve[{
    SX'[t] == -(SX[t]*forceXin[InfX[t]] + BX/nX *SX[t]*
            (forceXout[InfY[t], InfZ[t]] - forceXin[InfX[t]]
            *forceXout[InfY[t], InfZ[t]])) + Lambda*nX - mu*SX[t],
    SY'[t] == -(SY[t]*forceYin[InfY[t]] + BY/nY *SY[t]*
            (forceYout[InfX[t], InfZ[t]] - forceYin[InfY[t]]
            *forceYout[InfX[t], InfZ[t]])) + Lambda*nY - mu*SY[t],
    SZ'[t] == -(SZ[t]*forceZin[InfZ[t]] + BZ/nZ *SZ[t]*
            (forceZout[InfX[t], InfY[t]] - forceZin[InfZ[t]]
            *forceZout[InfX[t], InfY[t]])) + Lambda*nZ - mu*SZ[t],
    InfX'[t] == (SX[t]*forceXin[InfX[t]] + BX/nX *SX[t]*
            (forceXout[InfY[t], InfZ[t]] - forceXin[InfX[t]]
            *forceXout[InfY[t], InfZ[t]])) - Gamma*InfX[t] - mu*InfX[t],
    InfY'[t] == (SY[t]*forceYin[InfY[t]] + BY/nY *SY[t]*
            (forceYout[InfX[t], InfZ[t]] - forceYin[InfY[t]]
            *forceYout[InfX[t], InfZ[t]])) - Gamma*InfY[t] - mu*InfY[t],
    InfZ'[t] == (SZ[t]*forceZin[InfZ[t]] + BZ/nZ *SZ[t]*
            (forceZout[InfX[t], InfY[t]] - forceZin[InfZ[t]]
            *forceZout[InfX[t], InfY[t]])) - Gamma*InfZ[t] - mu*InfZ[t],
    SX[0] == SXstart,
    SY[0] == SYstart,
    SZ[0] == SZstart,
    InfX[0] == IXstart,
    InfY[0] == IYstart,
    InfZ[0] == IZstart},
   {SX, SY, SZ, InfX, InfY, InfZ}, {t, tstart, tfinish},
   MaxSteps -> \infty];
solIXF = Plot[Evaluate[{InfX[t]} /. solution], {t, tstart, tfinish},
   PlotRange -> {0, 15000},
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14], PlotStyle -> Purple];
solIYF = Plot[Evaluate[{InfY[t]} /. solution], {t, tstart, tfinish},
   PlotRange -> {0, 15000},
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14],
   PlotStyle -> {Purple, Dashed}];
solIZF = Plot[Evaluate[{InfZ[t]} /. solution], {t, tstart, tfinish},
   PlotRange -> {0, 15000},
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14],
   PlotStyle -> {Purple, Thick}];
solIALLF =
  Plot[Evaluate[{InfX[t] + InfY[t] + InfZ[t]} /. solution], {t,
```

```
    tstart, tfinish}, PlotRange -> {0, 20000},
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14], PlotStyle -> Purple];


Rho = .99;
solution = NDSolve[{
    SX'[t] == -(SX[t]*forceXin[InfX[t]] + BX/nX *SX[t]*
            (forceXout[InfY[t], InfZ[t]] - forceXin[InfX[t]]
            *forceXout[InfY[t], InfZ[t]])) + Lambda*nX - mu*SX[t],
    SY'[t] == -(SY[t]*forceYin[InfY[t]] + BY/nY *SY[t]*
            (forceYout[InfX[t], InfZ[t]] - forceYin[InfY[t]]
            *forceYout[InfX[t], InfZ[t]])) + Lambda*nY - mu*SY[t],
    SZ'[t] == -(SZ[t]*forceZin[InfZ[t]] + BZ/nZ *SZ[t]*
            (forceZout[InfX[t], InfY[t]] - forceZin[InfZ[t]]
            *forceZout[InfX[t], InfY[t]])) + Lambda*nZ - mu*SZ[t],
    InfX'[t] == (SX[t]*forceXin[InfX[t]] + BX/nX *SX[t]*
            (forceXout[InfY[t], InfZ[t]] - forceXin[InfX[t]]
            *forceXout[InfY[t], InfZ[t]])) - Gamma*InfX[t] - mu*InfX[t],
    InfY'[t] == (SY[t]*forceYin[InfY[t]] + BY/nY *SY[t]*
            (forceYout[InfX[t], InfZ[t]] - forceYin[InfY[t]]
            *forceYout[InfX[t], InfZ[t]])) - Gamma*InfY[t] - mu*InfY[t],
    InfZ'[t] == (SZ[t]*forceZin[InfZ[t]] + BZ/nZ *SZ[t]*
            (forceZout[InfX[t], InfY[t]] - forceZin[InfZ[t]]
            *forceZout[InfX[t], InfY[t]])) - Gamma*InfZ[t] - mu*InfZ[t],
    SX[0] == SXstart,
    SY[0] == SYstart,
    SZ[0] == SZstart,
    InfX[0] == IXstart,
    InfY[0] == IYstart,
    InfZ[0] == IZstart},
   {SX, SY, SZ, InfX, InfY, InfZ}, {t, tstart, tfinish},
   MaxSteps -> \infty];
solIXG = Plot[Evaluate[{InfX[t]} /. solution], {t, tstart, tfinish},
   PlotRange -> {0, 15000},
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14], PlotStyle -> Black];
solIYG = Plot[Evaluate[{InfY[t]} /. solution], {t, tstart, tfinish},
   PlotRange -> {0, 15000},
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14],
   PlotStyle -> {Black, Dashed}];
solIZG = Plot[Evaluate[{InfZ[t]} /. solution], {t, tstart, tfinish},
   PlotRange -> {0, 15000},
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14],
   PlotStyle -> {Black, Thick}];
solIALLG =
```

```
  Plot[Evaluate[{InfX[t] + InfY[t] + InfZ[t]} /. solution], {t,
    tstart, tfinish}, PlotRange -> {0, 20000},
   AxesLabel -> {"Time (months)", "Number of Infectives"},
   LabelStyle -> Directive[FontSize -> 14], PlotStyle -> Black];

Show[solIXA , solIYA, solIZA, solIXB , solIYB, solIZB, solIXC ,
solIYC, solIZC, solIXD , solIYD, solIZD, solIXE , solIYE, solIZE,
solIXF , solIYF, solIZF, solIXG , solIYG, solIZG]
(*Shows the plots for the number of infectives across time for
each of our communities and each value of Rho. *)

Show[solIALLA, solIALLB, solIALLC, solIALLD, solIALLE, solIALLF,
solIALLG]
(*Shows the plots for the number of infectives across time for the
total population and each value of Rho. *)
```

Appendix B. *Mathematica*® notebook: internal degree and boundary nodes

```
(* This notebook plots the average internal degree and the expected
number of boundary nodes in a sample network of three communities of
100 nodes as a function of the mixing parameter, Rho. *)

Clear[k, nX, nY, nZ, Ntot, M, P, kinX, BX];
k = 1.5; (* average degree *)
nX = 100; (* # of nodes in population X*)
nY = 100; (* # of nodes in population Y*)
nZ = 100; (* # of nodes in population Z*)
Ntot = 300; (* total network size *)

M = {{(Rho + (1-Rho)*nX/Ntot), (1-Rho)*nY/Ntot, (1-Rho)*nZ/Ntot},
    {(1 - Rho)*nX/Ntot, (Rho + (1 - Rho)*nY/Ntot), (1-Rho)*nZ/Ntot},
    {(1-Rho)*nX/Ntot, (1-Rho)*nY/Ntot, (Rho + (1-Rho)* nZ/Ntot)}};
(*Mixing matrix, Mij tells us the probability that a neighbor of mine
is from community j given that I am in community i*)


P = {{k*M[[1, 1]]/nX, k*M[[1, 2]]/nY, k*M[[1, 3]]/nZ},
    {k*M[[2, 1]]/nX, k*M[[2, 2]]/nY, k*M[[2, 3]]/nZ},
    {k*M[[3, 1]]/nX, k*M[[3, 2]]/nY, k*M[[3, 3]]/nZ}};
(*Mixing matrix, Pij tells us the probability that a node in the ith
communitiy is adjacent to a node in the jth community*)

kinX = P[[1, 1]]*nX; (*Average # inner community links in X*)
BX = nX (1 - ((1 - P[[2, 1]])^nY*(1 - P[[3, 1]])^nZ));\\
(*Expected value for # boundary nodes in X*)

Plot[kinX, {Rho, 0, 1}, PlotRange -> {0, 1.5},
AxesLabel -> {"Rho", "kin"},
LabelStyle -> Directive[FontSize ->14]]
(*Plots the internal degree in our communities as a function of Rho*)

Plot[BX, {Rho, 0, 1}, AxesLabel -> {"Rho", "B"},
LabelStyle -> Directive[FontSize -> 14]]
(*Plots the expected number of boundary nodes in our communities
as a function of Rho*)
```

Bibliography

[1] Adimora AA, Schoenbach VJ, Bonas DM, Martinson FEA, Donaldson KH, Stancil TR (2002) Concurrent sexual partnerships among women in the United States. Epidemiology 13.3: 320–327.

[2] Adimora AA, Schoenbach VJ (2005) Social context, sexual networks, and racial disparities in rates of sexually transmitted infections. J Infect Dis 191 Supplement 1: S115–S122.

[3] Apolloni A, Poletto C, Ramasco JJ, Jensen P, Colizza V (2014) Metapopulation epidemic models with heterogeneous mixing and travel behaviour. Theor Biol Med Model 11.3. doi:10.1186/1742-4682-11-3

[4] Bogart LM, Thorburn S (2005) Are HIV/AIDS conspiracy beliefs a barrier to HIV prevention among African Americans?. J Acq Immun Ded Synd 38.2:213–218.

[5] Boily M, Bastos FI, Desai K, Masse B (2004) Changes in the transmission dynamics of the HIV epidemic after the wide-scale use of antiretroviral therapy could explain increases in sexually transmitted infections. Sex Transm Dis 31.2:100–113. doi: 10.1097/01.OLQ.0000112721.21285.A2

[6] Brauer F (2008) Mathematical Epidemiology. Berlin: Springer.

[7] Catania JA, Coates TJ, Kegels S, Fullilove MT (1992) Condom use in multi-ethnic neighborhoods of San Francisco: the population-based AMEN (AIDS in Multi-Ethnic Neighborhoods) Study. Am J Public Health 82:284287. doi: 10.2105/AJPH.82.2.284

[8] Chen NE, Gallant JE, Page KR (2012) A systematic review of HIV/AIDS survival and delayed diagnosis among Hispanics in the United States. J Immigr Minor Health 14.1:65–81. doi: 10.1007/s10903-011-9497-y

[9] Clémençon S, De Arazoza H, Rossi F, Tran VC (2014) A statistical network analysis of the HIV/AIDS epidemics in Cuba. arXiv preprint arXiv:1401.6449.

[10] Coelho FC, Cruz OG, Codeço CT (2008) Epigrass: a tool to study disease spread in complex networks. Source code for biology and medicine 3.3. doi:10.1186/1751-0473-3-3

[11] David Hill (2011) Census: more people moving to DC. The Washington Times.

[12] District of Columbia Department of Health, HIV/AIDS, Hepatitis, STD, and TB Administration (HAHSTA) (2012) Annual epidemiology and surveillance report.

[13] District of Columbia Department of Health (2013) MSM in DC: A life long commitment to stay HIV free.

[14] Earnshaw VA, Bogart LM, Dovidio JF, Williams DR (2013) Stigma and racial/ethnic HIV disparities: moving toward resilience. Am Psychol 68.4:225–236. doi: 10.1037/a0032705

[15] The Henry J. Kaiser Family Foundation (2012) The HIV/AIDS epidemic in Washington, D.C..

[16] Huang L, Park K, Lai YC (2006) Information propagation on modular networks. Phys Rev E 73.3:035103. doi: 10.1103/PhysRevE.73.035103

[17] Hyman JM, Stanley EA (1988) Using mathematical models to understand the AIDS epidemic. Math Biosci 90:415–473.

[18] Jacquez JA, Simon CP, Koopman J, Sattenspiel L, Perry T (1988) Modelling and analysing HIV transmission: the effect of contact patterns. Math Biosci 92:119–199.

[19] Johnson MO, Catz SL, Remien RH, Rotheram-Borus MJ, Morin SF, Charlebois E, Gore-Felton C et al. (2003) Theory-guided, empirically supported avenues for intervention on HIV medication nonadherence: findings from the Healthy Living Project. AIDS Patient Care ST 17.12:645–656. doi:10.1089/108729103771928708

[20] Kenah E, Robins JM (2007). Network-based analysis of stochastic SIR epidemic models with random and proportionate mixing. J Theor Biol 249:706–722. doi:10.1016/j.jtbi.2007.09.011

[21] Kermack WO, McKendrick AG (1927) A contribution to the mathematical theory of epidemics. Proc R Soc Lond 115:700–721.

[22] Kerstetter K, Reed J, Lazere E (2009) New census data reveal growing income gaps in the district. DC Fiscal Policy Institute.

[23] Kiskowski MA (2014) A three-scale network model for the early growth dynamics of 2014 West Africa Ebola epidemic. PLoS Current Outbreaks. doi: 10.1371/currents.outbreaks.c6efe8274dc55274f05cbcb62bbe6070

[24] Kitchovitch S, Lió P (2011) Community structure in social networks: applications for epidemiological modelling. PLoS ONE 6.7:e22220. doi: 10.1371/journal.pone.0022220

[25] Laumann EO, Gagnon JH, Michael RT, Michaels S (1994) The social organization of sexuality. Chicago: University of Chicago Press.

[26] Laumann EO, Youm Y (1999) Racial/ethnic group differences in the prevalence of sexually transmitted diseases in the United States: a network explanation. Sex Transm Dis 26.5:250–261.

[27] Leserman J (2003) HIV disease progression: depression, stress, and possible mechanisms. Biol Psychiat 54.3:295–306. doi:10.1016/S0006-3223(03)00323-8

[28] Liljeros F, Edling CR, Amaral LAN, Stanley HE, Aberg Y (2001) The web of human sexual contacts. Nature 411:907–908. doi:10.1038/35082140

[29] Miller JC (2011) A note on a paper by Erik Volz: SIR dynamics in random networks. J Math Biol 62:349–358. doi: 10.1007/s00285-010-0337-9

[30] Moody J (2001) Race, school integration, and friendship segregation in America. Am J Sociol 107:679–716. doi:10.1086/338954

[31] Morris M, Kretzschmar M (1995) Concurrent partnerships and transmission dynamics in networks. Soc Networks 17.3:299–318. doi: 10.1016/0378-8733(95)00268-S

[32] Newman MEJ (2002) Spread of epidemic disease on networks. Phys Rev 66:016128. doi: 10.1103/PhysRevE.66.016128

[33] Newman MEJ (2003) The structure and function of complex networks. SIAM Rev 45.2:167–256. doi:10.1137/S003614450342480

[34] Newman MEJ (2010) Networks: An Introduction New York: Oxford University Press.

[35] Obermeyer CM, Osborn M (2007) The utilization of testing and counseling for HIV: a review of the social and behavioral evidence. Am J Public Health 97.10:1762–1774. doi:10.2105/AJPH.2006.096263

[36] Peng XL, Small M, Xu XJ, Fu X (2013) Temporal prediction of epidemic patterns in community networks. New J Phys 15:113033. doi:10.1088/1367-2630/15/11/113033

[37] Public Health Agency of Canada, http://www.phac-aspc.gc.ca/aids-sida/publication/hivtr-rtvih-eng.php 2013.

[38] Punyacharoensin N, Edmunds WJ, De Angelis D, White RG (2011) Mathematical models for the study of HIV spread and control amongst men who have sex with men. Eur J Epidemiol 26:695–709. doi: 10.1007/s10654-011-9614-1

[39] Ross HL, Sawhill IV, MacIntosh AR (1975) Time of transition: The growth of families headed by women. The Urban Institute.

[40] Rothenberg R (2009) HIV transmission networks. Curr Opin HIV AIDS 4.4:260–265. doi: 10.1097/COH.0b013e32832c7cfc

[41] Salathe M, Jones JH (2010) Dynamics and control of diseases in networks with community structure. PLoS Comput Biol 6.4:e1000736. doi: 10.1371/journal.pcbi.1000736

[42] Sattenspiel L, Simon CP (1988) The spread and persistence of infectious disease in structured populations Math Biosci 90:341–366. doi: 10.1016/0025-5564(88)90074-0

[43] Sattenspiel L, Koopman J, Simon C, Jacquez JA (1990) The effects of population structure on the spread of the HIV infection. Am J Phys Anthropol 82:421–429. doi: 10.1002/ajpa.1330820404

[44] Shapiro MF, Morton SC, McCaffrey DF, Senterfitt JW, Fleishman JA, Perlman JF, Athey LA et al. (1999) Variations in the care of HIV-infected adults in the United States: results from the HIV Cost and Services Utilization Study. JAMA 281.24:2305–2315. doi: 10.1001/jama.281.24.2305

[45] Tatian P, Lei S (2010) Washington, D.C., Our Changing City. The Urban Institute.

[46] Turner BJ, Cunningham WE, Duan N, Andersen RM, Shapiro MF, Bozzette SA, Nakazono T et al. (2000) Delayed medical care after diagnosis in a US national probability sample of persons infected with human immunodeficiency virus. Arch Intern Med 160.17:2614–2622.

[47] United States Census Bureau (2015) District of Columbia Quick Facts.

[48] United States Department of Health and Human Services (2013) Stages of HIV infection. www.aids.gov.

[49] Volz E (2008) SIR dynamics in random networks with heterogeneous connectivity. J Math Biol 56:293–310. doi 10.1007/s00285-007-0116-4

[50] World Health Organization (2014) HIV/AIDS. http://www.who.int/features/qa/71/en/.

[51] Zenilman JM, Ellish N, Fresia A, Glass G (1999) The geography of sexual partnerships in Baltimore: applications of core theory dynamics using a geographic information system. Sex Transm Dis 26.2:75–81.