NETWORK MODELING OF SEXUALLY TRANSMITTED DISEASES

A Thesis
Presented to
The Academic Faculty

by

Yao-Hsuan Chen

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
H. Milton Stewart School of Industrial and Systems Engineering

Georgia Institute of Technology
May 2014

NETWORK MODELING OF SEXUALLY TRANSMITTED DISEASES

Approved by:

Dr. Joel Sokol, Advisor
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Paul Griffin
Harold and Inge Marcus Department of Industrial and Manufacturing Engineering
Pennsylvania State University

Dr. Craig Tovey
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Thomas L. Gift
Division of Sexually Transmitted Diseases
Centers for Disease Control and Prevention

Dr. Dave Goldsman
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: April 2014

ACKNOWLEDGEMENTS

In this unexpectedly long and lonesome journey, beautiful people showed up and kindly gave me their hands to help me reach the end. Without them, I could not have finished what I started.

Dr. Gary Parker and Mrs. Pam Morrison protected me from being stopped by many hardships. My thesis advisor, Dr. Joel Sokol, not only shared with me in this thesis work his tremendous talent of making math and science conceivable and useful in the daily life, but inspired me to follow my heart with courage (and with wisdom). Besides their time and effort to improve my thesis manuscript, I hope to recognize the importance of great conversations with Dr. Paul Griffin, a life filled with kindness and spirit-lifting humor of Dr. Dave Goldsman and Dr. Gamze Tokol-Goldsman, and a shelter provided by Dr. Craig Tovey and his family, Gail and William, where I can rest to regain strength.

In addition, I am grateful to Dr. Samuel R. Friedman, Dr. Thomas Gift, Dr. H. Hunter Handsfield, and Dr. Jami Leichliter for their guidance, who made me think like a public health professional, not just a modeler. This work was partially supported by Battelle (contract 215031) and the Centers for Disease Control and Prevention (contract W911N F007-D-0001).

In the end, I want to thank a great friend, Dr. Tsung-Lin Wu, for always helping me in the right way at the right time, and my best friend, Melih Çelik, for his input of hope in my life. Last but not least, I can never thank my mother enough for enduring my foolishness at the cost of her increasing amount of white hair.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Compartmental models have long been used to represent disease progression and propagation. Since individuals and their relationships can be viewed as a network on top of which diseases spread over time, compartmental disease models have also been combined with static contact network models to capture heterogeneity in the number as well as in the duration of individual contacts. However, such models assume the contact network remains static over time. To replicate more closely the population network structures of interest, in this thesis we create a dynamic network model. The model is an extension of the static preferential attachment network model: it allows arcs to be dropped and added over time based on the nodes' demographic and sociological attributes.

Network, Norms and HIV/STI Risk Among Youth (NNAHRAY) is a community relationship survey data set recording 465 residents' demography, relationships, and blood testing results of Sexually Transmitted Diseases (STDs) in Bushwick, NY [19]. It provides a rare sample of a human risky-behavior contact network. Combining disease compartmental models with our dynamic network model, we simulate the spread of Human Immunodeficiency Virus (HIV) and Herpes Simplex Type 2 Virus (HSV2) with consideration of HSV2's synergistic impact on HIV's transmission in Bushwick from 1990 to 2002.

When our disease spread model parameters are set according to the observed data and the epidemiological literature such that the designed network metrics approximate the data well, the model reproduces HIV prevalence, HSV-2 prevalence, and the contact network close to those observed in NNAHRAY. Our model prevalence prediction results of HIV annual prevalence are closer to the estimated values from

the literature than those of any disease spread model based on static networks. Our work supports the hypothesis that considering the underlying contact dynamics as well as network structures is important for making good disease prevalence predictions. From our network model fitting experience, we demonstrate the need to model the data sampling process when validating against real-world data.

Our model, under certain conditions, has prevalence prediction results that are insensitive to changes in network size. The analysis of various prevention/intervention strategies targeting different risky groups gives important insights into strategy prioritization and illustrates how our model can be used to assist in making public health policy decisions in practice, both for individual diseases and in the more-recent area of study that considers synergy between two diseases.

CHAPTER I

RESEARCH BACKGROUND AND LITERATURE REVIEW

The complexity of disease spread, due to both development within a host and in transmission through a population, makes analyzing the impact of public health prevention/intervention programs a difficult task. The spread of disease has been traditionally modeled using differential equations, in which the human contact network is assumed to be homogeneous (at least within pairs of groups). This unrealistic assumption has spurred efforts to develop models that account for contact heterogeneity. One of these is the network model.

When a network, which consists of nodes and arcs between nodes, is applied to model population contacts, a node is regarded as an individual and an arc as the indicator that shows the linked individuals have contacted with each other. Unlike the traditional differential equation model, which assumes every infected person in the same compartment (normally characterized by age and disease status) has the same contacts as the others, in a contact network nodes from the same compartment may have different contacts and/or different numbers of contacts, and the disease can transmit from one node to another only when the two are linked by an arc.

Various network models have been developed to explain the attributes of contact networks observed in real life. In general, they can be partitioned into static and dynamic network models. In a static network model arcs never appear or disappear, while in the dynamic network model arcs are expected to appear and disappear over time. Due to the more-realistic representation of contact characteristics and therefore a more-realistic disease propagation process, a dynamic network model can be

preferable to a static network model in public intervention policy analysis.

We hypothesize that, to closely reflect the disease propagation process, a synthesized population model needs to include the dynamic nature of contacts and match well with the contact structures observed in real data.

To test the hypothesis, in Chapter 2 we introduce a new dynamic network model combined with the compartmental disease progression model. We show in Chapter 3 how to parameterize our network model with behavioral study data collected by public health workers as well as HIV and HSV-2 transmission data in the epidemiological literature. Our model-fitting experience shows that it can be important to model how the interviewee sampling was done in detail in order to match the network model structures with data.

After completing the parameterization of our model, in Chapter 4 we compare our model's annual HIV prevalence prediction with other static network models'. The comparison results show that inclusion of network dynamics and network structural properties is indeed necessary to closely replicate the estimated HIV prevalence from the literature. Also in Chapter 4 we confirm our model's scale-invariant property computationally and then illustrate our model's potential use to analyze intervention policy effectiveness. In Chapter 5 we conclude this thesis with a summary of research contributions and discussion of future research directions.

1.1 Literature Review

The concepts of compartmentalizing population and regarding diseases as the driving forces for individuals flowing through compartments were introduced as early as 1927 [28], but did not attract much research attention until 1982. Kermack and McKendrick [28] compartmentalized the population into three groups based on disease stages: (1) susceptible, (2) infected and infectious, and (3) recovered. The individuals flow through the compartments in order. The flow speed depends on how

infectious the disease of concern is: the more infectious the virus is, the shorter time it requires for all individuals in the population to move from the susceptible group to the infected and infectious group. The flow between compartments, therefore, represents the spread of disease among population. This type of model also was later referred to as an SIR model.

The model simplicity results from two major assumptions: individuals within the same compartment are equally infectious or equally susceptible, and all individuals in the infected compartment can reach all individuals in the susceptible compartment. In 1982, Anderson and May [2, 3, 4] successfully brought the SIR model back to the attention of the field of epidemiology. To accommodate the incubation period and the age-wise infection rates of diseases, they added to the SIR model one more disease stage compartment, infected but not infectious, and divided all disease stage compartments further by age. After the concept of population compartmentalization was widely accepted, more variants of the SIR model were developed to predict endemic outbreak potential [12, 43] and to assess public health policy effectiveness [5, 11, 39]. Since the models are usually formulated with ordinary differential equations, they are also called compartmental differential equation models of disease spread.

The assumption of equal infectivity, susceptibility, and connectivity in the compartmental model can be questioned easily, especially when it is used to replicate the spread of Sexually Transmitted Diseases (STDs). Because the number of individual sexual contacts ranges more widely than the casual contacts [31], and also because the infection likelihood varies greatly between individuals at high risk and the general public, the modeling of STDs naturally needs to take heterogeneity and connectivity into consideration so that it can be accurate enough to assist in analyzing intervention policies.

The emergence of large static network data [1, 27, 44, 47] and the development of

static contact network models, such as the configuration network model [33], small-world network model [44], and geometric preferential-attachment model [15], in the recent two decades enabled researchers to effectively avoid the homogeneity assumption. Many works, e.g., [30, 34, 37] have shown how to define heterogeneity in a population mathematically and further incorporate it with the compartmental model under various contact network settings. These contact networks are all static, partially for the benefit of analytical tractability. Since in a static contact network any contact is assumed to exist from the beginning of the disease spread until the end, it may lead to disease prevalence inacurracy.

To model the dynamics of contacts, research efforts have been put into the following three directions: (1) collection of larger and more-precise dynamic contact data, (2) development of sophisticated agent-based simulation models, and (3) construction of abstract dynamic network models. For example, [40] extracted an approximate dynamic sexual contact network from the massive number of prostitute rating comments by sex buyers. [9, 14] used census data to parameterize individuals' characteristics and behaviors in agent-based simulation models of measles and influenza respectively. [41] proposed a theoretical dynamic network model in which an individual has a fixed number of contacts but with a varying set of individuals over time.

To date, there is still no large real-life dynamic contact data available. Most existing contact network data are samples of the real-life dynamic contact network of certain focus groups, for example, young adult populations [29] and drug users [45]. Discussion of disease prevalence on the collected contact network data provides ad-hoc insights, difficult to generalize. To represent the whole population, dynamic network models therefore play an important role. Agent-based simulation models are capable of modeling dynamic interactions in detail, but when it comes to parameterizing the model, the amount of data needed, both the census and the field behavioral data, is generally not available.

Abstract dynamic contact network models, on the other hand, keep the number of parameters under control. They were built upon the establishment of static network models with additional parameters to control the dynamics of networks. For example, the models in [13, 42] started with the static configuration model algorithm. They both allow arcs to drop and form over time, but use different algorithms to maintain the chosen degree distribution. For example, the model in [25] adopted an algorithm that creates the small-world model when forming and dropping arcs. All these models have great potential of being incorporating into disease modeling. However, none of the dynamic contact network models have been validated either by large dynamic network data, which is still missing for comparison, or by the available static contact network data.

Compared to the progress made in population interaction modeling, the modeling of disease interaction has been less studied. The presence of STDs which cause genital ulceration, such as HSV-2 and syphilis, has long been suspected to be an important cofactor of HIV spread [16]. [23] built an agent-based simulation model to study the synergistic impact of syphilis on HIV transmission among men who have sex with men (MSM) in Australia. Their results show that it is possible and also effective to bring down HIV incidence by successfully implementing a national syphilis action plan. The same cost-effectiveness analysis would also be desirable for HSV-2, especially given that the estimated global prevalence of HSV-2 is 536 million, including 23.6 million new cases worldwide in 2003 [32], and also that HSV-2 has been found to be highly prevalent in places where HIV incidence is high [24].

CHAPTER II

DYNAMIC PREFERENTIAL NETWORK MODEL

In this chapter, we propose a general dynamic network model whose parameters allow us to study the dynamics of personal attributes, contacts, and contact network structures. Our model's outputs are undirected networks, but it can be flexibly extended to develop directed networks for other applications where they are more appropriate.

The model is simple and general enough to fit many situations. In this thesis we show its application to STD networks. In this chapter we first introduce the basics of our model, which replicate the dynamics of contact in a network. Then we introduce the more-detailed version of our model, designed to accommodate both personal attribute changes and disease spread over time.

2.1 The Model Basics

Starting from a single node, our model produces an undirected network whose number of nodes grows and whose arcs appear and disappear with time. Our model takes three kinds of actions to alter the network structure over time: increasing the network size, adding new arcs, and deleting arcs. When a new node is to be added, our model uses the preferential attachment algorithm (see Section 2.1.1) to select an existing node to attach to the new node. To add or delete an existing arc, it uses our arc change algorithm (see Section 2.1.2). The preferential algorithm, the arc change algorithm, and a detailed description of our overall model are illustrated in order in the next sections.

2.1.1 Preferential Attachment Algorithm

The power law distribution of degree has been widely observed in large network systems in real life, such as scientific collaboration networks [35], sexual networks [31], and the internet [1]. One family of methods to produce networks with a power law distribution is based on the preferential attachment mechanism, in which new nodes are more likely to connect to existing nodes of higher degree. Every newly added node goes through the preferential attachment algorithm to choose an existing node to form an arc with. Our model uses a similar preferential attachment formulation to that proposed in [15]: Given a vector c of the degrees of each existing node, a new node attaches to existing node i with probability $c_i^\alpha / (\sum_{k \in N} c_k^\alpha)$ for some exponent α, where N is the set of all existing nodes. However, assuming that isolated nodes are more eager to establish new arcs than connected nodes, which is usually the case with human partner-seeking behavior, in our algorithm we set the isolated nodes to be as popular as the most-connected node, i.e., if $c_i = 0$, we instead set $c_i = \max_k\{c_k\}$.

Figure 2.1.1 describes how the preferential attachment is implemented in our model. Input parameter α (exponential preference) will get its value from fits to actual data (see Chapter 3).

2.1.2 Arc Add Algorithm

The arc add algorithm is designed to capture the establishment of new relationships between existing nodes. When the network is scheduled to establish new relationships, the algorithm examines each existing node, and with probability P_{Add} it connects the examined node with another node using the same node selection process as the one implemented in the preferential attachment algorithm. Figure 2.1.2 describes how the arc add algorithm is implemented in our model. Input parameter P_{Add} (probability for a node to add a new arc) will get its value from fits to actual data (see Chapter 3).

Data: Set of nodes in the current network: N
Set of arcs in the current network: A

Input: Degree of connectivity: α

Function create_weights($N, A, \alpha, preferential\_weight[\ ]$)
begin
  Store the number of arcs of each existing node in $c[\ ]$;
  Let $maxc = max_i\{c[i]\}$;
  sum=0 ;
  for $i = 1$ ***to*** $|N|$ **do**
    if $c[i] = 0$ **then**
      $c[i] = maxc$;
      $sum \mathrel{+}= c[i]^\alpha$;
    end
  end
  for $i = 1$ ***to*** $|N|$ **do**
    $preferential\_weight[i] = c[i]^\alpha / sum$;
  end
end
Function preferential_attachment(N, A, α)
begin
  Call create_weights($N, A, \alpha, preferential\_weight[\ ]$) ;
  Choose node $n \in N$ using $preferential\_weight[\ ]$ as a probability mass function;
  Add node $|N| + 1$ to N ;
  Add arc $(|N| + 1, n)$ to A ;
end

Figure 2.1.1: Preferential attachment algorithm

```
Data: Set of nodes in the current network: N
      Set of arcs in the current network: A

Input: Degree of connectivity: α
       Probability for a node to add a new arc: P_Add

Function arc_add(N, A, α, P_Add)
begin
    for i = 1 to |N| do
        Call create_weights(N, A, α, preferential_weight[ ]) ;
        With probability P_Add do
            Choose a node n ∈ N using preferential_weight[ ] as a
            probability mass function ;
            if i ≠ n then
                add arc (i, n) to A ;
            end
        end
    end
end
```

Figure 2.1.2: Arc add algorithm for existing nodes

2.1.3 Arc Drop Algorithm

The arc drop algorithm is designed to model the disappearance of existing relationships. When the network is scheduled to abandon existing relationships, the algorithm examines each existing arc and with probability P_{Drop} it drops the examined arc from the current network. Figure 2.1.3 describes how the arc drop algorithm is implemented in our model. Input parameter P_{Drop} (probability for an existing arc to disappear) will get its value from fits to actual data (see Chapter 3).

2.1.4 Dynamic Network Model

Our general dynamic model is constructed based on the preferential attachment, arc add, and arc drop algorithms. The former guides the model to produce a power law degree distribution and the latter two algorithms equip the model with dynamic arc features. There are in total seven parameters in our model: besides the parameters

Data: Set of nodes in the current network: N
Set of arcs in the current network: A

Input: Probability for an existing arc to be dropped: P_{Drop}

Function arc_drop(N, A, P_{Drop})
begin
  for $a = 1$ ***to*** $|A|$ **do**
    With probability P_{Drop} **do**
      Delete arc a from A ;
    end
  end
end

Figure 2.1.3: Arc drop algorithm

α, P_{Add}, P_{Drop} explained as part of the preferential attachment, arc add, and arc drop algorithms, there are three more parameters, New, T_{Add}, and T_{Change}, that govern the frequency of dynamic behaviors, and one more parameter $WEEK$, which is just the number of time periods that the model is run for. The complete model description is given in Figure 2.1.4.

Our model works in the following way. It starts with only one node in the network. Every T_{Add} weeks it adds to the current model New number of new nodes. To assign arcs between new nodes and existing nodes, it follows the preferential attachment algorithm. Every T_{Change} weeks it scans through the current network to add new arcs and to drop arcs among existing nodes according to the arc add and arc drop algorithms. The algorithm terminates after $WEEK$ weeks.

2.2 Model Modification and Extension

Our dynamic network model can be modified and extended to accommodate the process of disease spreading, one of the interesting applications of contact network model research. We incorporate the idea of compartmentalization to track node attributes as well as arc attributes, and to include a slightly different node selection process in both

Data: Number of weeks to run our model: $WEEK$

Input: Degree of connectivity: α
Probability for a node to add a new arc: P_{Add}
Probability for an existing arc to be dropped: P_{Drop}
Number of new nodes added at the same time: New
Time interval between adding new nodes: T_{Add}
Time interval between changing current arcs: T_{Change}

Function
dynamic_preferential_model($WEEK, NEW, T_{Add}, T_{Change}, P_{Add}, P_{Drop}, \alpha$)
begin
  $N = 1$; $A = \emptyset$;
  $model\_time = 0$;
  while $model\_time \leq WEEK$ **do**
    if $(model\_time \ mod \ T_{Add}) = 0$ **then**
      Call preferential_attachment (N, A, α) New times;
    end
    if $(model\_time \ mod \ T_{Change}) = 0$ **then**
      Call arc_add (N, A, α, P_{Add}) ;
      Call arc_drop (N, A, P_{Drop}) ;
    end
    $model\_time = model\_time + 1$;
  end
end

Figure 2.1.4: Dynamic preferential model

the preferential attachment and arc add algorithms. The modification is essential to reflect two observations in reality. The first observation is that nodes of different attributes experience disease transmission differently. For example, very young people can be infected with seasonal influenza virus more easily than people of middle age, so differentiating between those nodes is important. The second observation is that arcs of different types transmit disease with different efficiencies. For example, HIV is more likely to transmit through an arc representing contact resulting from needle sharing than an arc representing contact resulting from handshaking.

Once we include node and arc attributes to accurately capture disease spreading, we need to also incorporate them in the node selection process to prevent unwanted arc assignment results, such as the case in which an injection drug user node has a needle-sharing contact with a non-injection-drug user node. On top of our dynamic network model, we include a disease spread algorithm to capture the dynamics of disease spreading between nodes.

In this section we first explain how the node selection process is modified in the preferential attachment and arc change algorithms. Then we introduce the disease spread algorithm. Last, we show a summary of our modified and extended model.

2.2.1 Modifying the Preferential Attachment and Arc Add Algorithms

Both in the preferential attachment and arc add algorithms, the node selection process needs to take node attributes and arc attributes into consideration. As shown in Figure 2.2.1, Figure 2.2.2, and Figure 2.2.3, three more vectors CT, NT, and $pn[\,]$ are now included in set of input data. CT stores the arc contact types and NT stores the node types accepted in our model. $pn[\,]$ stores the probability of having a node of each type. As we show in Chapter 3, the frequencies of node and arc types and their connectivity with one another can be estimated from observed data.

There are two more matrices included in the algorithms: $pnc[\,][\,]$ and $pnn[\,][\,]$,

which can also be estimated directly from data. $pnc[\,][\,]$ stores the probability for nodes of each type to have arcs of each type, while $pnn[\,][\,]$ stores the probability for nodes of each type to have an arc connecting to nodes of each type. $pnc[\,][\,]$ and $pnn[\,][\,]$ have entries equal to zero if the corresponding pair of node and arc or pair of nodes is unwanted or unrealizable.

After $pnc[\,][\,]$ and $pnn[\,][\,]$ are properly set, they are used to alter the node selection process in both algorithms. Before the modification is in place, we use the current node degree, c_i, as the only criterion to determine the popularity of a node i, and the probability that node i will form an arc with node j is equal to $c[j]^\alpha / \sum_{k \in N} c[k]^\alpha$. The new probability is now modified to be equal to $(c[j]^\alpha / \sum_{k \in N} c[k]^\alpha) \times pnn[j\text{'s node type}][i\text{'s node type}] \times pnc[j\text{'s node type}][a]$, where a is the type of arc which node i chooses in advance using $pnc[i\text{'s node type}][\,]$ as a probability mass function. In this way we still maintain the preferential attachment mechanism based on node degree while ruling out unwanted arc formation between any incompatible pair of nodes. More generally, we account for node preference in terms of contacted node types.

We can view the modified node selection process as first choosing the type of arc and then preferentially selecting the corresponding node. If no node in the current network is compatible with the chosen arc type in the first attempt, we set the arc type equal to 1 and repeat the node selection process once more.

2.2.2 Disease Spread Algorithm

In the contact network setting, a disease is transmitted through arcs, with varying transmission rates whose values can change with node types and arc types (including gender, health conditions, etc., of the connected pair of nodes). We create the disease spread algorithm, detailed in Figure 2.2.4, to reflect this aspect of disease transmission.

In the disease spread algorithm the vector $S[\,]$ stores the stages of disease, the

Data: Set of nodes in the current network: N
Set of arcs in the current network: A
Set of contact types in our model: CT
Set of node types in our model: NT
Probability of having a type A node: $pn[A]$
Probability of a type A node having a type C contact: $pnc[A][C]$
Probability of a type A node choosing a type B node: $pnn[A][B]$

Input: Degree of connectivity: α

Function create_weights($N, A, CT, NT, pnc[\,][\,], pnn[\,][\,], node\_type, preferential\_weight[\,]$)
begin
  Store the number of contacts of each existing node in $c[\,]$;
  Let $maxc = max_i\{c[i]\}$;
  sum=0 ;
  for $i = 1$ ***to*** $|N|$ **do**
    if $c[i] = 0$ **then**
      $c[i] = maxc$;
      $sum \mathrel{+}= c[i]^{\alpha}$;
    end
  end
  Store the node type of each existing node in $nt[\,]$;
  Choose an $arc\_type \in CT$ using $pnc[nt][\,]$ as a probability mass function ;
  for $j = 1$ ***to*** $|N|$ **do**
    $preferential\_weight[j] =$
    $(c[j]^{\alpha}/sum) \times pnn[nt[i]][nt[j]] \times pnc[nt[j]][arc\_type]$;
  end
  if *preferential_weight[ ] are all* $0s$ **then**
    $arc\_type = 1$;
    for $j = 1$ ***to*** N **do**
      $preferential\_weight[k] =$
      $(c[j]^{\alpha}/sum) \times pnn[nt[i]][nt[j]] \times pnc[nt[j]][arc\_type]$;
    end
  end
end

Figure 2.2.1: Preferential attachment algorithm - Part I

```
Function preferential_attachment(N, A, CT, NT, pnc[ ][ ], pnn[ ][ ])
begin
    Choose node_type ∈ NT using pn[ ] as a probability mass function;
    Call create_weights(N, A, CT, NT, pnc[ ][ ], pnn[ ][ ], node_type,
    preferential_weight[ ]) ;
    Choose node n ∈ N using preferential_weight[ ] as a probability mass
    function ;
    Add node |N| + 1 to N ;
    Add type arc_type arc (|N| + 1, n) to A ;
end
```

Figure 2.2.2: Preferential attachment algorithm - Part II

```
Data: Set of nodes in the current network: N
      Set of arcs in the current network: A
      Set of contact types in our model: CT
      Set of node types in our model: NT
      Probability of a type A node having a type B contact: pnc[A][B]
      Probability of a type A node choosing a type B node: pnn[A][B]

Input: Degree of connectivity: α
       Probability for a node to add a new arc: P_Add

Function arc_add(N, A, α, P_Add, CT, NT, pnc[ ][ ], pnn[ ][ ])
begin
    Store the type of each existing node in nt[ ] ;
    for i = 1 to |N| do
        With probability P_Add do
            Call create_weights(N, A, CT, NT, pnc[ ][ ], pnn[ ][ ], nt[i],
            preferential_weight[ ]) ;
            Choose a node n ∈ N using preferential_weight[ ] as a
            probability mass function;
            if i ≠ n then
                Add type arc_type arc (i, n) to A ;
            end
        end
    end
end
```

Figure 2.2.3: Arc add algorithm

matrix $p[\,][\,][\,]$ stores the probability of infection depending on characteristics like gender and disease stage of both nodes, the matrix $s[\,][\,]$ stores the cofactor scaling effect of one disease on another, and the vector $F[\,]$ stores the frequency of each arc type. The first three are estimated from the epidemiological literature, and $F[\,]$ is estimated from data.

When the algorithm is called upon to spread a disease d within our model, it finds arcs that link an infected node i with an susceptible node j for disease d. If such a pair of nodes is found, the algorithm examines the attributes of both nodes and their arc to determine the likelihood of transmission, scaled up by both nodes' cofactors from every other disease concurrently considered in the model. For example, if the model has only two diseases, say d and d^+, then the likelihood of transmission is then multiplied by $s[i$'s d^+ stage$][i$'s d stage$]$ and $s[j$'s d^+ stage$][j$'s d stage$]$. To determine the final likelihood of transmission of disease d from node i to node j, the likelihood is adjusted by the frequency of the arc linking both nodes. This is to say, in the previous example, node i transmits disease d to node j with likelihood equal to $1-(1-p[i$'s gender$][i$'s stage of $d] \times s[i$'s d^+ stage$][i$'s d stage$] \times s[j$'s d^+ stage$][j$'s d stage$]\,)^{F[arc_type]}$.

2.2.3 Dynamic Network Model with Disease Spread

Our full dynamic network model with disease spread is constructed using our extended preferential attachment, arc add, arc drop, and disease spread algorithms. Besides the parameters explained in the original model and the extended versions of the previously mentioned algorithms, there are four more inputs, $pnt[\,]$, $pnd[\,][\,]$, $d[\,][\,]$, and T_{Spread}. $pnt[\,]$ is the vector of node type frequencies, $pnd[\,][\,]$ is the initial prevalence of each disease in each node type, $d[\,][\,]$ is the duration of each stage of each disease, and T_{Spread} is the frequency with which the model will spread disease. The first two parameters can be derived from the data, the third can be taken from the epidemiology

Data: Set of arcs in current network: A
Set of nodes in current network: N
Set of genders for nodes: G
Set of diseases spreading in our model : D
Set of disease A's infected stages in our model: S_A
Probability of an infected node of gender A with disease B to transmit: $p[A][B][S_B]$
Scaling effect of disease A on disease B's transmission: $s[S_A][S_B]$

Input: Frequency of a type A arc: $F[A]$

Function spread_disease($A, N, G, D, S_1, \ldots, S_D, p[\,][\,][\,], s[\,][\,]$)
begin
  Store the stage of each disease of each node in $stage[\,][\,]$;
  Store the gender of each node in $gender[\,]$;
  for $a = 1$ ***to*** $|A|$ **do**
    for $d_1 = 1$ ***to*** $|D|$ **do**
      if arc a links an infected node n_1 with disease d_1 and an uninfected node n_2 **then**
        arc_type stores a's type ;
        $infection\_probability = p[gender[n_1]][stage[n_1][d_1]]$;
        for $d_2 = 1$ ***to*** $|D|$ **do**
          $infection\_probability = infection\_probability \times s[stage[n_1][d_2]][stage[n_1][d_1]] \times s[stage[n_2][d_2]][stage[n_2][d_1]]$
        end
        if $random(0,1) \geq (1 - infection\_probability)^{F[arc\_type]}$ **then**
          n_2 becomes infected with disease d_1 ;
        end
      end
    end
  end
end

Figure 2.2.4: Disease spread algorithm

literature, and the last can be fitted to observed data. The complete model description is detailed in Figure 2.2.5 and Figure 2.2.6.

Our model works in the following way: it starts with only one node in the network. Every T_{Add} weeks it adds to the current model New new nodes, each of which is assigned with a node type based on $pnt[\ ]$. Based on $pnd[\ ][\ ]$, each new node's initial status of all the diseases of concern are determined. To assign arcs between new nodes and existing nodes, the algorithm uses the preferential attachment algorithm. Every T_{Change} weeks it scans through the current network to add new arcs and to drop arcs among existing nodes following the arc change algorithm. Every T_{Spread} weeks it checks the spread of all diseases from infected nodes to susceptible nodes on arcs in the network using the spread algorithm. Every week, the disease status of every node will be examined. If it is their time to advance to the next stage, our model updates them and records their next advancement times based on $d[\ ][\ ]$. After $WEEK$ weeks, the run terminates.

Although the number of model parameters appears large, only 3 of them can not be directly taken from data or the epidemiology literature. The remaining 3 parameters can be fit to data, as we show in the next chapter.

Data: Set of contact types in our model: CT
Set of node types in our model: NT
Set of disease types in our model: D
Set of genders for nodes: G
Set of disease A's infected stages in our model: S_A
Probability of a node being type A: $pnt[A]$
Probability of a type A node with a disease B: $pnd[A][B]$
Probability of an infected node of gender A with disease B to transmit: $p[A][B][S_B]$
Scaling effect of disease A on disease B's transmission: $s[S_A][S_B]$
Duration of disease A's stages: $d[A][S_A]$
Number of weeks to run our model: WEEK

Input: Degree of connectivity: α
Probability for a node to add a new arc: P_{Add}
Probability for an existing arc to be dropped: P_{Drop}
Number of new nodes added at the same time: New
Time interval between adding new nodes: T_{Add}
Time interval between changing current arcs: T_{Change}
Time interval between spreading diseases: T_{Spread}
Frequency of a type A arc: $F[A]$

Figure 2.2.5: Dynamic preferential model - Part I

```
Function
dynamic_preferential_model
(WEEK, New, T_Add, T_Change, T_Spread, P_Add, P_Drop, α, F[A])
begin
    model_time = 0 ;
    N = 1; A = ∅ ;
    while model_time ≤ WEEK do
        if (model_time mod T_Add) = 0 then
            for i = 1 to NEW do
                Choose node (|N| + i)'s type, node_type, ∈ NT using pnt[ ] as
                a probability mass function ;
                for d = 1 to |D| do
                    With probability pnd[node_type][d] do
                        Choose (|N| + i)'s disease d's status ∈ S_d uniformly at
                        random ;
                    end
                end
            end
            Call preferential_attachment
            (N, A, New, CT, NT, pnc[ ][ ], pnn[ ][ ]) New times;
        end
        if (model_time mod T_Change) = 0 then
            Call arc_add (N, A, α, P_Add, CT, NT, pnc[ ][ ], pnn[ ][ ]);
            Call arc_drop (N, A, P_Drop);
        end
        if (model_time mod T_Spread = 0 then
            Call spread_disease (A, N, G, D, S_1, ..., S_D, p[ ][ ], s[ ][ ]) ;
        end
        for i = 1 to |N| do
            for d_1 = 1 to |D| do
                if model_time = node i's disease d_1's status advancement time
                then
                    stage = node i's disease d_1's status ;
                    Advance node i's disease d_1's status ;
                    Change node i's disease d_1's status advancement time
                    based on d[d_1][stage] ;
                end
            end
        end
        model_time = model_time + 1 ;
    end
end
```

Figure 2.2.6: Dynamic preferential model - Part II

CHAPTER III

DYNAMIC PREFERENTIAL NETWORK MODEL IN SEXUALLY TRANSMITTED DISEASES

Previously we introduced a basic framework for dynamic network modeling. In this chapter we are going to demonstrate how to calibrate our model to replicate both the real-life risky behavioral contact network and STD prevalence results.

Due to privacy concerns and the complexity of keeping track of human relationships, to date researchers have not been able to accurately depict a dynamic contact network (based on real human activity) on the scale of thousands of persons. The best dataset we have found comes from the fields of Social Science and Public Health, where researchers have been able to track static contact networks representing no more than hundreds of people (and there, only a portion of the relationships were tracked).

In this chapter we briefly describe a network study [19] titled "Networks, Norms, and HIV Risk Among Youth" (NNAHRAY). Following the introduction we present the static contact network constructed from the NNAHRAY data as well as the procedure of construction. We take the following steps to conduct our model parameterization. First we estimate some of the network parameters from NNAHRAY data and the other by fitting the network metrics in the NNAHRAY constructed network. Then we estimate the HIV and HSV-2 transmission parameters similarly. Most of them are extracted from the epidemiological literature, except for the risky behavioral contact frequencies. They are estimated by fitting the HIV and HSV-2 prevalence records in NNAHRAY data.

3.1 NNAHRAY

Public health workers conducted the NNAHRAY project in Bushwick, Brooklyn, NY between 2002 and 2004. In 1981, AIDS was first identified in [22], and there were 53 cases reported in New York City with HIV-related symptoms [38]. The antibody detection test for HIV was developed in 1984. In the following 10 years the incidence of AIDS in adults in New York City reported per year increased from fewer than 1000 cases to 9000 cases. However, both the incidence and the mortality rates declined dramatically around 1996 and remained constant after 1999 [18]. Prevention measures such as screening of the blood supply, free needle exchange, and safer sexual practices may have resulted in the drastic slowdown in the epidemic. More details of the HIV epidemic in New York City between 1985 and 2001 are reported in [18].

In its early years, the New York City AIDS epidemic was mostly localized in neighborhoods which suffered both economic and social marginalization [17]. Bushwick was one such community. Conducted between 2002 and 2004, the NNAHRAY Project [20, 21] was aimed at studying the risky relationships critical to the spread of HIV, including sexual relationships, needle sharing relationships and group sex event attendance relationships in Bushwick.

The interviewers started with an index group of injection drug users (IDUs) and a sample of young adults living in the area, and then found as many as possible of the interviewee's partners up to the fourth level in the relationship chain. In other words, the partner of the partner of the partner of the partner of the index group member was the last level to be interviewed in this study. Since not all partners could be reached, and some refused to be interviewed, some relationships either among the interviewees or between the interviewees and non-interviewees were missing in the study. All partners were involved directly or indirectly with at least one index group member in a risky relationship during the previous three months before being recruited.

The total number of initial group members was 112. Among them, 40 were IDUs.

The total number of first level, second level, third level, fourth and fifth level partners was 133, 118, 67, 34,and 1 respectively. In total, 465 interviewees were recruited and 424 risky relationships among them were recorded. The interviewees' self-reported records about the type of their relationships in the study fall into the following nine categories: "Group Sex and IDU", "Group Sex", "IDU and Sex", "IDU", "Sex and Group Sex", "Sex and IDU", "Sex", "Sex and Group Sex and IDU", and "Missing", where the category "Missing" means that the interviewee's comment on the type of relationship was not available.

The project also tested the interviewees for HIV and various other STDs, including herpes simplex virus-2 (HSV-2), syphilis, chlamydia, and gonorrhea. The laboratory testing showed that HIV and HSV-2 were the two most prevalent diseases in the testing, at 9% and 48% respectively.

3.2 Contact Network in NNAHRAY

To construct from NNAHRAY the contact network on top of which the spread of HIV and other STDs can be modeled, we disregarded 36 relationships in which both partners reported "Group Sex", because knowing that any two interviewees who were in such relationship assures only that they have had attended a group sex event together but not that they have had sexual contact with each other in the event. Furthermore, we exclude 2 relationships with both partners' comments on the type of relationship "Missing". The resulting NNAHRAY static contact network has 465 nodes and $424 - 36 - 2 = 386$ arcs.

Figure 3.2.1 shows the NNAHRAY static contact network. All nodes in green belong to the largest connected component. Table 3.2.1 summarizes the component size distribution in the network. Among 107 components, there are 53 isolated nodes and 28 dyads. This means that $(53 + (28 \times 2))/465 = 23\%$ of the interviewees possibly had either no relationships or one monogamous relationship in the three

months before being recruited. Compared with the largest component, which consists of ($206/465 \approx 44\%$) of all interviewees, the second largest component, is relatively small, $15/206 \approx 6\%$ of all interviewees.

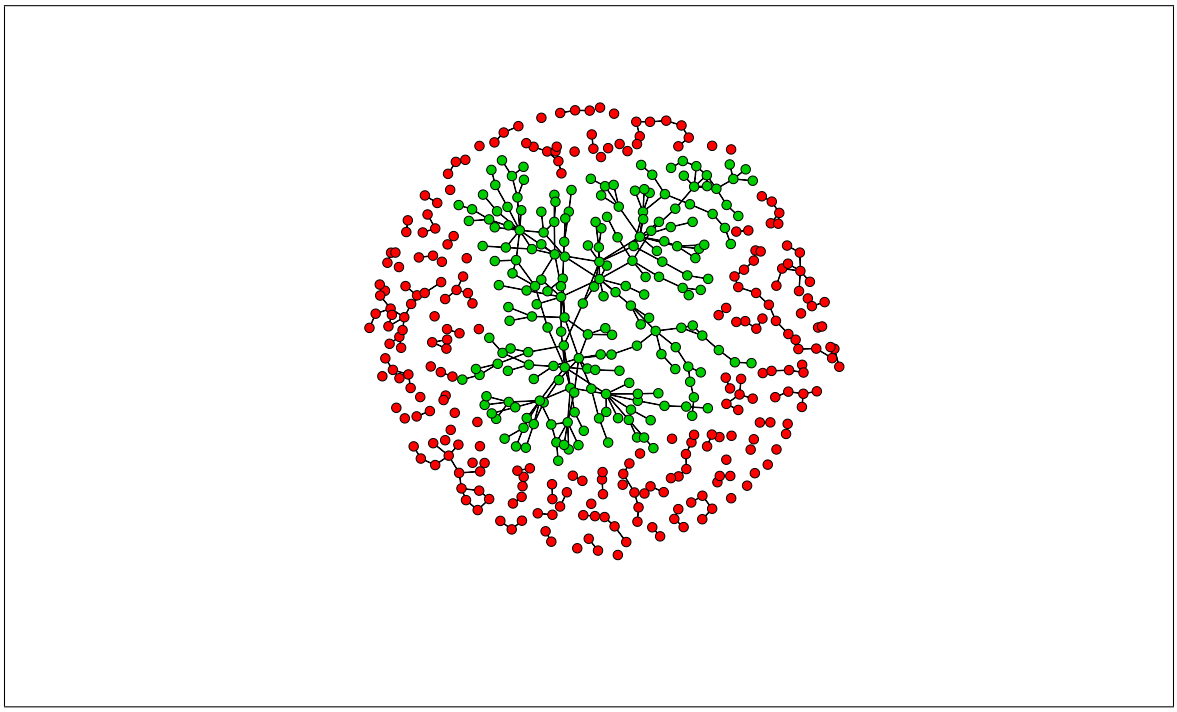

Figure 3.2.1: NNANRAY static contact network

Based on the self-reported attribute records of gender, injection drug using history, and sexual orientation of the interviewees, we can categorize the nodes in the contact network into 12 types: *FBN*, *FHN*, *FSN*, *FBI*, *FHI*, *FSI*, *MBN*, *MHN*, *MSN*, *MBI*, *MHI*, and *MSI*. *F* and *M* indicate a female and a male, respectively. *B* stands for bisexual orientation, *H* means homosexual orientation, and *S* represents straight (heterosexual) sexual orientation. *N* indicates a non injection drug user and *I* indicates an injection drug user. For example, a node with type *MBI* corresponds to a male bisexual who has had injected drugs in the previous three months before being recruited in the NNAHRAY study.

Ideally when we examine the answers from partners with regard to the type of relationship they share with each other, we expect they are the same. However, out of 386 relationships, we found that 105 relationships have one partner's comment on

Table 3.2.1: NNAHRAY component distribution

Component size	Number of components in the network
1	53
2	28
3	10
4	3
5	6
6	1
7	2
8	2
13	1
14	1
15	1
206	1

the relationship type while the other partner's comment is "Missing". Acknowledging that this may not be the ideal case, we use the one-sided comment as the type of those relationship arcs in the contact network.

The preliminary static contact network we thus constructed from the NNAHRAY data has each node belonging to one of the 12 types, $\{F, M\} \times \{B, H, S\} \times \{I, N\}$, and each arc belonging to one of the 3 types, $\{Sex, IDU, Sex\&IDU\}$, as mentioned above. However, we found three kinds of inconsistencies when we carefully examined the matching of node and arc types in the network.

The first kind of inconsistency could result from under-reporting of one partner. An example from the data is that one partner reported "Sex and IDU" while the other reported either only "Sex" or only "IDU". We deal with such inconsistency by assigning the corresponding arcs a type based on the more complete account.

The second kind of inconsistency could result from the interviewee's forgetfulness. For example, an interviewee may have self reported that he/she is homosexual but is found having a sexual relationship with another interviewee of the opposite gender. We fix this kind of inconsistency by updating the attribute from "homosexual" to

"bisexual" [Note that this attribute is intended to reflect their behavior, not their self-identification].

The third kind of inconsistency could result from the interviewee's intention to hide his/her identity. The first example from the data is that an interviewee self reported being an "NIDU" but is found being in an "IDU" relationship. Another example is that an interviewee self reported being "Straight" but is found in a "Sex" relationship with another of the same gender. We reassign the attribute records of these interviewees from "straight" to "bisexual" and from "NIDU" to "IDU".

In total we modified the type of 15 arcs, updated the type of 23 nodes from NIDU to IDU, and changed the type of 30 nodes from straight/homosexual to bisexual in the contact network.

For the resulting final static NNAHRAY contact network, we summarize the arc distribution of nodes in 12 types in Table 3.2.2. Since there are no nodes of type FHI after the type adjustment, their incidence rate is listed as NA in the table. For example, adjacent to nodes of type FSI, 27% of arcs are "Sex", 33% are "IDU", and 40% are "Sex & IDU".

Table 3.2.3 and Table 3.2.4 record the mixing ratio of nodes in one type with nodes in another type. For example, in the contact network, for nodes of type FBN, 18% of arcs link with FBN nodes, 2% with FHN nodes, 6% with FBI nodes, 8% with MBN nodes, 46% with MSN nodes, 3% with MBI nodes, and 17% with MSI nodes. After our type modification, nodes whose sexual orientation and drug using history are not compatible will not have any contact among them. Therefore in Table 3.2.3, for example, nodes of type FSN have no arcs to nodes of type FSN, since the former, who are female, straight, and non-injection drug users, do not have either sexual contact or needle sharing contact with nodes who are also female, straight, and non-injection drug users.

Table 3.2.2: Static NNAHRAY network node-arc type incidences

Type	Arc type	Fraction	Type	Arc type	Fraction
FBN	SEX	1.00	MBN	SEX	1.00
FBN	IDU	0.00	MBN	IDU	0.00
FBN	SE&ID	0.00	MBN	SE&ID	0.00
FHN	SEX	1.00	MHN	SEX	1.00
FHN	IDU	0.00	MHN	IDU	0.00
FHN	SE&ID	0.00	MHN	SE&ID	0.00
FSN	SEX	1.00	MSN	SEX	1.00
FSN	IDU	0.00	MSN	IDU	0.00
FSN	SE&ID	0.00	MSN	SE&ID	0.00
FBI	SEX	0.50	MBI	SEX	0.27
FBI	IDU	0.27	MBI	IDU	0.32
FBI	SE&ID	0.23	MBI	SE&ID	0.42
FHI	SEX	NA	MHI	SEX	0.42
FHI	IDU	NA	MHI	IDU	0.38
FHI	SE&ID	NA	MHI	SE&ID	0.19
FSI	SEX	0.27	MSI	SEX	0.27
FSI	IDU	0.33	MSI	IDU	0.56
FSI	SE&ID	0.40	MSI	SE&ID	0.17

3.3 Dynamic Preferential Network Model Calibration Metrics

We would like to use the snapshot of a real contact network provided by NNAHRAY to find reasonable model parameter values of α, P_{Add}, P_{Drop}, and $F[\,]$ so that our basic dynamic network model described in Chapter 2 can generate results similar to NNAHRAY, in terms of both network structure and STD prevalence. In this section we describe the metrics we use to measure network structure similarity.

3.3.1 Network Structure Metrics

Three interesting characteristics of the contact network in NNAHRAY are difficult to duplicate using network growth models: it has one dominantly large component and a few other components of size ≥ 3, in its largest component the nodes are not much more connected than they are in a tree structure, and certain nodes in it are much

Table 3.2.3: Static NNAHRAY Network node-node type incidences - I

Type	Type	Fraction	Type	Type	Fraction
FBN	FBN	0.18	FBI	FBN	0.06
FBN	FHN	0.02	FBI	FHN	0.02
FBN	FSN	0.00	FBI	FSN	0.00
FBN	FBI	0.06	FBI	FBI	0.13
FBN	FHI	0.00	FBI	FHI	0.00
FBN	FSI	0.00	FBI	FSI	0.06
FBN	MBN	0.08	FBI	MBN	0.05
FBN	MHN	0.00	FBI	MHN	0.00
FBN	MSN	0.46	FBI	MSN	0.15
FBN	MBI	0.03	FBI	MBI	0.18
FBN	MHI	0.00	FBI	MHI	0.02
FBN	MSI	0.17	FBI	MSI	0.34
FHN	FBN	0.50	FHI	FBN	NA
FHN	FHN	0.00	FHI	FHN	NA
FHN	FSN	0.00	FHI	FSN	NA
FHN	FBI	0.50	FHI	FBI	NA
FHN	FHI	0.00	FHI	FHI	NA
FHN	FSI	0.00	FHI	FSI	NA
FHN	MBN	0.00	FHI	MBN	NA
FHN	MHN	0.00	FHI	MHN	NA
FHN	MSN	0.00	FHI	MSN	NA
FHN	MBI	0.00	FHI	MBI	NA
FHN	MHI	0.00	FHI	MHI	NA
FHN	MSI	0.00	FHI	MSI	NA
FSN	FBN	0.00	FSI	FBN	0.00
FSN	FHN	0.00	FSI	FHN	0.00
FSN	FSN	0.00	FSI	FSN	0.00
FSN	FBI	0.00	FSI	FBI	0.05
FSN	FHI	0.00	FSI	FHI	0.00
FSN	FSI	0.00	FSI	FSI	0.10
FSN	MBN	0.11	FSI	MBN	0.02
FSN	MHN	0.00	FSI	MHN	0.00
FSN	MSN	0.64	FSI	MSN	0.15
FSN	MBI	0.03	FSI	MBI	0.11
FSN	MHI	0.00	FSI	MHI	0.02
FSN	MSI	0.22	FSI	MSI	0.55

more connected than the others. We want our model to be able to create network samples that look similar to the contact network in NNAHRAY, possessing these

Table 3.2.4: Static NNAHRAY Network node-node type incidences -II

Type	Type	Fraction	Type	Type	Fraction
MBN	FBN	0.10	MBI	FBN	0.03
MBN	FHN	0.00	MBI	FHN	0.00
MBN	FSN	0.24	MBI	FSN	0.05
MBN	FBI	0.06	MBI	FBI	0.18
MBN	FHI	0.00	MBI	FHI	0.00
MBN	FSI	0.04	MBI	FSI	0.15
MBN	MBN	0.24	MBI	MBN	0.07
MBN	MHN	0.08	MBI	MHN	0.02
MBN	MSN	0.00	MBI	MSN	0.00
MBN	MBI	0.08	MBI	MBI	0.17
MBN	MHI	0.14	MBI	MHI	0.17
MBN	MSI	0.00	MBI	MSI	0.17
MHN	FBN	0.00	MHI	FBN	0.00
MHN	FHN	0.00	MHI	FHN	0.00
MHN	FSN	0.00	MHI	FSN	0.00
MHN	FBI	0.00	MHI	FBI	0.04
MHN	FHI	0.00	MHI	FHI	0.00
MHN	FSI	0.00	MHI	FSI	0.08
MHN	MBN	0.33	MHI	MBN	0.27
MHN	MHN	0.50	MHI	MHN	0.04
MHN	MSN	0.00	MHI	MSN	0.00
MHN	MBI	0.08	MHI	MBI	0.38
MHN	MHI	0.08	MHI	MHI	0.15
MHN	MSI	0.00	MHI	MSI	0.04
MSN	FBN	0.25	MSI	FBN	0.06
MSN	FHN	0.00	MSI	FHN	0.00
MSN	FSN	0.57	MSI	FSN	0.12
MSN	FBI	0.08	MSI	FBI	0.11
MSN	FHI	0.00	MSI	FHI	0.00
MSN	FSI	0.10	MSI	FSI	0.24
MSN	MBN	0.00	MSI	MBN	0.00
MSN	MHN	0.00	MSI	MHN	0.00
MSN	MSN	0.00	MSI	MSN	0.00
MSN	MBI	0.00	MSI	MBI	0.05
MSN	MHI	0.00	MSI	MHI	0.01
MSN	MSI	0.00	MSI	MSI	0.42

characteristics and matching the NNAHRAY snapshot in other metrics as well.

We use 13 network structure metrics in total, defined as follows. When a metric

has more than one dimension, such as the ordered node degree list of a network, we quantify closeness by measuring the 1-norm distance between points in the high dimensional space.

1. *Size*: the total number of nodes in a network.

2. *Edge*: the total number of edges in a network.

3. *Lgcsize*: the total number of nodes in the largest component of a network.

4. *Lgcedge*: the total number of edges in the largest component of a network.

5. *Lgcdensity*: the density of the largest component of a network, equal to *lgcedge*/*lgcsize*.

6. *Arcremove*: the minimum number of arcs needed to be removed so that a network has no cycles. Let C be the set of components in a network, and let n_i and $a_i$be the number of nodes and arcs in component $i \in C$ respectively. *Arcremove* is equal to $\sum_{i \in C} max\{0, a_i - (n_i - 1)\}$.

7. *Componentdistmean*: the node-wise average component size. Let N be the total number of nodes, C be the set of components, and n_i be the number of nodes in each component $i \in C$. *Componentdistmean* is equal to $(\sum_{i \in C} n_i^2)/N$.

8. *Componentdistsd*: the standard deviation of the node-wise component size. Using the same notation as that in *Componentdistmean*, *Componentdistsd* is equal to $(\sum_{i \in C} n_i(n_i - Componentdistmean)^2)/(N-1)$.

9. *Degreemax*: the maximum node degree in a network.

10. *Degreemean*: the average node degree in a network.

11. *Degreesd*: the standard deviation of node degrees in a network.

12. *Orderdegreediffmean*: the average of the differences between the ordered node degree of a network and the ordered node degree of the contact network in

NNAHRAY. Let N be the total number of nodes, O be an ordering of nodes by decreasing degree (O_i is the node with i_{th} highest degree), and O' be a similar ordering of the 465 nodes of the NNAHRAY network. Let $d(j)$ be the degree of node j in the test network and $d'(j)$ be the degree of node j in the NNAHRAY newtork. *Orderdegreediffmean* is equal to $\sum_{j=1}^{465}(d'(O'_i) - d(O_i))/465$, where $d(j) = 0, \forall j \in \{N+1, \ldots, 465\}$.

13. *Orderdegreediffsd*: the standard deviation of nodes' ordered degree difference between the test network and the NNAHRAY network. Using the same notation as that in *Orderdegreediffmean*, *Orderdegreediffsd* is equal to $\sum_{j=1}^{465}((d'(O'_i) - d(O_i)) - \textit{Orderdegreediffmean})^2/(465-1)$

In addition to structural similarity, we also measure the similarity in HIV and HSV-2 prevalence between test networks and the NNAHRAY networks.

3.4 Dynamic Preferential Network Model Fitting

In this section we first introduce the disease parameters used in order to model HIV and HSV-2 prevalence. We then give a detailed summary of our adjusted network model to be used to fit the NNAHRAY contact network both structurally and in disease prevalence. We briefly describe the difficulty with fitting models to data following the traditional direct approach, and demonstrate the need to also model the study's network sampling method in order to have a good comparision. Finally, we demonstrate our model's ability to satisfactorily fit the NNAHRAY data.

3.4.1 The Spread of HIV and HSV-2

The parameters needed to model the spread of HIV and HSV-2 are taken directly from the epidemiological literature, adopted from the stage-based disease models suggested in [36] and [46]. Their results are summarized in Table 3.4.1 for HIV and Table 3.4.2 for HSV-2.

As Table 3.4.1 shows, any patient infected with HIV goes through four stages: primary, asymptomatic, symptomatic, and AIDS. The patient's transmission probability per coital act differs with gender and stage. Men are twice as likely to transmit the disease than women in general. The duration of each stage assumes a Weibull distribution with shape parameter 2.

In Table 3.4.2 we see that an HSV-2 infected individual experiences four major stages in order: primary ulcers, early latent with recurrent ulcers, latent with recurrent ulcers, and late latent without ulcers. The patient's transmission probability per coital act differs with gender and stage. The duration of each stage follows a Weibull distribution with shape parameter 2, except for the duration of the early latent stage, latent stage, and the intervals between recurrent ulcers, which follow an exponential distribution instead. The mean interval length between ulcer occurrence is 10 weeks for males and 12 weeks for females in the early latent stage, and 24 weeks for males and 32 weeks for females in the latent stage.

Also in Table 3.4.2 we note that this HSV-2 model considers HSV2 synergistic impact on HIV transmission. During the stages of primary ulcers and recurrent ulcers, patients are much more susceptible to HIV transmission than the other stages, by the multiplicative factor shown in the rightmost column.

Table 3.4.1: HIV Model

HIV stages	Mean duration (weeks)	Transmission probability per coital act	Transmission probability per coital act
		Male to Female	Female to Male
Primary	10	0.028	0.014
Asymptomatic	260	0.002	0.001
Symptomatic	208	0.006	0.003
AIDS	40	0.014	0.007

Table 3.4.2: HSV2 Model

HSV2 stages	Mean duration (weeks)	Transmission probability per coital act	Transmission probability per coital act	Cofactor on HIV transmission
		Male to Female	Female to Male	
Primary ulcers	3	0.300	0.150	25
Early latent	104	0.010	0.005	1
Latent	520	0.005	0.003	1
Late Latent	Life time	0.000	0.000	1
Recurrent Ulcers	1	0.200	0.100	10

3.4.2 Dynamic Preferential Network Model Description

With the HIV and HSV-2 parameters taken from the literature and the NNAHRAY attributes taken from that data set, four parameters remain, with three pertaining to our basic dynamic network model: α, P_{Add}, and P_{Drop} and the frequency parameter $F[\,]$ related to the disease spread algorithm. Before reporting on the parameter fitting, we first review our model, putting together all of the various pieces. All the other (non-fit) parameter values and the source from which we obtained their estimated values are summarized in Table 3.4.3 and Table 3.4.4 .

In the very beginning our model starts with one node and progresses through 10 years of time, week by week. Every $T_{Add} = 4$ weeks, it adds $New = 10$ nodes to the existing contact network. Upon entry to the network, after being assigned an age and gender according to the NNAHRAY distribution, each node is initialized to be one of $NT = 12$ types according again to the NNAHRAY distribution. Each node is distinguished by gender (Female and Male), sexual orientation (Bisexual, Homosexual, and Straight), and history of drug use (IDU and Non-IDU).

Table 3.4.3: Dynamic network model parameter value - I

Definition of parameter	Parameter notation	Value	Value source
Dynamic Network Model			
Set of contact types	AT	3 (SEX, IDU, SEX and IDU)	NNAHRAY [19]
Set of node types	NT	12 ($FBI, FHI, FSI, \ldots, MSN$)	NNAHRAY [19]
Set of diseases	D	2 (HIV and HSV-2)	NNAHRAY [19]
Age	Age	Uniformly distributed between 18 and 66	NNAHRAY [19]
Gender	$Gender$	57% male and 43% female	NNAHRAY [19]
Group type proportion within women:			
FBN	$pnt[FBN]$	22%	NNAHRAY [19]
FHN	$pnt[FHN]$	1%	NNAHRAY [19]
FSN	$pnt[FSN]$	43%	NNAHRAY [19]
FBI	$pnt[FBI]$	11%	NNAHRAY [19]
FHI	$pnt[FHI]$	0%	NNAHRAY [19]
FSI	$pnt[FSI]$	22%	NNAHRAY [19]
Group type proportion within men:			
MBN	$pnt[MBN]$	12%	NNAHRAY [19]
MHN	$pnt[MHN]$	3%	NNAHRAY [19]
MSN	$pnt[MSN]$	36%	NNAHRAY [19]
MBI	$pnt[MBI]$	9%	NNAHRAY [19]
MHI	$pnt[MHI]$	3%	NNAHRAY [19]
MSI	$pnt[MSI]$	37%	NNAHRAY [19]
Probability of a type A node having a type B contact	$pnc[A][B]$	See Table 3.2.2	NNAHRAY [19]
Probability of a type A node choosing a type B node	$pnn[A][B]$	See Table 3.2.3	NNAHRAY [19]
Number of weeks to run our model	$WEEK$	$(10 \times 52 = 520)$ weeks	Model setup
Number of new nodes added at the same time	New	10	Model setup
Time interval between adding new nodes	T_{Add}	4 weeks	Model setup
Time interval between changing current arcs	T_{Change}	4 weeks	Model setup
Time interval between spreading diseases	T_{Spread}	1 week	Model setup

Table 3.4.4: Dynamic network model parameter value - II

Parameter definition	Parameter Notation	Value	Value source
HIV and HSV-2			
HIV prevalence rate of newcomers from 1990 to 2002:			
Set of HIV's infected stages in our model	S_{HIV}	4	[18, 26]
IDU:	$pnd[IDU][HIV]$	50% with 3% decrease each year	[18, 26]
MSM:	$pnd[MSM][HIV]$	47% with 3.3% decrease each year	[18, 26]
General population:	$pnd[General][HIV]$	9 % with 0.6% decrease each year	[18, 26]
Transmission rate of an infected node of gender A with HIV	$p[A][HIV][S_{HIV}]$	See Table 3.4.1	[36, 46]
Duration of HIVs stages	$d[HIV][S_{HIV}]$	See Table 3.4.1	[36, 46]
HSV-2 prevalence rate of newcomers from 1990 to 2002:			
Set of HSV-2's infected stages in our model	S_{HSV-2}	5	[36, 46]
General population:	$pnd[General][HSV_2]$	21% before 1996 and 17% after 1996	[48]
Transmission rate of an infected node of gender A with HSV-2	$p[A][HSV-2][S_{HSV-2}]$	See Table 3.4.2	[36, 46]
Duration of HIVs stages	$d[HSV-2][S_{HSV-2}]$	See Table 3.4.2	[36, 46]
Scaling effect of HSV-2 on HIV's transmission	$s[S_{HSV-2}][S_{HIV}]$	See Table 3.4.2	[36, 46]

Based on its node type, each new node's infection status of HIV and HSV-2 is set according to both diseases' prevalence rates at the time [48, 18, 26]. An infected node's disease stage is selected randomly according to a uniform distribution. The HSV-2 prevalence rates before and after 1996 are taken from [48], and the HIV prevalence rates (and their change) among the general, IDU, and MSM populations are taken from [18, 26].

After the initialization of a new node's attributes and disease statuses is completed, our model assigns it to form an arc with an existing node. The assignment is conducted preferentially, according to the NNAHRAY incidences, by degree of connectivity α, and the current degree distribution in the network (see Section 2.2.1).

Every $T_{Change} = 4$ weeks, our model examines each existing arc and drops it with a probability P_{Drop}. It also examines each existing node and adds a new arc between it and another existing node (chosen the same way as arcs from new nodes) with probability P_{Add}.

Every $T_{Spread} = 1$ weeks, potential HIV and HSV-2 transmission is checked from the infected node to each susceptible node along an arc linking them. The transmission rate for HSV-2 in each unprotected sex contact is taken from [36, 46], and depends on gender and disease stage. For HIV, the transmission rate in each unprotected sex contact or needle-sharing contact is taken from [36, 46]. The duration of each stage of HIV and HSV-2 follows the distribution functions shown in Table 3.4.1 and Table 3.4.2.

To properly reflect the fact that many intervention policies targeting HIV took place and successfully diminished its spread in NYC between 1990 and 2000 [48], the frequency parameter $F[\,]$ is extended to four possible values in the disease spread algorithm, $F[SEX_{before}]$, $F[SEX_{after}]$, $F[IDU_{before}]$, and $F[IDU_{after}]$, in our model to capture the difference in HIV and HSV-2 transmission probabilities before and after 1996 among the population in NYC.

After $WEEK = 520$ weeks, we take a network snapshot of nodes and arcs that have existed in the last 3 months (similar to NNAHRAY's snapshot).

3.4.3 Initial Model Fitting Results

The contact network metric values measured in the NNAHRAY contact network, summarized in Table 3.4.5, are used concurrently to define closeness of a simulated network to the NNAHRAY contact network.

Table 3.4.5: Network metric and disease prevalence results of NNAHRAY

Definition of parameter	Parameter notation	Value
Disease prevalence:		
HIV prevalence		9 %
HSV-2 prevalence		48 %
Network structure:		
The number of nodes in a network sample	$Size$	465
The total number of edges in a network sample	$Edge$	386
The number of nodes of the largest component in a network sample	$Lgcsize$	206
The number of edges of the largest component in a network sample	$Lgcedge$	231
The density of the largest component in a network sample	$Lgcdensity$	1.12
The minimum number of arcs to be removed so that a network sample has no cycles	$Arcremove$	30
The average component weight of the nodes	$Component-distmean$	94.07
The standard deviation of component weight of the nodes	$Component-distsd$	99.99
The maximum node degree in a network sample	$Degreemax$	12
The average node degree in a network sample	$Degreemean$	1.66
The standard deviation of node degrees in a network sample	$Degreesd$	1.49

A contact network is considered close to the NNAHRAY contact network if all its network metric values fall within the ranges listed in Table 3.4.6. For high dimension network metrics, such as $Orderdegreedisffsd$, their ranges of closeness are set wider

than single dimension network metrics': their range are chosen with a 50% deviation while the range of most other metrics are set with a 10% deviation, except for Arcremove. Its range is set with a 25% deviation.

Table 3.4.6: Network metric close range

Network Metric	Specification	Value
$Size$	$465 \pm (465 \times 10\%)$	(512,419)
$Edge$	$386 \pm (386 \times 10\%)$	(424,347)
$Lgcsize$	$206 \pm (206 \times 10\%)$	(185,227)
$Lgcedge$	$254 \pm (254 \times 10\%)$	(229,279)
$Lgcdensity$	$1.12 \pm (1.12 \times 10\%)$	(1.00,1.23)
$Arcremove$	$30 \pm (30 \times 25\%)$	$(37.5, 22.5)$
$componentdistmean$	$94.07 \pm (94.07 \times 10\%)$	(85,103)
$componentdistsd$	$100 \pm (100 \times 10\%)$	(90,110)
$Degreemax$	$12 \pm (12 \times 10\%)$	(10,13)
$Degreemean$	$1.66 \pm (1.66 \times 10\%)$	(1.5,1.8)
$Desgreesd$	$1.49 \pm (1.49 \times 10\%)$	(1.34,1.64)
$Orderdegreediffemean$	$0.00 \pm (0.00 \times 50\%)$	$(-0.5, 0.5)$
$Orderdegreedisffsd$	$0.00 \pm (0.00 \times 50\%)$	$(-0.5, 0.5)$

Our modeling fitting takes place in two steps. First, we focus on fitting the model parameters related to network structures. Then we fit the parameters related to disease prevalence. In other words, after choosing the right structural parameter set of (α, P_{Add}, P_{Drop}), we continue on finding the frequency parameter set of ($F[SEX_{before}]$, $F[Sex_{later}]$, $F[IDU_{before}]$, $F[IDU_{later}]$).

The simulation experiments are programmed in C. At the end of each simulation experiment network structure analysis is performed in R using the network package [8] and the sna [7] package.

When we tune our model so that the size (number of nodes) of network snapshots are the same as that of the NNAHRAY contact network, many points in the $(\alpha, P_{Add}, P_{Drop})$ parameter space can produce network snapshots that match one target network metric value. When we demand the network snapshots and the NNAHRAY contact network to be structurally closer and closer, implying more

and more network metrics being fitted at the same time, there are fewer and fewer points in the parameter space that can make it happen. Eventually, we found that there is no point that produces snapshots with *Lgcsize*, *Lgcdensity*, *Degreemax*, and *Arcremove* network metric values concurrently close to those measured in the NNAHRAY contact network. Table 3.4.7 and Table 3.4.8 show an example of the mentioned fitting results.

This observation, on the one hand, points out an alarming insufficiency of the traditional single network metric fitting method. It can be used to produce "fitted" network results are in reality structurally far from the target network, as is the case with our model. On the other hand the lack of satisfying points in the $(\alpha, P_{Add}, P_{Drop})$ parameter space signals a missing ingredient in our model that needs to be included to completely capture the mechanism behind the NNAHRAY contact network structures. That ingredient turns out to be modeling how the NNAHRAY sample was collected.

3.4.4 Modeling the Study Sampling Procedure

The NNAHRAY data collection details show that interviewees were sampled with emphasis on IDUs, whose sexual/needle-sharing relationships were hard to be completely traced down since not all participants were willing to admit all relationships. The NNAHRAY contact network, as a result, is a biased/partial sample of the real contact network. This means that the network's real structure may be different from the network discovered by the sampling method deployed in the NNAHRAY study. In order to match the NNAHRAY contact network structures, we modeled the sampling algorithm, and used it to extract a contact network from our model network snapshots.

The main idea behind the sampling algorithm is to model the progression of discovery in the NNAHRAY dataset, so we keep the same number of nodes in each recruitment level as those in the corresponding level of the sampling process in NNAHRAY. The specifics of the sampling algorithm are summarized in Figure 3.4.1.

The sampling algorithm starts by randomly choosing the same number of NIDU nodes and IDU nodes as in NNAHRAY's initial sample. Those nodes (level one nodes) are stored in both $NodeList$ and $Current$, and one after another examined in the algorithm in order to find no more than $i[2]$ number of level two nodes. When a node is examined, the set $Neighbor$ stores its adjacent nodes and the set Arc stores its adjacent arcs. If the adjacent nodes are in $Neighbor$ but not in $NodeList$, they have not been recruited by the sampling algorithm so far. They are then added to New and $NodeList$ and their adjacent arcs are added to $EdgeList$. When either the number of level two nodes equals $i[2]$ ($|New| = i[2]$) or there are no more nodes left to be examined in $Current$ ($|Current| = 0$), the algorithm stops the current level examination and repeats the same procedure to find the next level nodes.

When the algorithm completes the node recruiting iterations with $|EdgeList|$ greater than A, it stops. Otherwise, the sampling algorithm randomly chooses, two at a time, non-adjacent nodes in the sample contact network using the component weight $c[\,]$ as the probability mass function and adds the arc (i, j) to $EdgeList$. The arc adding process stops when $|EdgeList| = A$, and the algorithm completes its iterations.

When the sampling algorithm is used to extract contact networks from our model network snapshots in comparison with the NNAHRAY contact network, we use parameter values based on the NNAHRAY study, as summarized in Table 3.4.9.

Table 3.4.7: Initial model fitting result example - I

Network Metric			size	edge	lgcsize	lgcedge	lgcdensity	arcremove	component-distmean	component-distsd	degree-max	degree-mean	degree-sd	orderdegree-diffemean
NNAHRAY			465	386	206	231	1.12	30.00	94.07	99.99	12.00	1.66	1.49	0.00
α	P_{Add}	P_{Drop}												
0	0.02	0.02	465	390	335	350	1.05	16.50	243.31	147.25	11.40	1.68	1.73	-0.02
0	0.04	0.02	465	390	338	360	1.06	22.90	247.11	149.28	11.70	1.68	1.82	-0.02
0	0.04	0.04	465	383	316	329	1.04	13.10	217.92	144.19	10.70	1.65	1.61	0.01
0	0.06	0.02	465	401	346	374	1.08	29.20	258.23	149.41	12.90	1.72	1.92	-0.06
0	0.06	0.04	465	393	328	347	1.06	20.00	233.33	146.54	11.10	1.69	1.69	-0.03
0	0.06	0.06	465	382	293	304	1.04	12.30	194.73	133.48	8.60	1.64	1.50	0.02
0	0.08	0.02	465	393	345	368	1.07	24.60	256.29	149.36	12.90	1.69	1.94	-0.03
0	0.08	0.04	465	391	324	343	1.06	20.60	227.85	145.52	11.40	1.68	1.75	-0.02
0	0.08	0.06	465	387	313	328	1.05	16.10	212.81	143.65	9.70	1.66	1.58	0.00
0	0.08	0.08	465	382	302	312	1.03	11.30	199.65	139.76	9.00	1.64	1.47	0.02
0.25	0.02	0.02	465	393	329	351	1.07	23.20	235.20	146.96	14.40	1.69	1.88	-0.03
0.25	0.04	0.02	465	394	346	367	1.06	21.90	258.76	148.79	15.30	1.70	1.90	-0.04
0.25	0.04	0.04	465	387	326	341	1.05	16.00	230.32	146.12	11.90	1.67	1.71	-0.01
0.25	0.06	0.02	465	396	346	371	1.07	26.10	258.34	149.66	15.60	1.70	2.04	-0.04
0.25	0.06	0.04	465	395	343	366	1.07	23.70	253.84	149.24	12.50	1.70	1.82	-0.04
0.25	0.06	0.06	465	386	318	330	1.04	13.70	219.25	145.12	11.40	1.66	1.60	0.00
0.25	0.08	0.02	465	400	345	374	1.08	30.00	257.86	148.75	15.30	1.72	2.10	-0.06
0.25	0.08	0.04	465	389	324	340	1.05	17.40	228.63	145.26	12.10	1.67	1.80	-0.01
0.25	0.08	0.06	465	385	317	332	1.05	16.50	217.33	145.25	10.60	1.66	1.65	0.00
0.25	0.08	0.08	465	377	284	294	1.03	11.00	178.15	133.96	10.50	1.62	1.53	0.04
0.5	0.02	0.02	465	396	330	354	1.07	25.50	234.77	148.01	21.50	1.70	2.17	-0.04
0.5	0.04	0.02	465	402	351	376	1.07	26.00	265.27	149.35	28.50	1.73	2.35	-0.07
0.5	0.04	0.04	465	392	323	340	1.05	17.90	226.02	146.17	18.40	1.69	1.89	-0.03
0.5	0.06	0.02	465	401	347	374	1.08	28.10	260.50	148.84	19.10	1.73	2.20	-0.07
0.5	0.06	0.04	465	387	330	347	1.05	17.70	235.33	148.08	15.70	1.67	1.88	-0.01
0.5	0.06	0.06	465	382	294	306	1.04	13.40	190.58	137.36	13.70	1.64	1.71	0.02
0.5	0.08	0.02	465	406	352	384	1.09	33.20	267.24	149.62	19.60	1.75	2.30	-0.09
0.5	0.08	0.04	465	395	342	364	1.06	22.40	253.27	148.90	14.70	1.70	1.94	-0.04
0.5	0.08	0.06	465	393	331	350	1.06	19.90	237.79	147.54	14.10	1.69	1.82	-0.03
0.5	0.08	0.08	465	385	308	323	1.05	15.60	206.27	142.83	11.80	1.66	1.64	0.00

Table 3.4.8: Initial model fitting result example - II

Network Metric			size	edge	lgcsize	lgcedge	lgcdensity	arcremove	component-distmean	component-distsd	degree-max	degree-mean	degree-sd	orderdegree-diffemean
NNAHRAY			465	386	206	231	1.12	30.00	94.07	99.99	12.00	1.66	1.49	0.00
α	P_{Add}	P_{Drop}												
0.75	0.02	0.02	465	393	329	349	1.06	20.70	234.49	147.09	28.20	1.69	2.44	-0.03
0.75	0.04	0.02	465	400	333	359	1.08	26.90	240.81	147.34	35.70	1.72	2.69	-0.06
0.75	0.04	0.04	465	394	325	344	1.06	19.80	229.09	146.54	24.30	1.70	2.15	-0.04
0.75	0.06	0.02	465	405	343	376	1.10	33.80	254.76	148.54	29.00	1.74	2.56	-0.08
0.75	0.06	0.04	465	399	337	361	1.07	24.70	245.68	148.16	23.90	1.72	2.25	-0.06
0.75	0.06	0.06	465	389	317	334	1.05	17.50	218.10	145.39	17.50	1.67	1.82	-0.01
0.75	0.08	0.02	465	416	359	394	1.10	35.90	279.23	146.80	44.20	1.79	3.17	-0.13
0.75	0.08	0.04	465	399	333	357	1.07	24.60	240.04	148.13	33.70	1.71	2.57	-0.05
0.75	0.08	0.06	465	390	323	344	1.07	22.70	225.90	145.89	20.40	1.68	2.09	-0.02
0.75	0.08	0.08	465	387	312	325	1.04	14.00	211.61	143.84	15.50	1.66	1.79	0.00
1	0.02	0.02	465	435	374	404	1.08	31.50	301.99	144.68	153.30	1.87	7.58	-0.21
1	0.04	0.02	465	421	344	377	1.09	33.80	259.50	143.38	96.00	1.81	5.14	-0.15
1	0.04	0.04	465	414	340	368	1.08	28.10	252.91	144.88	82.90	1.78	4.55	-0.12
1	0.06	0.02	465	425	361	401	1.11	40.20	282.39	147.48	90.00	1.83	5.02	-0.17
1	0.06	0.04	465	419	351	380	1.08	30.40	267.04	146.67	101.30	1.80	5.16	-0.14
1	0.06	0.06	465	402	310	329	1.06	19.80	210.89	141.04	59.80	1.73	3.42	-0.07
1	0.08	0.02	465	446	371	425	1.14	55.00	297.98	144.51	135.20	1.92	7.00	-0.26
1	0.08	0.04	465	412	347	377	1.08	30.80	260.38	147.96	81.50	1.77	4.65	-0.11
1	0.08	0.06	465	400	329	353	1.07	25.10	234.94	145.74	47.40	1.72	3.13	-0.06
1	0.08	0.08	465	392	307	322	1.05	15.70	205.30	142.40	63.60	1.68	3.49	-0.02
1.25	0.02	0.02	465	439	344	379	1.10	36.10	262.19	137.65	145.10	1.89	7.25	-0.23
1.25	0.04	0.02	465	428	329	360	1.08	32.10	250.49	127.14	128.50	1.84	6.65	-0.18
1.25	0.04	0.04	465	404	282	299	1.05	18.10	188.83	126.48	126.20	1.74	6.36	-0.08
1.25	0.06	0.02	465	417	333	364	1.09	32.30	246.72	138.12	108.80	1.79	5.73	-0.13
1.25	0.06	0.04	465	450	387	426	1.10	40.30	322.75	142.14	245.00	1.94	11.52	-0.28
1.25	0.06	0.06	465	404	312	331	1.06	20.20	217.34	137.27	118.10	1.74	5.92	-0.07
1.25	0.08	0.02	465	446	363	407	1.11	44.40	292.33	132.20	161.00	1.92	8.18	-0.26
1.25	0.08	0.04	465	424	339	368	1.08	30.30	254.67	139.93	140.10	1.82	7.00	-0.16
1.25	0.08	0.06	465	405	306	320	1.04	15.00	208.01	138.97	135.10	1.74	6.56	-0.08
1.25	0.08	0.08	465	403	294	308	1.04	14.90	198.08	132.62	153.10	1.73	7.31	-0.07

Data: Set of nodes in the network: N
Total number of target arcs: A
Component weight in the network: $c[\ ]$
Level of sampling : L
Total number of index sampling nodes : I
Total number of index IDU sampling nodes: II
Total number of interviewees for each sampling level: $i[L]$

$NodeList = \{\emptyset\}$; $EdgeList = \{\emptyset\}$; $Current = \{\emptyset\}$;

Choose uniformly at random II IDU nodes and $(I - II)$ NIDU nodes from N;
Add them to $NodeList$ and to $Current$;

for $k = 1$ ***to*** L **do**
  $Neighbor = \{\emptyset\}$; $New = \{\emptyset\}$; $Arc = \{\emptyset\}$;
  $position = 1$;
  while $|New| \leq (i[k])$ *AND* $position \leq |Current|$ **do**
    Store all neighbors of the $position^{th}$ node in $Current$ in $Neighbor$;
    Store arcs of the $position^{th}$ node in $Current$ in Arc ;
    Add nodes in ($Neighbor$ - ($NodeList \cap Neighbor$)) to New ;
    Add nodes in ($Neighbor$ - ($NodeList \cap Neighbor$)) to $NodeList$;
    Add arcs in (Arc - ($EdgeList \cap Arc$)) to $EdgeList$;
    $Neighbor = \{\emptyset\}$;
    $Arc = \{\emptyset\}$;
    $position = position + 1$;
  end
  $Current = New$;
end
while $|EdgeList| \leq A$ **do**
  Choose any two nodes, i and j, from $NodeList$ using $c[\ ]$ as a probability mass function ;
  if *arc* (i, j) *is not in EdgeList* **then**
    arc (i, j) is added to $EdgeList$;
  end
end

Figure 3.4.1: Sampling algorithm

Table 3.4.9: Sampling algorithm parameter value

Parameter definition	Parameter notation	Value	Value source
Number of target arcs	A	386	NNAHRAY [19]
Level of sampling	L	4	NNAHRAY [19]
Number of index sampling nodes	I	112	NNAHRAY [19]
Number of index IDU sampling nodes	II	40	NNAHRAY [19]
Limit of interviewees for each sampling level	i[L]	$(133, 118, 67, 34)$	NNAHRAY [19]

3.4.5 Sample Model Fitting Results

In Table 3.4.10 and Table 3.4.11 we first show all the sample network structural fitting results without the arc adding step. The results where $P_{Add} < P_{Drop}$ are omitted because their sample networks are more loosely connected than the NNAHRAY contact network. Each row in Table 3.4.10 lists the average network structure metric values of the 100 samples, 10 samples from each one of 10 model contact networks, at the indicated value of (α, P_{Add}, P_{Drop}). In the sample networks without additional arc adding, the number of arcs in sample networks is often slightly lower than that in the NNAHRAY contact network and the largest component is often smaller and less dense. None of the points produces satisfying structural fitting results.

In Table 3.4.12 and Table 3.4.13 the sample network fitting results with the arc adding step in the sampling algorithm. Compared to the previous fitting results in Table 3.4.10, the results here are closer to the NNAHRAY contact network. There are two sets of parameters,$(\alpha, P_{Add}, P_{Drop}) = (0.25, 0.02, 0, 02)$ and $(0.50, 0.06, 0.06)$, meeting the overall structural fitting requirement.

The parameter set $(\alpha, P_{Add}, P_{Drop}) = (0.50, 0.06, 0, 06)$ has closer structural metrics values to the NNAHRAY contact network than the set $(\alpha, P_{Add}, P_{Drop})$ = (0.25,0.02, 0,02) does, as summarized in Table 3.4.14. For this reason we consider it the best parameter set to model the NNAHRAY contact network. Figure 3.4.2 shows one sample network produced by this particular set of parameters. In this graph, and

many other graphs produced by the chosen parameter set, we can now see three unique characteristics of the contact network in NNAHRAY: one dominantly large component and a few other smaller components, a largest component which is slightly more connected than a tree, and some densely-connected nodes in the largest component.

The frequency of unprotected SEX before 1996, $F[SEX_{before}]$, the frequency of unprotected SEX after 1996, $F[SEX_{later}]$, the frequency of needle-sharing before 1996, $F[IDU_{before}]$, and the frequency of needle-sharing after 1996, $F[IDU_{later}]$ are estimated by fitting the HIV prevalence and HSV-2 prevalence rates at the same time. The ($F[SEX_{before}]$,$F[SEX_{later}]$, $F[IDU_{before}]$, $F[IDU_{later}]$) are estimated to be (0.4, 0.1, 0.25, 0.1) times/week, because when combined with the parameter set $(\alpha, P_{Add}, P_{Drop})$ = (0.50, 0.06, 0.06), their prevalence results of both diseases at the end of 10 years, 12 years, and 15 years, as shown in Table 3.4.15, converge to the target disease prevalence values.

We conclude this section with a final note. Although the model requires a large number of parameters, the large majority of our model parameters are estimated directly from the data and from the epidemiological literature, and require no fitting. Our model fits just seven parameters from the data while meeting thirteen target metrics, whose interactions are too complicated to manipulate; in fact, until we modeled the sampling procedure it was impossible to fit. Our model's ability to overcome the data fitting challenge, together with its realistic model assumptions, such as the dynamics of nodes and arcs, leads us to believe that it may be more suitable than previous models for studying the interaction between disease spread and human behavior.

Table 3.4.10: Network metric values of sampling without the adding arc step - I

Network metric			size	edge	lgcsize	lgcedge	lgcdensity	arcremove	component-distmean	component-distsd	degree-max	degree-mean	degree-sd	orderdegreediff-diffmean	oderdegree-diffsd
NNAHRAY contact network			465	386	206	231	1.12	30.00	94.07	99.99	12.00	1.66	1.49	0.00	0.00
α	P_{Add}	P_{Drop}													
0	0.02	0.02	465	356	108	107	1.00	1.30	36.32	42.23	8.90	1.53	1.29	0.13	0.56
0	0.04	0.02	465	373	224	227	1.01	4.30	121.70	105.41	9.60	1.60	1.44	0.06	0.49
0	0.04	0.04	465	359	92	91	0.99	0.70	33.26	37.35	7.60	1.54	1.23	0.12	0.53
0	0.06	0.02	465	382	280	288	1.03	9.50	176.13	130.80	9.50	1.64	1.49	0.02	0.52
0	0.06	0.04	465	368	192	195	1.01	4.70	97.19	88.95	8.80	1.58	1.36	0.08	0.51
0	0.06	0.06	465	358	100	99	0.99	1.00	35.81	40.22	7.30	1.54	1.20	0.12	0.61
0	0.08	0.02	465	385	310	323	1.04	14.40	207.87	144.31	9.80	1.65	1.60	0.01	0.57
0	0.08	0.04	465	373	245	250	1.02	6.70	138.53	115.45	8.90	1.61	1.44	0.06	0.51
0	0.08	0.06	465	372	191	194	1.01	4.40	94.83	89.11	8.70	1.60	1.36	0.06	0.54
0	0.08	0.08	465	361	175	175	1.00	1.60	81.88	81.03	7.10	1.55	1.24	0.11	0.54
0.25	0.02	0.02	465	361	125	125	1.00	1.50	48.76	54.64	11.00	1.55	1.40	0.11	0.43
0.25	0.04	0.02	465	370	204	208	1.02	5.40	104.78	95.27	12.00	1.59	1.51	0.07	0.53
0.25	0.04	0.04	465	361	146	146	1.00	1.50	56.57	62.06	8.60	1.55	1.28	0.11	0.48
0.25	0.06	0.02	465	378	271	279	1.03	9.10	166.77	126.67	11.50	1.63	1.63	0.03	0.49
0.25	0.06	0.04	465	373	250	254	1.02	5.30	139.30	120.46	9.60	1.60	1.43	0.06	0.45
0.25	0.06	0.06	465	361	145	146	1.01	3.00	55.59	64.23	8.50	1.55	1.27	0.11	0.56
0.25	0.08	0.02	465	387	308	325	1.06	18.30	206.93	141.86	12.50	1.66	1.76	0.00	0.62
0.25	0.08	0.04	465	376	261	266	1.02	6.50	157.98	122.38	10.10	1.62	1.51	0.05	0.50
0.25	0.08	0.06	465	367	187	190	1.01	4.00	87.41	88.03	9.30	1.58	1.39	0.08	0.46
0.25	0.08	0.08	464	363	129	129	1.00	1.40	54.05	56.38	8.30	1.56	1.27	0.10	0.65
0.5	0.02	0.02	465	356	136	137	1.00	1.50	53.16	59.75	18.20	1.53	1.64	0.13	0.63
0.5	0.04	0.02	465	370	200	202	1.01	4.10	99.53	92.93	16.00	1.59	1.65	0.07	0.53
0.5	0.04	0.04	465	360	125	126	1.00	2.10	47.13	54.09	11.30	1.55	1.37	0.11	0.48
0.5	0.06	0.02	465	385	292	302	1.03	11.50	187.82	136.26	17.30	1.66	1.81	0.00	0.60
0.5	0.06	0.04	465	372	198	202	1.02	4.60	100.25	91.54	12.00	1.60	1.48	0.06	0.47
0.5	0.06	0.06	465	362	168	170	1.01	2.40	70.62	77.10	10.80	1.56	1.36	0.10	0.45
0.5	0.08	0.02	465	391	315	330	1.05	16.80	214.99	144.42	17.80	1.68	1.90	-0.02	0.69
0.5	0.08	0.04	465	373	241	247	1.03	8.40	138.62	113.58	13.20	1.61	1.65	0.06	0.52
0.5	0.08	0.06	465	368	201	206	1.02	5.50	99.45	94.93	10.00	1.58	1.47	0.08	0.46
0.5	0.08	0.08	464	363	153	153	1.00	1.40	65.93	68.32	9.70	1.56	1.34	0.10	0.54

Table 3.4.11: Network metric values of sampling without the adding arc step - II

Network metric			size	edge	lgcsize	lgcedge	lgcdensity	arcremove	component-distmean	component-distsd	degree-max	degree-mean	degree-sd	orderdegreediff-diffmean	oderdegree-diffsd
NNAHRAY contact network			465	386	206	231	1.12	30.00	94.07	99.99	12.00	1.66	1.49	0.00	0.00
α	P_{Add}	P_{Drop}													
0.75	0.02	0.02	464	357	123	123	1.00	1.70	48.41	53.28	22.60	1.54	1.81	0.12	0.94
0.75	0.04	0.02	465	378	262	269	1.02	7.70	151.93	126.33	40.50	1.63	2.52	0.03	1.51
0.75	0.04	0.04	464	357	129	130	1.00	2.30	58.60	58.02	18.20	1.54	1.58	0.13	0.78
0.75	0.06	0.02	465	389	312	325	1.04	14.10	212.27	143.62	25.80	1.67	2.20	-0.01	0.99
0.75	0.06	0.04	465	367	182	185	1.01	4.10	87.95	84.55	19.00	1.58	1.70	0.08	0.67
0.75	0.06	0.06	465	358	161	162	1.00	1.80	68.78	72.68	12.20	1.54	1.40	0.12	0.53
0.75	0.08	0.02	465	393	315	332	1.05	18.40	214.74	144.68	27.40	1.69	2.32	-0.03	1.18
0.75	0.08	0.04	465	379	287	298	1.04	12.40	179.95	135.89	20.70	1.63	1.90	0.03	0.71
0.75	0.08	0.06	465	372	205	210	1.02	5.50	104.04	97.06	15.00	1.60	1.55	0.06	0.50
0.75	0.08	0.08	464	361	128	129	1.00	2.70	51.02	55.33	12.20	1.56	1.39	0.11	0.61
1	0.02	0.02	464	363	165	167	1.00	2.90	82.80	75.18	67.90	1.56	3.73	0.10	3.03
1	0.04	0.02	464	384	223	231	1.02	8.40	133.64	101.78	97.00	1.65	5.03	0.01	4.35
1	0.04	0.04	464	355	161	163	1.01	2.80	69.82	75.43	30.10	1.53	2.06	0.14	1.24
1	0.06	0.02	465	384	266	276	1.03	10.40	157.69	127.06	58.60	1.65	3.39	0.01	2.42
1	0.06	0.04	465	370	237	243	1.02	7.10	129.18	113.97	54.10	1.59	2.99	0.07	2.09
1	0.06	0.06	463	350	155	156	1.01	2.70	63.49	70.87	38.70	1.51	2.26	0.16	1.70
1	0.08	0.02	465	403	328	348	1.06	21.40	232.68	146.67	71.80	1.73	4.03	-0.07	3.02
1	0.08	0.04	465	379	275	285	1.04	11.70	166.23	131.51	63.70	1.63	3.47	0.03	2.62
1	0.08	0.06	465	368	212	217	1.02	6.10	105.77	101.27	27.50	1.58	2.01	0.08	1.07
1	0.08	0.08	465	365	176	178	1.01	3.30	77.51	82.05	21.10	1.57	1.66	0.09	0.73
1.25	0.02	0.02	427	327	129	130	0.99	1.50	81.86	51.63	105.30	1.52	5.30	0.26	5.51
1.25	0.04	0.02	463	378	198	201	1.00	4.00	122.20	83.37	128.30	1.63	6.30	0.03	5.75
1.25	0.04	0.04	441	348	167	170	1.01	4.80	93.24	77.35	104.40	1.57	5.12	0.16	5.14
1.25	0.06	0.02	465	409	287	300	1.04	14.20	197.14	125.31	166.80	1.76	8.03	-0.10	7.35
1.25	0.06	0.04	465	399	262	271	1.03	10.00	168.72	118.80	169.30	1.72	8.03	-0.06	7.41
1.25	0.06	0.06	463	353	126	128	1.00	2.50	51.00	55.30	42.00	1.52	2.43	0.14	1.80
1.25	0.08	0.02	465	398	290	304	1.04	15.30	191.96	128.14	109.10	1.71	5.49	-0.05	4.70
1.25	0.08	0.04	465	406	299	314	1.05	16.10	207.14	130.43	147.40	1.74	7.15	-0.08	6.43
1.25	0.08	0.06	465	369	194	200	1.02	6.70	103.16	90.33	64.20	1.59	3.43	0.07	2.72
1.25	0.08	0.08	461	347	122	123	1.00	1.70	45.12	53.19	29.60	1.51	1.95	0.17	1.50

Table 3.4.12: Network metric values of sampling with the adding arc step - I

Network metric			size	edge	lgcsize	lgcedge	lgcdensity	arcremove	component-distmean	component-distsd	degree-max	degree-mean	degree-sd	orderdegreediff-diffmean	oderdegree-diffsd
NNAHRAY contact network			465	386	206	231	1.12	30.00	94.07	99.99	12.00	1.66	1.49	0.00	0.00
α	P_{Add}	P_{Drop}													
0	0.02	0.02	465	386	197	218	1.11	22.13	88.82	93.99	9.91	1.66	1.38	0.00	0.46
0	0.04	0.02	465	386	236	255	1.08	20.00	125.14	114.20	9.36	1.66	1.45	0.00	0.46
0	0.04	0.04	465	386	213	234	1.10	22.00	102.81	102.96	8.08	1.66	1.34	0.00	0.50
0	0.06	0.02	465	387	277	294	1.06	18.25	168.98	132.27	10.31	1.66	1.56	0.00	0.50
0	0.06	0.04	465	386	255	272	1.07	18.55	144.45	122.75	9.26	1.66	1.45	0.00	0.47
0	0.06	0.06	465	386	223	244	1.09	21.73	112.02	107.89	8.43	1.66	1.36	0.00	0.49
0	0.08	0.02	465	388	302	319	1.06	18.49	198.99	140.66	10.23	1.67	1.64	-0.01	0.54
0	0.08	0.04	465	387	282	299	1.06	17.89	174.90	133.95	9.66	1.66	1.51	0.00	0.51
0	0.08	0.06	465	386	247	265	1.08	19.38	135.81	118.97	8.22	1.66	1.38	0.00	0.52
0	0.08	0.08	465	386	222	242	1.09	21.19	111.50	107.28	7.66	1.66	1.33	0.00	0.53
0.25	0.02	0.02	465	386	208	230	1.11	22.83	98.22	100.10	11.03	1.66	1.47	0.00	0.45
0.25	0.04	0.02	465	386	253	272	1.08	19.78	141.43	122.56	12.48	1.66	1.58	0.00	0.46
0.25	0.04	0.04	465	386	213	236	1.11	23.86	103.59	102.70	9.97	1.66	1.41	0.00	0.49
0.25	0.06	0.02	465	387	278	294	1.06	17.69	170.40	132.15	11.40	1.66	1.62	0.00	0.49
0.25	0.06	0.04	465	386	244	263	1.08	19.47	132.89	118.33	9.91	1.66	1.45	0.00	0.46
0.25	0.06	0.06	465	386	227	248	1.09	21.58	116.71	109.83	8.85	1.66	1.37	0.00	0.51
0.25	0.08	0.02	465	389	306	323	1.06	18.15	204.25	142.06	12.81	1.67	1.75	-0.01	0.57
0.25	0.08	0.04	465	387	276	292	1.06	17.28	168.09	131.92	11.77	1.66	1.57	0.00	0.50
0.25	0.08	0.06	465	386	259	277	1.07	19.54	148.83	124.79	9.16	1.66	1.46	0.00	0.50
0.25	0.08	0.08	464	386	226	247	1.09	21.38	115.15	109.50	8.39	1.66	1.37	0.00	0.54
0.5	0.02	0.02	465	386	210	232	1.11	23.29	99.47	101.18	15.17	1.66	1.62	0.00	0.53
0.5	0.04	0.02	465	386	262	280	1.07	18.48	152.01	126.44	20.89	1.66	1.85	0.00	0.68
0.5	0.04	0.04	465	386	214	236	1.11	22.94	103.59	103.40	14.18	1.66	1.52	0.00	0.44
0.5	0.06	0.02	465	388	293	309	1.06	17.80	187.80	137.84	19.54	1.67	1.91	-0.01	0.71
0.5	0.06	0.04	465	386	253	271	1.07	19.06	142.82	122.40	12.15	1.66	1.57	0.00	0.45
0.5	0.06	0.06	465	386	218	240	1.10	22.76	107.47	105.40	10.96	1.66	1.43	0.00	0.46
0.5	0.08	0.02	465	391	307	326	1.06	20.29	204.98	142.30	16.33	1.68	1.90	-0.02	0.66
0.5	0.08	0.04	465	387	289	305	1.06	17.17	182.30	136.80	13.09	1.67	1.69	-0.01	0.56
0.5	0.08	0.06	465	386	247	265	1.08	19.67	135.51	119.55	10.91	1.66	1.49	0.00	0.47
0.5	0.08	0.08	465	386	236	256	1.09	21.49	124.69	114.06	9.65	1.66	1.41	0.00	0.48

Table 3.4.13: Network metric values of sampling with the adding arc step - II

Network metric			size	edge	lgcsize	lgcedge	lgcdensity	arcremove	component-distmean	component-distsd	degree-max	degree-mean	degree-sd	orderdegreediff-diffmean	oderdegree-diffsd
NNAHRAY contact network			465	386	206	231	1.12	30.00	94.07	99.99	12.00	1.66	1.49	0.00	0.00
α	P_{Add}	P_{Drop}													
0.75	0.02	0.02	465	386	221	243	1.10	23.00	110.15	106.61	34.91	1.66	2.29	0.00	1.33
0.75	0.04	0.02	465	387	259	278	1.07	19.63	149.26	124.90	28.73	1.67	2.16	-0.01	1.04
0.75	0.04	0.04	464	386	210	234	1.12	25.60	99.19	101.34	23.31	1.66	1.79	0.00	0.83
0.75	0.06	0.02	465	390	293	311	1.06	18.60	188.35	138.29	24.97	1.68	2.17	-0.02	0.98
0.75	0.06	0.04	465	386	248	267	1.08	19.74	137.00	120.04	18.26	1.66	1.76	0.00	0.61
0.75	0.06	0.06	464	386	203	229	1.13	26.42	93.61	98.02	12.45	1.66	1.48	0.00	0.56
0.75	0.08	0.02	465	394	313	333	1.06	20.90	213.79	144.04	34.46	1.70	2.54	-0.03	1.39
0.75	0.08	0.04	465	388	290	306	1.06	17.48	183.48	137.26	27.73	1.67	2.09	-0.01	0.97
0.75	0.08	0.06	465	386	249	268	1.08	20.22	137.92	120.33	16.69	1.66	1.65	0.00	0.56
0.75	0.08	0.08	465	392	314	333	1.06	19.81	214.03	144.37	14.80	1.69	1.77	-0.03	0.56
1	0.02	0.02	462	390	238	262	1.11	24.76	130.13	113.77	87.50	1.69	4.56	-0.02	3.91
1	0.04	0.02	465	396	274	293	1.07	19.98	168.76	128.25	73.63	1.70	4.03	-0.04	3.12
1	0.04	0.04	462	386	194	224	1.16	31.18	85.92	93.04	25.35	1.67	1.91	0.00	1.10
1	0.06	0.02	465	404	301	320	1.06	20.20	200.44	138.42	100.40	1.74	5.10	-0.08	4.27
1	0.06	0.04	465	387	247	268	1.08	21.28	136.00	120.22	44.79	1.66	2.73	0.00	1.75
1	0.06	0.06	461	386	211	241	1.15	31.26	100.98	102.25	33.70	1.67	2.20	0.00	1.51
1	0.08	0.02	465	415	337	362	1.07	26.38	247.51	145.54	108.34	1.79	5.62	-0.12	4.73
1	0.08	0.04	465	391	288	306	1.06	19.00	181.71	136.68	50.76	1.68	2.94	-0.02	1.97
1	0.08	0.06	465	386	244	265	1.09	22.62	131.83	118.36	39.96	1.66	2.47	0.00	1.50
1	0.08	0.08	462	386	200	228	1.14	28.33	91.41	96.44	18.71	1.67	1.61	0.00	0.76
1.25	0.02	0.02	441	399	271	304	1.19	34.69	175.38	123.74	184.42	1.82	8.93	-0.05	8.73
1.25	0.04	0.02	463	410	283	303	1.09	21.15	186.71	126.32	167.92	1.77	8.12	-0.10	7.52
1.25	0.04	0.04	450	392	232	267	1.19	35.66	128.03	110.78	98.63	1.74	5.12	-0.02	4.92
1.25	0.06	0.02	464	416	301	324	1.08	24.28	208.24	129.40	159.17	1.79	7.79	-0.13	7.03
1.25	0.06	0.04	464	394	255	276	1.09	22.53	146.69	121.33	89.76	1.69	4.62	-0.03	3.84
1.25	0.06	0.06	457	386	212	247	1.17	36.06	103.44	102.34	67.12	1.69	3.65	0.00	3.17
1.25	0.08	0.02	464	397	275	296	1.08	22.28	171.20	127.20	80.13	1.71	4.39	-0.05	3.55
1.25	0.08	0.04	465	404	284	305	1.08	21.70	183.10	129.36	123.93	1.74	6.15	-0.07	5.34
1.25	0.08	0.06	464	393	240	262	1.10	23.40	132.91	113.78	75.65	1.69	4.11	-0.03	3.31
1.25	0.08	0.08	463	386	213	243	1.14	30.70	102.22	103.83	58.76	1.67	3.07	0.00	2.41

Table 3.4.14: Closest parameter sets

Network metric			size	edge	lgc-size	lgc-edge	lgc-density	arc-remove	component-distmean	component-distsd
NNAHRAY			465	386	206	231	1.12	30.00	94.07	99.99
After adding arcs										
α	P_{Add}	P_{Drop}								
0.25	0.2	0.2	465	386	208	230	1.11	22.83	98.22	100.10
0.5	0.6	0.6	465	386	218	240	1.10	22.76	107.47	105.40
Before adding arcs										
α	P_{Add}	P_{Drop}								
0.25	0.02	0.02	465	361	125	125	1.00	1.50	48.76	54.64
0.5	0.06	0.06	465	362	168	170	1.01	2.40	70.62	77.10

Network metric			degree-max	degree-mean	degree-sd	orderdegree-diffmean	oderdegree-diffsd
NNAHRAY			12.00	1.66	1.49	0.00	0.00
After adding arcs							
α	P_{Add}	P_{Drop}					
0.25	0.2	0.2	11.03	1.66	1.47	0.00	0.45
0.5	0.6	0.6	10.96	1.66	1.43	0.00	0.46
Before adding arcs							
α	P_{Add}	P_{Drop}					
0.25	0.02	0.02	11.00	1.55	1.40	0.11	0.43
0.5	0.06	0.06	10.80	1.56	1.36	0.10	0.45

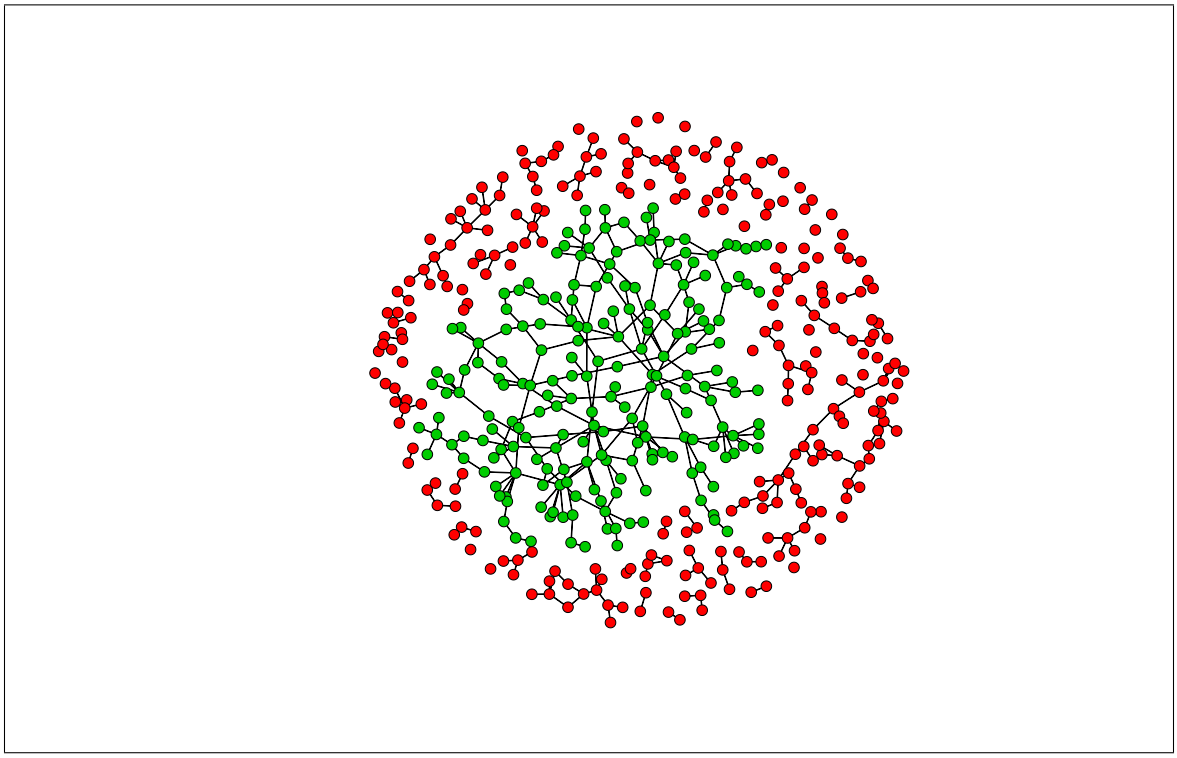

Figure 3.4.2: Model contact network sample

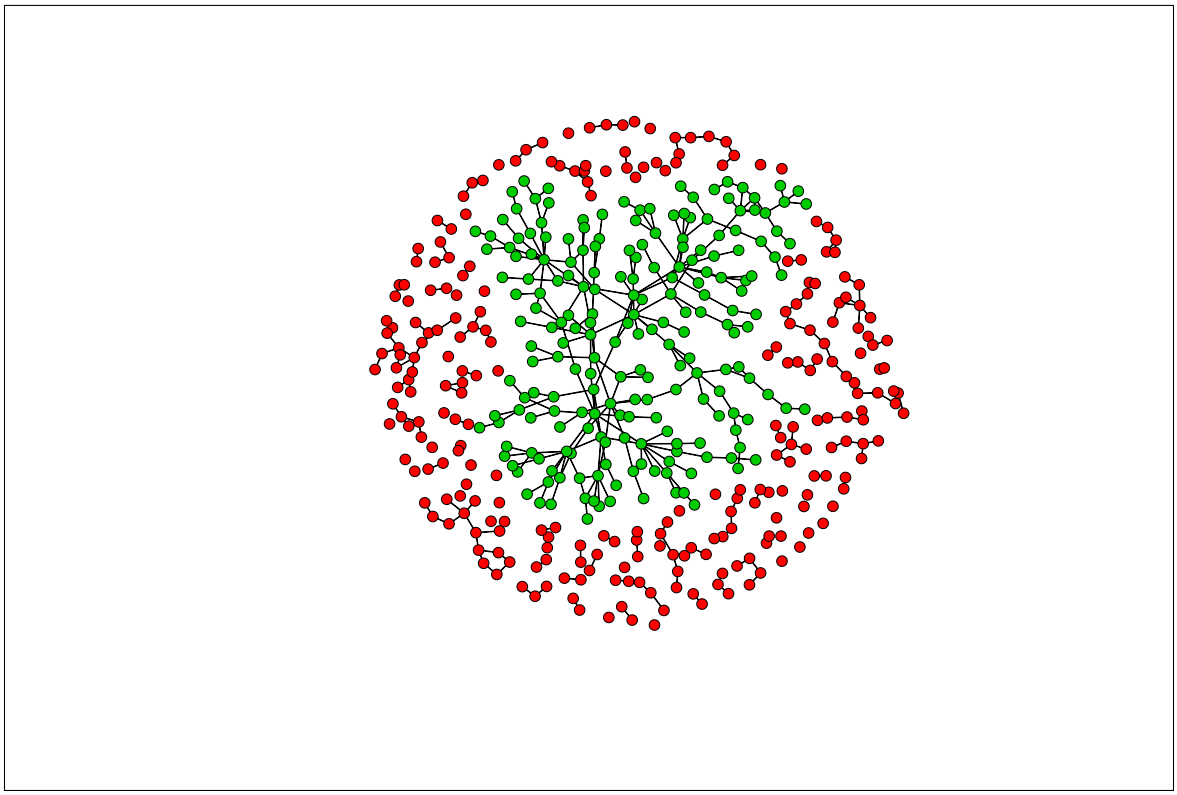

Figure 3.4.3: NNANRAY contact network

Table 3.4.15: Prevalence Prediction versus Simulation Time

Number of years	α	P_{Add}	P_{Drop}	populationhiv	populationhsv_2	samplehiv	samplehsv_2
NNAHRAY				NA	NA	0.09	0.48
10	0.5	0.06	0.06	0.13	0.39	0.15	0.42
12	0.5	0.06	0.06	0.10	0.38	0.13	0.42
15	0.5	0.06	0.06	0.07	0.41	0.09	0.45

CHAPTER IV

DYNAMIC NETWORK MODEL COMPUTATIONAL RESULTS

With our dynamic network model in hand, we now test two important hypotheses of disease spread modeling. The first hypothesis is that the dynamics of the network affect the prediction of disease spread. The second hypothesis is that the structure of a contact network on top of which a disease spreads affects the prediction of disease spread.

In this chapter we first test both hypotheses by comparing the network structures and the HIV prevalence prediction of our model and other static network models all calibrated with NNAHRAY data. We later test the hypothesis regarding the network structures by observing the prediction change within our model when using various single metric fitting methods and show how they differ.

To conclude this chapter, we illustrate the potential use of our model to analyze the efficacy of prevention/intervention policies for controlling HIV spread. This part can be extended to conduct cost effectiveness analysis of prevention/intervention policies when desired.

4.1 Disease Prevalence and Network Dynamic

We use NNAHRAY data to calibrate four static models and our dynamic network model and compare their HIV prevalence simulation results with the HIV prevalence record in New York City (NYC) from 1990 to 2002 in three different testings. The static models are named as follows: Preferential, Grandom, Configuration, and Compartmental. The first three are static contact network models, and each aims

at capturing at least one structural property discovered in the contact network in NNAHRAY. The last one is a variant of the known compartmental disease model that disregards the contact network structures of any population. Three different testings are set up: compare the results found in the model samples with NYC record considering HSV-2 synergy; compare the results in the model with NYC record; compare the model samples with NYC record without considering HSV-2 synergy.

All the simulation coding is done in C programming language and the pseudocode is included in Appendix A. We first explain each model's definition and high-level implementation scheme. Then we discuss the results of HIV prevalence comparison for each modeling strategy. We conclude this section with a summary of the comparison results and the implications for the hypothesis that if the network dynamic is incorporated into modeling, then a more accurate disease prediction result can be obtained.

4.1.1 Dynamic

Dynamic is our model calibrated as in Chapter 3 to fit the network structures and both HIV and HSV-2 prevalence results in NNHARAY. When running our model to extract HIV prevalence results between 1990 and 2002, we use the fitted parameter values, summarized in Table 4.1.1, and the extracted parameter values in Table 3.4.3. The implementation details of Dynamic are explained in Section 2.2.3.

Table 4.1.1: Dynamic network model parameter fitting results with NNAHRAY

Definition of parameter	Parameter Notation	Value
Degree of connectivity	α	0.50
Probability for a node to add a new arc	P_{Add}	0.06
Probability for an existing arc to be dropped	P_{Drop}	0.06
Frequency of unprotected SEX before 1996	$F[SEX_{before}]$	0.4 times/week
Frequency of unprotected SEX after 1996	$F[SEX_{later}]$	0.1 times/week
Frequency of needle-sharing before 1996	$F[IDU_{before}]$	0.25 times/week
Frequency of needle-sharing after 1996	$F[IDU_{later}]$	0.1 times/week

4.1.2 Preferential

Preferential is one of the static preferential attachment network models, originally proposed in [6]. The static preferential attachment network model in general aims at producing networks with the power law degree distribution by making certain nodes more attractive to add arcs with than the others.

As shown in Figure A.1 and Figure A.2 in Appendix A, to initialize the static contact network, Preferential starts with only one node in the network and adds $(WEEK/T_{Add} \times New)$ number of new nodes consecutively. Upon addition to the existing network each node is assigned with a node type based on $pnt[\,][\,]$. Based on $pnd[\,][\,]$, each new node's initial status of all diseases of concern are determined. Also upon addition to the existing network each new node is attached to an existing node following the preferential attachment algorithm. Values of the Preferential parameters related to network formation are listed in Table A.1 in Appendix A.

Every T_{Spread} weeks, the algorithm checks the spread of all diseases from infected nodes to the connected susceptible nodes in the network using the spread algorithm. Each week, the disease status of every node will be examined. If it is their time to advance to the next stage, our model updates them and records their next advancement time based on $d[\,][\,]$. Note that the the disease spread algorithm used in Preferential is the same as that used in Dynamic. Values of the Preferential parameters related disease spread are listed in Table A.2 in Appendix A.

There are two main differences between Preferential and Dynamic. The first difference is that although Preferential considers the node dynamics, such as the progression of disease stage in each node, Preferential does not incorporate arc dynamics, such as arc removal and creation as relationships change over time. Dynamic includes both node and arc dynamics. The second difference is regarding to the initial disease status. Given that all nodes are assumed to be present at the same time (and before the diseases start spreading) in Preferential, the proportion of the nodes initially

infected with HIV or HSV-2 is based on the levels of the population to be modeled. Since in our model nodes are added over time, the proportions of newly added nodes infected with HIV or HSV-2 differ each year.

4.1.3 Grandom

Grandom is a variant of the static general random network model. The static general random network model aims at producing networks whose expected degrees equal a chosen degree sequence [10]. Specifically, in an N-node graph with degrees $w = (w_1, w_2, ..., w_N)$, the probability of having an arc between nodes i and j is $(w_i \times w_j)/\sum_{k=1}^{N} w_k$. Note that the chosen degree sequence w needs to satisfy conditions that $(max_k\{w_k\})^2 \leq \sum_{k=1}^{N} w_k$ as well as that $\sum_{k=1}^{N} w_k$ is an even number to avoid self-loops and hyper arcs in a network.

Figure A.3 in Appendix A summarizes the algorithm used to add arcs between a set of nodes in Grandom. The algorithm *Grandom_AddArc* has two components. The first component generates an appropriate degree sequence, $n[\,]$, based on the input degree distribution w. The second component asigns arcs between all pairs of nodes based on $grandom\_weight[\,]$, which is calculated based on the degree sequence $n[\,]$, the node types of both nodes, the probability $pnn[\,][\,]$, and the probability $pnc[\,][\,]$. For example, when calculating the probability of an IDU node i having an IDU arc with another IDU node j, the *grandom_weight* is equal to $(n[i] \times n[j])/\sum_{k \in N} n[k] \times pnn[IDU][IDU] \times pnc[IDU][IDU]$, where N is the total number of nodes in the network of concern. If two nodes are not compatible regarding the chosen arc type, for example an IDU node and a non-IDU node with an IDU arc, their *grandom_weight* is equal to zero.

Details of Grandom are summarized in Figure A.4 and Figure A.5 in Appendix A. To initialize the static contact network, Grandom starts with only one node in

the network and adds ($WEEK/T_{Add} \times New$) new nodes consecutively. Upon addition to the network each node is assigned with a type based on $pnt[\,][\,]$. Based on $pnd[\,][\,]$, each new node's initial statuses of all the diseases of concern are determined. As soon as the nodes are all in place, the arc assignment among them follows the *Grandom_AddArc* algorithm. Values of the Grandom parameters related network formation are listed in Table A.3 in Appendix A.

Once the contact network is initialized, Grandom adopts the same disease spreading process as that in Preferential. The related parameter values are also listed in Table A.3.

4.1.4 Configuration

Configuration is a modified version of the static configuration network models, which aim at producing networks whose node degree ideally follows a chosen degree sequence. A widely-adopted algorithm proposed in [10] used to create networks with node degree sequences close to the desired one is defined as follows. Suppose the chosen degree sequence in a N-node graph is equal to $w = (w_1, w_2, ..., w_N)$. To determine how the arcs are distributed among the n nodes in the network, the algorithm creates another network with $\sum_{i=1}^{N} w_i$ nodes that are partitioned into N groups. Each group i consists of w_i nodes. After finding a random perfect matching for the $\sum_{i=1}^{N} w_i$ nodes in the created network, the algorithm adds an arc to the original network between each pair of node i and j if there is a node in group i matched with a node in group j in the created network.

Figure A.6 in Appendix A summarizes the algorithm used to add arcs between a set of nodes in Configuration. Similar to *Grandom_AddArc*, the first component of $Configuration\_AddArc$ is to generate an appropriate degree sequence, $n[\,]$, based on the input degree distribution w. The second component, however, is different from that in *Grandom_AddArc*. To determine which node to add an arc to,

$Configuration\_AddArc$ relies on the $configuration\_weight[\,]$, consisting of the number of unmatched nodes of each group in the created network $n_1[\,]$, the node types of both nodes, the probability $pnn[\,][\,]$, and the probability $pnc[\,][\,]$. For instance, when $Grandom\_AddArc$ calculates the $configuration\_weight$ of all the other nodes for an IDU node to form an IDU arc with, the resulting $configuration\_weight$ for any non-IDU node's is equal to zero, and for an IDU node j's $configuration\_weight$ is equal to $(n_1[j])/\sum_{k\in N} n_1[k] \times pnn[IDU][IDU] \times pnc[IDU][IDU]$.

Details of Configuration are summarized in Figure A.7 and Figure A.8 in Appendix A. As soon as the nodes are properly initialized and in place, the arc assignment among them adopts the $Configuration\_AddArc$ algorithm detailed in Figure A.6. Values of the Configuration parameters related network formation are listed in Table A.5 in Appendix A. Configuration adopts the same disease spreading process as that in Preferential and Grandom. The related parameter values are also listed in Table A.5.

4.1.5 Compartmental

Compartmental is a network model version of the traditional epidemiological differential equation model. Since in Compartmental any two nodes in compatible compartments are linked by an arc, it has a much denser network than the other network models have. We discount the disease transmission probability in Compartmental by a constant equal to the graph density in the NNAHRAY network to make a closer disease prevalence comparison with the other network models.

Figure A.9 in Appendix A describes the algorithm used to add arcs between a set of nodes in Compartmental. The algorithm $Compartmental\_AddArc$ in general examines all pairs of nodes. If a pair of nodes i and j are type compatible judged by the arc type randomly chosen based on the node i's type$pnc[nt[i]][\,]$, then the algorithm $Compartmental\_AddArc$ forms an arc between them.

Figure A.10 in Appendix A describes the algorithm used to spread disease in Compartmental. The algorithm *Compartmental_SpreadDisease* is similar to that in the previously mentioned models except that each attempt to transmit a disease from the infected to the susceptible is discounted by a scaling parameter p. For our data set, we estimate p to be equal to $386/(465 \times 464/2) \approx 0.0035$, since there are 386 arcs in the contact network in NANHRAY, and the network can accommodate no more than $465 \times 464/2 = 107880$ arcs.

Details of Compartmental are summarized in Figure A.11 and Figure A.12 in Appendix A. As soon as the nodes are properly initialized and in place, the arc assignment among them follows the *Compartmental_AddArc* in A.9. When the contact network is formed, *Compartmental* adopts *Compartmental_SpreadDisease* in Figure A.10 to spread the disease among the nodes. The parameter values are all listed in Table A.5 in Appendix A.

4.1.6 Computational Results And Discussion

Three testings are used to analyze the HIV prevalence results in the models. Testing 1 intends to simulate HIV transmission within a network of larger size than that of the contact network in NNAHRAY and compare the HIV prevalence results in the network model samples. Testing 2 aims at simulating HIV transmission within a network of size approximately equal to the contact network in NNAHRAY, then sample from the network using a similar approach to that of NNAHRAY, and analyze directly the HIV prevalence results in the networks. Testing 3 follows testing 1 but ignores HSV-2's effect on HIV spreading. We run 100 independent computational experiments for each model to simulate HIV transmission from 1990 to 2002. For testing 1 and 3, we draw a network sample from each of the 100 networks generated by the model.

We estimate the HIV prevalence of a population whose composition is similar to

that of NNAHRAY using the NYC HIV prevalence record between 1990 and 2002. We first assume a network/population in which nodes join at the same rate as that in Dynamic over the years. The proportion of node types is set to be the same as that in NNAHRAY. Each node's HIV and HSV-2 status is probabilistically determined by its timing of joining the network and its type [18, 26]. For all the infected nodes, their disease stages advance with time. Therefore an HIV-infected node may die at any time before 2002. We obtained our estimate from the constructed network between 1990 and 2002.

Table 4.1.2: HIV annual prevalence results of all models

		1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Testing 1	Dynamic	0.30	0.29	0.27	0.26	0.26	0.25	0.24	0.23	0.20	0.18	0.16	0.14	0.12
	Preferential	0.31	0.39	0.40	0.40	0.40	0.40	0.40	0.39	0.38	0.36	0.34	0.32	0.29
	Grandom	0.31	0.34	0.34	0.34	0.33	0.33	0.32	0.31	0.29	0.28	0.26	0.23	0.21
	Configuration	0.32	0.46	0.47	0.48	0.49	0.49	0.49	0.48	0.47	0.46	0.44	0.41	0.38
	Compartmental	0.32	0.49	0.52	0.54	0.56	0.57	0.57	0.57	0.57	0.56	0.54	0.51	0.47
Testing 2	Dynamic	0.34	0.30	0.28	0.27	0.26	0.25	0.25	0.23	0.20	0.18	0.16	0.14	0.12
	Preferential	0.31	0.39	0.40	0.40	0.40	0.40	0.40	0.39	0.38	0.36	0.34	0.31	0.29
	Grandom	0.31	0.34	0.34	0.34	0.33	0.33	0.32	0.31	0.29	0.28	0.26	0.24	0.21
	Configuration	0.31	0.46	0.47	0.48	0.48	0.48	0.48	0.48	0.47	0.45	0.43	0.40	0.37
	Compartmental	0.31	0.37	0.39	0.39	0.40	0.40	0.40	0.39	0.39	0.37	0.35	0.33	0.30
Testing 3	Dynamic	0.29	0.27	0.25	0.24	0.22	0.21	0.20	0.19	0.17	0.15	0.13	0.11	0.09
	Preferential	0.31	0.35	0.35	0.35	0.35	0.35	0.34	0.33	0.32	0.31	0.29	0.27	0.24
	Grandom	0.31	0.32	0.32	0.32	0.32	0.31	0.30	0.29	0.28	0.26	0.24	0.22	0.20
	Configuration	0.32	0.37	0.38	0.38	0.39	0.39	0.39	0.38	0.37	0.36	0.35	0.33	0.30
	Compartmental	0.31	0.37	0.39	0.40	0.41	0.42	0.43	0.43	0.43	0.42	0.41	0.39	0.37

Table 4.1.2 summarizes the annual HIV prevalence results of all models under the different strategies. We notice that the models' prevalence prediction results do not vary much between testing 1 and testing 2, except for Compartmental, whose prediction results in testing 2 is 36% lower than those in testing 1. This observation implies the disease prevalence prediction using the traditional compartmental models of disease spreading may be sensitive to the population size. The results obtained in contact network models, either static or dynamic, appear insensitive to population size.

All the models have lower prevalence prediction results in testing 3 than in the other strategies, as would be expected if a spread cofactor is ignored. Comparing the

results in testing 1 and testing 3, the amount of prediction reduction due to neglecting HSV-2's effect on HIV across all models is between 5% and 25%. Grandom is the least affected model while Dynamic is the most sensitive one among all.

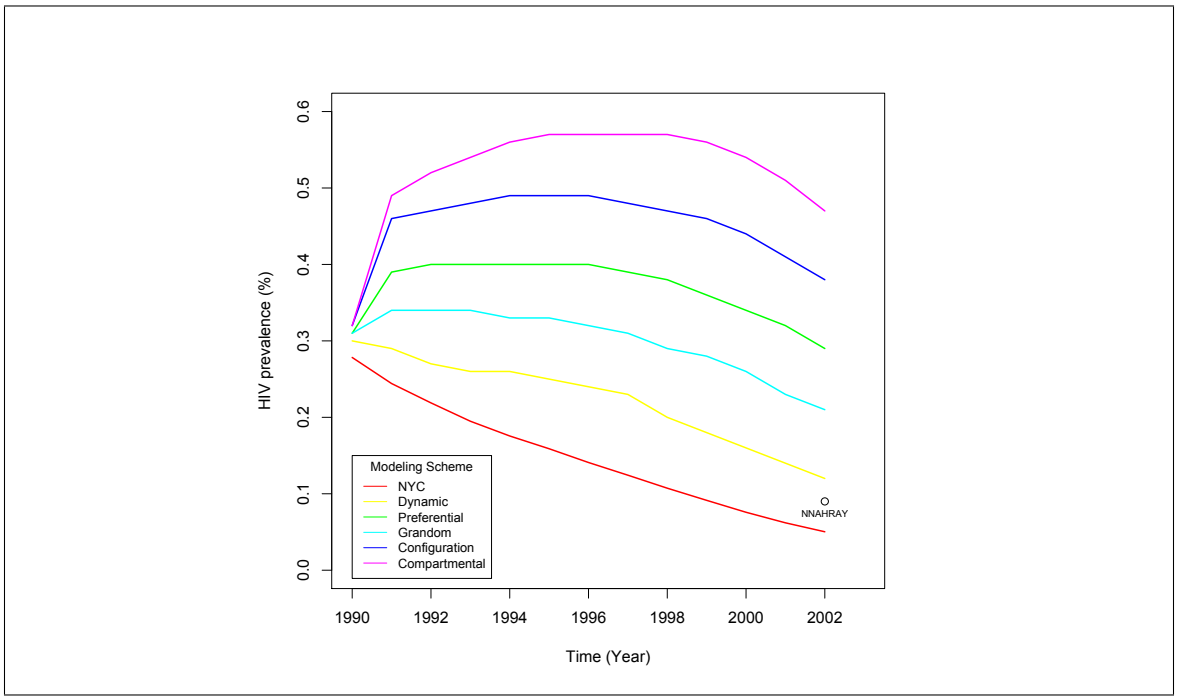

Figure 4.1.1: HIV prevalence of testing 1

Results from the table for each testing along with the previously-mentioned HIV prevalence estimate, NYC, are plotted in Figure 4.1.1, Figure 4.1.2, and Figure 4.1.3 for testing 1, testing 2, and testing 3 respectively. All figures clearly show that across three testings the relative prediction magnitude of all network models is the same. The order, from the highest to the lowest, is Configuration, Preferential, Grandom, and Dynamic, which is also the closest to the estimate for NYC from the literature. The prediction results from Compartmental may be more sensitive to the network size change than to the transmission rate change, as its relative magnitude with network models changes in testing 2 but not in testing 3.

Besides observing the HIV prevalence difference in models and in testings, we would also like to learn about their network structural differences. Since network structure measurement for models in testing 1 and in testing 3 is performed on samples

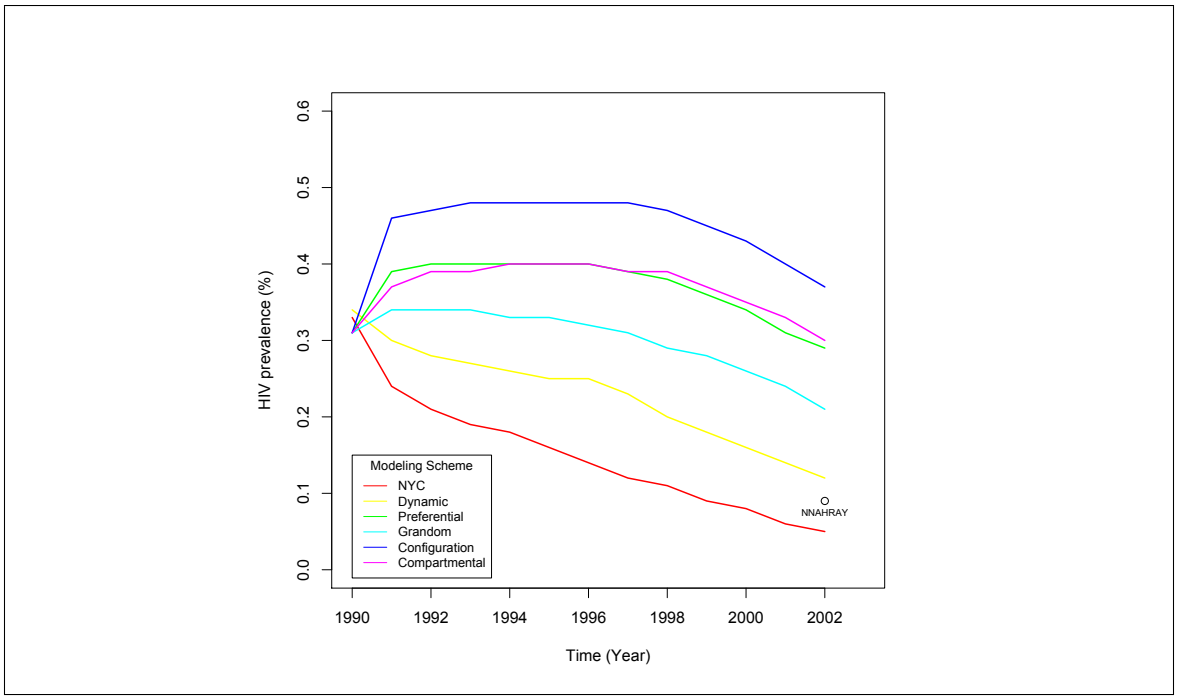

Figure 4.1.2: HIV prevalence of testing 2

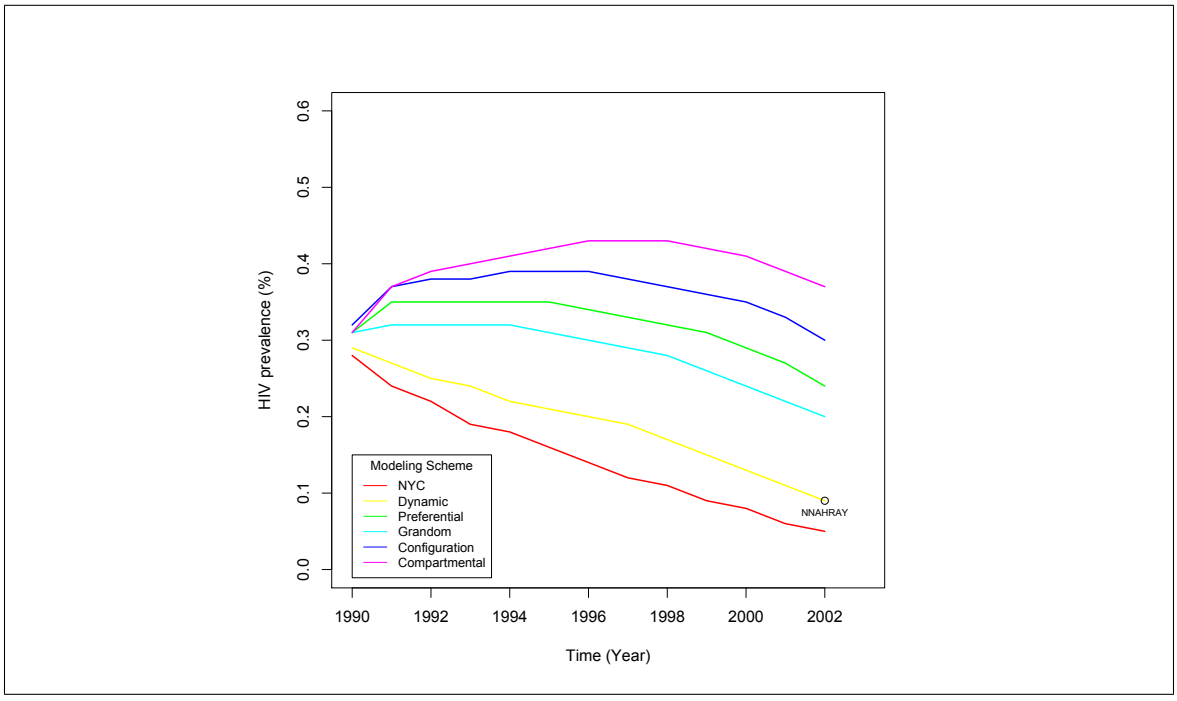

Figure 4.1.3: HIV prevalence of testing 3

from the network of the same size, the network structural results in testing 1 are the same as those in testing 3. Table 4.1.3 summarizes the network structure measurement results in all the model samples in testing 1 and testing 2.

Table 4.1.3: Network structures of all models

		size	edge	lgc-size	lgc-edge	lgc-density	arcremove	
	NNAHRAY	465	386	206	231	1.12	30.00	
Testing 1	Dynamic	465	386	203	230	1.13	27.62	
	Preferential	416	386	189	261	1.39	73.29	
	Grandom	235	386	120	360	3.04	241.53	
	Configuration	442	387	277	318	1.15	41.64	
	Compartmental	465	1012	460	1012	2.20	554.06	
		component-distmean	component-distsd	degree-max	degree-mean	degree-sd	orderdegree-diffmean	orderdegree-diffsd
	NNAHRAY	94.07	99.99	12.00	1.66	1.49	0.00	0.00
Testing 1	Dynamic	93.04	97.81	9.16	1.66	1.39	0.00	0.46
	Preferential	88.59	91.57	16.70	1.86	1.82	0.00	2.13
	Grandom	62.03	58.97	13.74	3.31	3.43	0.00	3.61
	Configuration	176.84	130.78	12.14	1.75	1.70	0.00	1.67
	Compartmental	454.02	40.67	412.65	4.35	26.12	-2.69	25.40
		size	edge	lgc-size	lgc-edge	lgc-density	arcremove	
	NNAHRAY	465	386	206	231	1.12	30.00	
Testing 2	Dynamic	591	451	278	289	1.04	11.81	
	Preferential	442	245	52	51	0.97	0.00	
	Grandom	478	138	23	22	0.95	0.25	
	Configuration	404	252	178	189	1.05	11.63	
	Compartmental	444	46917	444	46917	105.57	46474.17	
		component-distmean	component-distsd	degree-max	degree-mean	degree-sd	orderdegree-diffmean	orderdegree-diffsd
	NNAHRAY	94.07	99.99	12.00	1.66	1.49	0.00	0.00
Testing 2	Dynamic	140.27	135.08	8.70	1.52	1.22	0.21	1.99
	Preferential	12.71	17.97	11.67	1.11	1.24	0.57	1.54
	Grandom	3.63	5.51	5.22	0.58	0.83	-0.75	1.21
	Configuration	86.88	85.96	10.02	1.24	1.48	0.58	1.92
	Compartmental	444.08	0.00	342.12	211.13	52.83	-178.32	66.96

The results of testing 1 in Table 4.1.3 show that with a sampling method similar to the real sampling procedure used in NNAHRAY study, only Dynamic has network samples close to the NNAHRAY contact network. All the static network models are far from being close.

For testing 2, Table 4.1.3 shows that none of the network models can replicate the contact network in NNAHRAY closely, especially for the property of having a largest component which is denser than a tree. Note that Configuration and Dynamic are similar in their network measurements in testing 2. The similarity and the difference in sampling network structural measure results between these two models highlight the importance of network sampling method.

The above observations bring us to question the general network modeling practice of using population sample data to directly calibrate models. They also speak to the potential benefit of modeling the sampling method into network models.

4.1.7 Conclusion

The computational HIV prevalence prediction results of all models show that our dynamic network model predicts the spreading of HIV in a NNAHRAY- like population over 12 years more closely than any other model does in 3 kinds of comparison strategies. The results support the hypothesis that inclusion of arc dynamics and consideration of network structure is important to closely predict the prevalence of diseases that spread in a heterogeneous manner, such as HIV and STDs.

There are a few more network structural comparison results that are noteworthy:

- With a sampling method similar to the real sampling procedure used in NNAHRAY study, only Dynamic has network samples close to the NNAHRAY contact network. All the static network models are far from being close.
- Without any sampling method, none of the network models can replicate the contact network in NNAHRAY closely, especially for the property of having a largest component which is modestly denser than a tree.
- Without any sampling method, Configuration and Dynamic may have similar averaged network structural measurement results.

The results point out the importance of considering the network sampling process when adopting network models to represent populations. Our model's success in reproducing a contact network sample collected in the field by public health workers suggests that the practice of modeling the sampling method can be an important key.

4.2 Disease Prevalence and Network Structure

In the epidemiological literature, network models are commonly used to describe the population contact pattern, and are usually parameterized by fitting a small number of network metrics of the data. For example, [13] and [42] both fit their models to the degree distribution estimated from the data. The latter additionally fits the model to the number of cliques (a clique is a completely connected subgraph). As important as the degree distribution or the number of cliques may be, we hypothesize that a model may need to be fit to many network metrics in order to successfully reproduce a contact network on top of which the disease spreads similarly to the data. We test this hypothesis by simulating networks with a wide range of parameter values. If different parameter values can generate networks that are similar to the data in one metric and that have varying HIV spreads, then it will suggest that fitting to a single metric is insufficient to find a model that will closely match the disease reality.

There are three parameters, $(\alpha, P_{Add}, P_{Drop})$, in our model to be fitted. In Section 3.3, we fit our model so that all 13 network metrics of the network samples are close to those of the NNAHRAY contact network. To understand the necessity of fitting a large number of metrics at the same time and to examine the adequacy of the single metric fitting method, in this section, we fit our model with 8 different single metric fitting methods. The selected metrics are HIV prevalence and 7 network metrics, including Lgcsize, Lgcdensity, Arcremove, Componentdistmean, Ordereddegreediffmean, Degreesd, Degreemax. After comparing their 13 network metric values with those of the NNAHRAY contact network, we examine whether the best fitted results predict similar HIV prevalence within our model. For ease of terminology, from now on we refer to the single metric fitting method by the metric in use (e.g., Lgcsize) and our proposed 13-metric fit by Complete.

4.2.1 Computational Results and Discussion

We examine the same range of $(\alpha, P_{Add}, P_{Drop})$ values as that in Section 3.4.5, using the network metric results in Table 3.4.12 and Table 3.4.13 to look for the best fit results for all single metric methods.

In the disease simulation experiments, most model parameter values are the same as those used in Section 3.4.5, except for the risk behavior frequencies. In Section 3.4.5, the values of $F[SEX_{later}]$ and $F[IDU_{later}]$ are lower than those of $F[SEX_{before}]$ and $F[IDU_{before}]$ in order to reflect the emergence of intervention policy effects after 1996. To avoid confounding our results with intervention policies, in this section we set $F[IDU_{later}] := F[IDU_{before}]$ and $F[SEX_{later}] := F[SEX_{before}]$. For each set of specific $(\alpha, P_{Add}, P_{Drop})$ values, 10 networks are generated by the model. On top of them the spread of HIV and HSV-2 is simulated. Table 4.2.1 records their average HIV prevalence result.

For all single network metric methods, a contact network is considered close to the NNAHRAY contact network if the selected network metric falls within the specific range listed in Table 4.2.2. The ranges were chosen with varying degrees of deviation from the mean, from 40% to 1.7%. For the HIV prevalence metric, a contact network is considered close to NNAHRAY contact network if it has HIV prevalence within that of Complete, since the increases in $F[IDU_{later}]$ and $F[SEX_{later}]$ values lead to a higher HIV prevalence prevalence than in NNAHRAY.

Table 4.2.3 and Table 4.2.4 summarize the best fit results' 13 network metric statistics and HIV prevalence, averaged over the 10 instances, in each single network metric method. We do not find any single metric method whose 13 average network metrics are concurrently close to those of NNAHRAY contact network. Although HIV prevalence has the closest fit results to Complete, the former has larger Degreemax and Orderdegreediffsd than the latter. The HIV statistics in Table 4.2.4 show that although some single network metric methods have wider range of HIV prevalence

Table 4.2.1: HIV prevalence result

α	P_{Add}	P_{Drop}	HIV	α	P_{Add}	P_{Drop}	HIV
0	0.02	0.02	0.21	0.75	0.02	0.02	0.22
0	0.04	0.02	0.23	0.75	0.04	0.02	0.25
0	0.04	0.04	0.21	0.75	0.04	0.04	0.21
0	0.06	0.02	0.26	0.75	0.06	0.02	0.26
0	0.06	0.04	0.23	0.75	0.06	0.04	0.24
0	0.06	0.06	0.21	0.75	0.06	0.06	0.21
0	0.08	0.02	0.27	0.75	0.08	0.02	0.29
0	0.08	0.04	0.24	0.75	0.08	0.04	0.26
0	0.08	0.06	0.23	0.75	0.08	0.06	0.24
0	0.08	0.08	0.21	0.75	0.08	0.08	0.22
0.25	0.02	0.02	0.22	1	0.02	0.02	0.24
0.25	0.04	0.02	0.25	1	0.04	0.02	0.26
0.25	0.04	0.04	0.21	1	0.04	0.04	0.21
0.25	0.06	0.02	0.26	1	0.06	0.02	0.28
0.25	0.06	0.04	0.24	1	0.06	0.04	0.25
0.25	0.06	0.06	0.21	1	0.06	0.06	0.21
0.25	0.08	0.02	0.28	1	0.08	0.02	0.30
0.25	0.08	0.04	0.26	1	0.08	0.04	0.26
0.25	0.08	0.06	0.24	1	0.08	0.06	0.24
0.25	0.08	0.08	0.21	1	0.08	0.08	0.21
0.5	0.02	0.02	0.22	1.25	0.02	0.02	0.25
0.5	0.04	0.02	0.25	1.25	0.04	0.02	0.28
0.5	0.04	0.04	0.21	1.25	0.04	0.04	0.23
0.5	0.06	0.02	0.28	1.25	0.06	0.02	0.29
0.5	0.06	0.04	0.24	1.25	0.06	0.04	0.26
0.5	0.06	0.06	0.21	1.25	0.06	0.06	0.23
0.5	0.08	0.02	0.29	1.25	0.08	0.02	0.33
0.5	0.08	0.04	0.26	1.25	0.08	0.04	0.28
0.5	0.08	0.06	0.24	1.25	0.08	0.06	0.23
0.5	0.08	0.08	0.21	1.25	0.08	0.08	0.21

Table 4.2.2: Single metric close range

Network Metric	Specification	Value
$Lgcsize$	$206 \pm (14)$	(182, 230)
$Lgcdensity$	$1.12 \pm (0.02)$	(1.10, 1.14)
$Arcremove$	$30 \pm (8)$	(22,38)
$Componentdistmean$	$94.07 \pm (13)$	(81,107.47)
$Degreemax$	$12 \pm (2)$	(10,14)
$Desgreesd$	$1.49 \pm (0.6)$	(1.43,1.54)
$Orderdegreediffmean$	$0.00 \pm (0)$	(0.00, 0.00)

than the others, in general their prevalence results are close to that of Complete.

In order to see how consistently and closely the single network metric methods can

Table 4.2.3: Structure comparison - I

	α	P_{Add}	P_{Drop}	size	edge	lgc-size	lgc-edge	lgc-density	arcremove	component-distmean	component-distsd
NNAHRAY				465	386	206	231	1.12	30	94.07	99.99
Complete											
	0.5	0.06	0.06	465	386	218	239.8	1.10	22.76	107.47	105.40
Lgcsize											
max	1.25	0.080	0.080	465	386	227	248	1.17	36.06	116.71	109.83
min	0.00	0.020	0.020	457	386	194	218	1.09	21.19	85.92	93.04
average	0.54	0.051	0.051	464	386	212	237	1.12	25.17	102.43	102.53
Lgcdensity											
max	1.25	0.080	0.080	465	393	240	262	1.14	30.70	132.91	113.78
min	0.00	0.020	0.020	462	386	197	218	1.10	22.00	88.82	93.99
average	0.63	0.044	0.043	464	387	214	238	1.11	24.43	104.54	103.11
Componentdistmean											
max	1.25	0.080	0.080	465	386	218	247	1.17	36.06	107.47	105.40
min	0.00	0.020	0.020	457	386	194	218	1.10	22.00	85.92	93.04
average	0.64	0.047	0.047	463	386	208	234	1.12	26.38	98.62	100.50
Arcremove											
max	1.25	0.080	0.080	465	416	337	362	1.19	36.06	247.51	145.54
min	0.25	0.020	0.020	441	386	194	224	1.07	22.76	85.92	93.04
average	0.88	0.048	0.042	462	391	229	255	1.13	27.23	122.18	108.09
Orderdegreediffmean											
max	1.25	0.080	0.080	465	387	282	299	1.17	36.06	174.90	133.95
min	0.00	0.020	0.020	457	386	194	218	1.06	17.28	85.92	93.04
average	0.46	0.058	0.046	464	386	233	255	1.10	22.40	123.11	112.47
Degreesd											
max	0.75	0.080	0.060	465	387	282	299	1.13	26.42	174.90	133.95
min	0.00	0.020	0.020	464	386	203	229	1.06	17.89	93.61	98.02
average	0.30	0.058	0.044	465	386	237	257	1.09	21.01	126.46	114.05
Degreemax											
max	0.75	0.080	0.060	465	389	306	323	1.13	26.42	204.25	142.06
min	0.00	0.020	0.020	464	386	203	229	1.06	17.17	93.61	98.02
average	0.33	0.063	0.035	465	387	259	278	1.08	19.80	151.00	123.66
HIV Prevalence											
max	1.25	0.080	0.080	465	386	236	256	1.16	31.26	124.69	114.06
min	0.00	0.020	0.020	461	386	194	218	1.09	21.19	85.92	93.04
average	0.50	0.058	0.058	464	386	214	237	1.11	24.66	103.73	103.25

predict HIV prevalence compared to Complete, we run disease spread experiments with different magnitudes of two parameters, the unprotected sex frequency and the synergy of HSV-2 on HIV transmission. In half of the experiments, for the previous best fitted results of each method, HIV is transmitted with the frequency of unprotected sex modified to be 0.4, 0.3, 0.2, and 0.1 times/week. In the other half, HIV is transmitted when the synergy of HSV-2 on HIV's transmission, is discounted by scalars 100%, 50% and 10%. The prevalence comparison results are shown in Figure

Table 4.2.4: Structure comparison - II

	α	P_{Add}	P_{Drop}	degree-max	degree-mean	degree-sd	orderdegree-diffmean	orderdegree-diffsd	HIV
NNAHRAY				12	1.66	1.49	0.00	0.00	NA
Complete									
	0.5	0.06	0.06	10.96	1.66	1.43	0.00	0.46	0.21
Lgcsize									
max	1.25	0.080	0.080	67.12	1.69	3.65	0.00	3.17	0.23
min	0.00	0.020	0.020	7.66	1.66	1.33	0.00	0.44	0.21
average	0.54	0.051	0.051	20.37	1.66	1.77	0.00	0.90	0.21
Lgcdensity									
max	1.25	0.080	0.080	87.50	1.69	4.56	0.00	3.91	0.24
min	0.00	0.020	0.020	8.08	1.66	1.34	-0.03	0.44	0.21
average	0.63	0.044	0.043	27.90	1.67	2.08	0.00	1.18	0.22
Componentdistmean									
max	1.25	0.080	0.080	67.12	1.69	3.65	0.00	3.17	0.23
min	0.00	0.020	0.020	8.08	1.66	1.34	0.00	0.44	0.21
average	0.64	0.047	0.047	22.76	1.67	1.85	0.00	0.98	0.21
Arcremove									
max	1.25	0.080	0.080	184.42	1.82	8.93	0.00	8.73	0.30
min	0.25	0.020	0.020	9.97	1.66	1.41	-0.13	0.44	0.21
average	0.88	0.048	0.042	55.23	1.69	3.24	-0.02	2.46	0.23
Orderdegreediffmean									
max	1.25	0.080	0.080	67.12	1.69	3.65	0.00	3.17	0.26
min	0.00	0.020	0.020	7.66	1.66	1.33	0.00	0.44	0.21
average	0.46	0.058	0.046	17.89	1.66	1.72	0.00	0.77	0.23
Degreesd									
max	0.75	0.080	0.060	14.18	1.66	1.52	0.00	0.56	0.24
min	0.00	0.020	0.020	9.16	1.66	1.43	0.00	0.44	0.21
average	0.30	0.058	0.044	10.69	1.66	1.47	0.00	0.48	0.23
Degreemax									
max	0.75	0.080	0.060	13.09	1.67	1.75	0.00	0.57	0.28
min	0.00	0.020	0.020	10.23	1.66	1.43	-0.01	0.45	0.21
average	0.33	0.063	0.035	11.63	1.66	1.57	0.00	0.50	0.25
HIV prevalence									
max	1.25	0.080	0.080	58.76	1.67	3.07	0.00	2.41	0.21
min	0.00	0.020	0.020	7.66	1.66	1.33	0.00	0.44	0.21
average	0.50	0.058	0.058	16.77	1.66	1.62	0.00	0.75	0.21

4.2.1 and in Figure 4.2.2. Both graphs indicated that the single network metric fitting methods are consistently close to that of Complete in predicting HIV prevalence, noting that the HIV prevalence in the best fitted results of some single network fitting methods (e.g. Degreemax) gets further from that of Complete as the HIV prevalence increases.

Observing that the single metric fitting methods perform closely in average results but with considerable variance in HIV prevalence prediction, we hypothesize that the noise of HIV prevalence in our model can be eliminated by increasing the number of

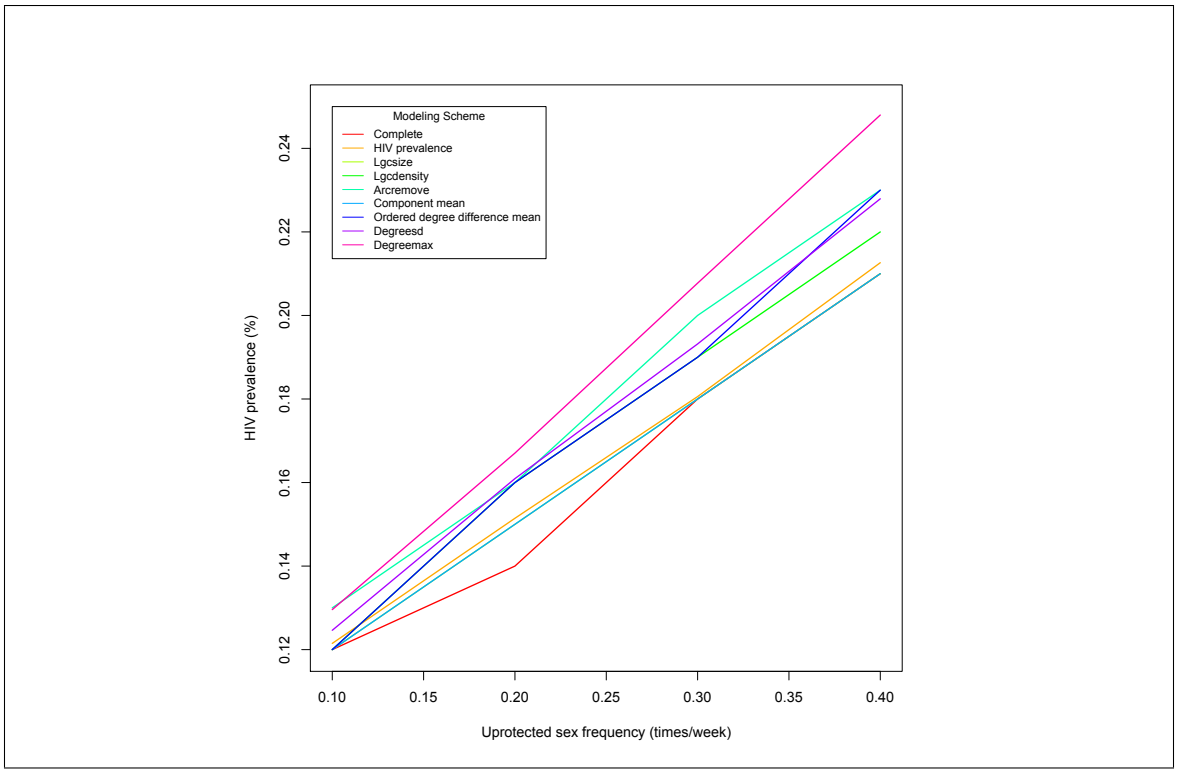

Figure 4.2.1: Unprotected sex frequency vs HIV prevalence

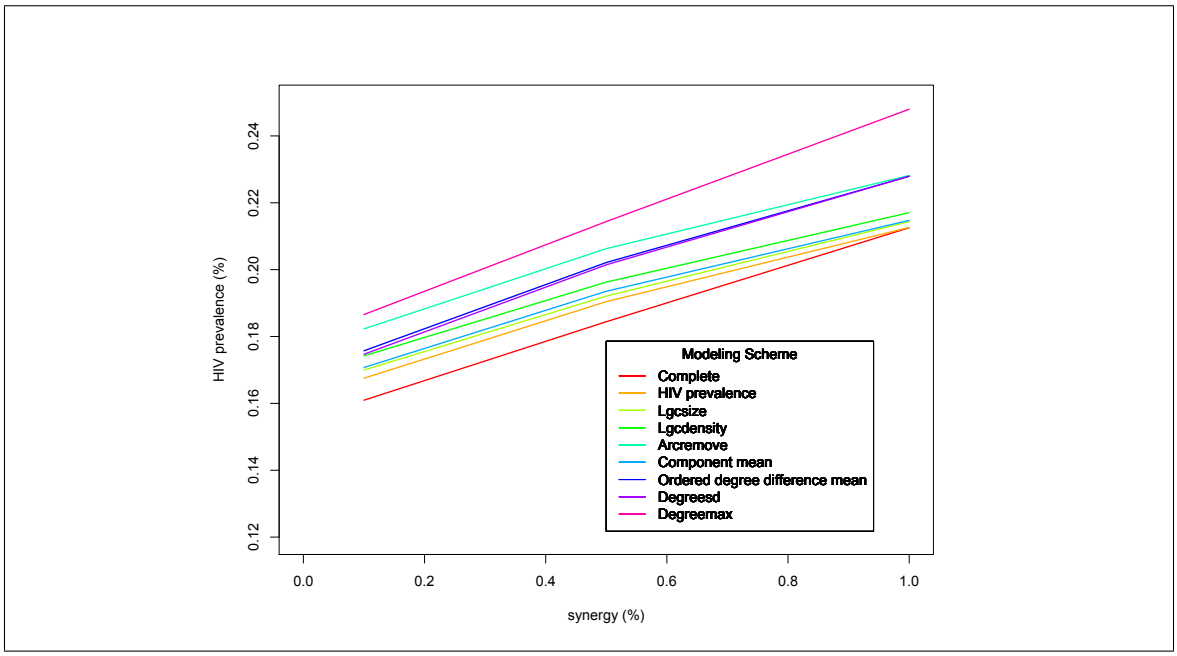

Figure 4.2.2: Synergy of HSV-2 vs HIV prevalence

network metrics to be fitted. Table 4.2.5 summarizes the number of fitted network metrics of the selected parameters among all single network metric methods. Figure

4.2.3 plots all the parameter sets in Table 4.2.5 based on the the number of fitted metrics and the HIV prevalence. The clear decreasing trend of HIV prevalence with increasing number of fitted metrics supports the hypothesis.

Table 4.2.5: The number of fitted network metrics for the selected parameters

α	P_{Add}	P_{Drop}	**# of fitted network metrics**	**HIV prevalence**	α	P_{Add}	P_{Drop}	**# of fitted network metrics**	**HIV prevalence**
0	0.2	0.2	3	0.21	0.5	0.2	0.2	6	0.22
0	0.4	0.2	2	0.23	0.5	0.4	0.4	6	0.21
0	0.4	0.4	4	0.21	0.5	0.6	0.4	2	0.24
0	0.6	0.2	2	0.26	0.5	0.6	0.6	7	0.21
0	0.6	0.4	2	0.23	0.5	0.8	0.4	1	0.26
0	0.6	0.6	2	0.21	0.5	0.8	0.6	3	0.24
0	0.8	0.2	1	0.27	0.5	0.8	0.8	1	0.21
0	0.8	0.4	2	0.24	0.75	0.2	0.2	4	0.22
0	0.8	0.6	1	0.23	0.75	0.4	0.4	6	0.21
0	0.8	0.8	2	0.21	0.75	0.6	0.4	1	0.24
0.25	0.2	0.2	7	0.22	0.75	0.6	0.6	7	0.21
0.25	0.4	0.2	2	0.25	0.75	0.8	0.6	1	0.24
0.25	0.4	0.4	5	0.21	1	0.2	0.2	2	0.24
0.25	0.6	0.2	2	0.26	1	0.4	0.4	4	0.21
0.25	0.6	0.4	2	0.24	1	0.6	0.4	1	0.25
0.25	0.6	0.6	2	0.21	1	0.6	0.6	4	0.21
0.25	0.8	0.2	1	0.28	1	0.8	0.2	1	0.3
0.25	0.8	0.4	2	0.26	1	0.8	0.6	1	0.24
0.25	0.8	0.6	2	0.24	1	0.8	0.8	5	0.21
0.25	0.8	0.8	2	0.21	1.25	0.2	0.2	1	0.25
0.5	0.2	0.2	6	0.22	1.25	0.6	0.6	4	0.23
0.5	0.4	0.2	1	0.25	1.25	0.8	0.6	2	0.23
0.5	0.4	0.4	6	0.21	1.25	0.8	0.8	5	0.21

4.2.2 Conclusion

In this section, we test various single metric fitting methods, and find that none can reproduce the NNAHRAY contact network closely. Although their structural fitting results are not satisfying, their HIV prevalence predictions are consistently close to that of our proposed 13 network metric fitting method, with considerably large variance in the results. Furthermore, we show that among the best fit parameters of all single network metric methods, the more network metrics that the parameters fit, the closer their HIV prevalence results are to the estimated prevalence in real life. This suggests that single-metric fitting methods should be used with caution,

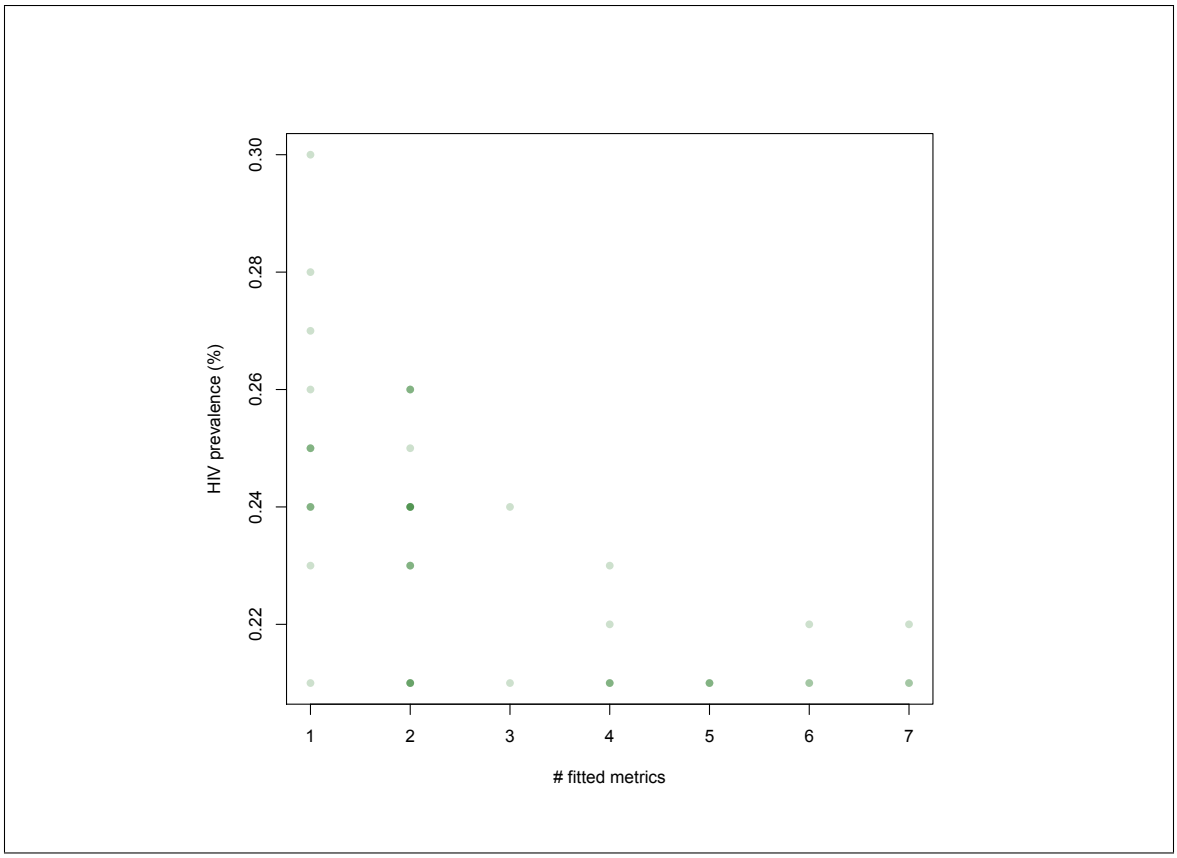

Figure 4.2.3: Number of fitted network metric on HIV prevalence

and that the effort of fitting more network metrics eliminates the unwanted disease prevalence prediction noise .

4.3 Model Application to Public Health Policy Analysis

HIV prevalence in the population can be reduced by decreasing risky contact frequency and host infectivity per contact, or by modifying the contact network structure. In this section we use our model to analyze the effect of the intervention measures, focusing more on the straightforward type.

First, we run computational experiments to test whether our model possesses the scale invariant property, which will enable us to apply the policy analysis result to a population of different size from our runs but with similar characteristics. Second, using our model calibrated to reproduce a population similar to Bushwick, we examine the effectiveness of HIV intervention policies implemented in different target groups.

4.3.1 Size Invariant Properties

In our model there are at least seven critical parameters to be considered for reducing HIV transmission. They are listed as follows:

1. HSV-2 synergistic impact on HIV
2. HIV transmission rate
3. HSV-2 transmission rate
4. Frequency of unprotected sex
5. Frequency of needle sharing
6. Frequency of adding a new partner
7. Probability that a new partner is the most popular, isolated, etc

In the following section we examine the impact in our model of changing these seven parameters on HIV prevalence for networks of various sizes.

4.3.2 Computational Results and Discussion

In the base case scenario, most model parameter values are the same as those used in Section 3.4.5, except for the risk behavior frequencies. Their values are selected in the same way as in Section 4.2 to exclude the effect of policy implementation after 1996. Both HIV and HSV-2 spread are simulated for 12 years on networks generated by our model with $(\alpha, P_{Add}, P_{Drop}) = (0.50, 0.06, 0.06)$, the Complete best fit result. In total 700 simulation experiments are conducted, with 100 experiments each on networks of 7 different sizes. The average number of HIV infections in the base case simulation experiments are recorded in Table 4.3.1.

700 simulation experiments, with one reduced value of each selected parameter, are conducted for the 7 parameters mentioned earlier. Among them, five parameters

have half of the base case values. The other two, the frequency of unprotected sex and needle sharing, have values cutting down to 0.1 times/week. The average relative rate of HIV infections compared to the base case for these simulation experiments are summarized in Table 4.3.1.

Table 4.3.1: HIV incidence

Network size: (total number of people)	1440	2880	4320	5760	7200	8640	10080	14400
Without intervention: (total number of infection)	18	36	53	71	87	105	122	179
	Relative rate (%)							
Intervention 1	85	82	84	84	84	83	83	83
Intervention 2	51	50	51	51	52	51	51	51
Intervention 3	94	92	94	92	93	94	94	93
Intervention 4	33	32	32	31	32	32	32	32
Intervention 5	92	91	92	90	92	92	91	91
Intervention 6	83	82	83	82	84	83	83	85
Intervention 7	100	97	99	97	100	99	100	99
	Relative rate summary for all network sizes(%)							
	Average	Max	Min					
Intervention 1	83	85	82					
Intervention 2	51	52	50					
Intervention 3	93	94	92					
Intervention 4	32	33	31					
Intervention 5	91	92	90					
Intervention 6	83	85	82					
Intervention 7	99	100	97					
Intervention 1	Cutting HSV-2 synergistic impact on HIV by half							
Intervention 2	Cutting HIV transmission rate by half							
Intervention 3	Cutting HSV-2 transmission rate by half							
Intervention 4	Decreasing frequency of unprotected sex from 0.4 to 0.1							
Intervention 5	Decreasing frequency of needle sharing from 0.25 to 0.1							
Intervention 6	Cutting the probability of adding a new partner in a month by half							
Intervention 7	Cutting α by half							

From the relative rate summary in Table 4.3.1, we find that the effect of cutting down the selected parameter value on reducing HIV incidence is essentially constant regardless of the network size. It is equivalent to say that our model, under this specific setting, possesses a scale invariant property for these critical parameters up to a network size of 14400 nodes.

4.3.3 STD/HIV Prevention Policy Analysis

We notice that STD/HIV prevention policies in the literature typically aim at reducing frequencies of transmission acts and probabilities of transmission through either pharmaceutical or non-pharmaceutical measures, and it prompts us to take a detailed look into the first five critical parameters. The synergistic parameter is tested in order to assess the indirect effectiveness of HSV-2 policies.

We investigate reducing the values of five parameters, including (i) HSV-2 synergistic impact on HIV, (ii) HIV transmission probability, (iii) HSV-2 transmission probability, (iv) the frequency of unprotected sex, and (v) the frequency of needle sharing. We test reductions in four groups: 100% of the population, a randomly-selected 50% of the population, IDU nodes, and lesbian, gay, and bisexual sex orientation(LGB) nodes. We run 100 simulation experiments for each of four magnitudes of parameter reduction in each group.

4.3.4 Computational Results and Discussion

We simulate the disease spread using the same parameter settings as in Section 4.2, to exclude the effect of policy implementation after 1996. Under this setting we simulate simultaneously the spread of HIV and HSV-2 from 1990 to 2002 in networks whose node and arc types are similar to those of the contact network in NNAHRAY. The target network size in these experiment is 1440.

Table 4.3.2 summarizes the average population composition at the end of the simulation (2002). The population is partitioned into groups based on sexual and injection drug behavior. On average 29% of nodes are Straight and IDU, 42% are Straight and NIDU, 10% are LGB and IDU, and 19% are LGB and NIDU.

Table 4.3.3 summarizes the prevalence results of both HIV and HSV-2 at the end of the simulation (2002) without reducing any of the critical parameters. LGB nodes and IDU nodes have similar, higher average HIV prevalence than Straight and

Table 4.3.2: Population composition

	Straight	LGB
IDU	29%	10%
NIDU	42%	19%

NIDU nodes, although the latter have the highest average HSV-2 prevalence. Without any intervention policies, applying our model to represent a population with similar composition to that in NNAHRAY results in an HIV prevalence prediction equal to 19% and HSV-2 prevalence prediction equal to 54%. Note that these are higher than the actual prevalence in NNAHRAY because, as we mentioned earlier, we run these tests ignoring the real-life-intervention-induced decreases, so they would not confound our results.

Table 4.3.3: Disease prevalence

	HIV Prevalence within group(%)	HSV2 Prevalence within group(%)
Straight and Non-IDU	18	57
IDU	21	49
LGB	22	56
Whole population	19	54

Table 4.3.4 records the HIV and HSV-2 prevalence results in 2002 for HIV transmission probability reductions. When the discount scalar is equal to 0.25, it means the HIV transmission probability is reduced to the original parameter value times 0.25. In the most extreme case in which transmission probability is reduced to zero for all nodes, the HIV prevalence is reduced from 19% to 4%. The prevalence is reduced to 11% if the transmission probability reduction takes place only in 50% randomly-chosen nodes. A similar or even lower HIV prevalence rate can be achieved by applying the probability reduction to only IDU nodes or LGB nodes, both fewer than 40% population. This unusual effectiveness of targeting at IDU or LGB nodes

indicates, in certain settings, how efficiently HIV spreads among (Straight and NIDU) nodes after relatively few contacts with the infected IDU or LGB nodes.

Table 4.3.4: HIV transmission probability

	Discount scalar	HIV prevalence (in whole population)	HSV2 prevalence (in whole population)
Without intervention	N.A.	0.19	0.54
Target group			
100% population	0	0.04	0.54
	0.25	0.08	0.54
	0.50	0.11	0.54
	0.75	0.15	0.53
50% population	0	0.11	0.54
	0.25	0.13	0.54
	0.50	0.15	0.53
	0.75	0.16	0.53
IDU	0	0.08	0.54
(39 % population)	0.25	0.10	0.54
	0.50	0.13	0.53
	0.75	0.16	0.53
LGB	0	0.11	0.54
(29% population)	0.25	0.13	0.54
	0.50	0.15	0.53
	0.75	0.17	0.53

Table 4.3.5 records the HIV and HSV-2 prevalence results in 2002 for HSV-2 synergistic impact on HIV reduction for each target group at two levels. The middle levels were not tested because there was so little difference between the extreme levels. In the most extreme case in no synergistic impact is considered in our model, the HIV prevalence rate is significantly less ($19\% - 13\% = 6\%$). Note that targeting at IDU nodes or LGB nodes are similarly effective as targeting randomly at 50% of the population.

Table 4.3.6 records the HIV and HSV-2 prevalence results in 2002 for HSV-2 transmission probability reduction. In the most extreme case in which HSV-2 transmission probability is reduced to zero for all nodes, HSV-2 prevalence rate drops from 54%

Table 4.3.5: HSV-2 synergistic impact on HIV

Target group	Discount scalar (%)	HIV prevalence rate (in whole population)	HSV2 prevalence rate (in whole population)
Without intervention	N.A.	0.19	0.54
Target group			
100% population	0	0.13	0.54
	25	-	-
	50	-	-
	75	0.15	0.54
50% population	0	0.16	0.54
	25	-	-
	50	-	-
	75	0.16	0.54
IDU (39 % population)	0	0.16	0.53
	25	-	-
	50	-	-
	75	0.17	0.54
LGB (29 % population)	0	0.16	0.53
	25	-	-
	50	-	-
	75	0.17	0.53

to 19%. If the reduction takes place only in 50% of randomly-chosen nodes, HSV-2 prevalence rate drops to 44%. As seen in the case with HIV transmission probability reduction, a lower HSV-2 prevalence rate can be achieved by applying the probability reduction to only IDU nodes or LGB nodes, both fewer than 40% of the population.

Earlier in Table 4.3.5 we see that HIV prevalence rate is sensitive to consideration of synergistic impact. Results in Table 4.3.6, on the other hand, show that HIV prevalence is less sensitive to the change of HSV-2 prevalence when the latter is above 33%. The results here shed light on the importance of considering the synergy of two diseases when at least one of them is moderately prevalent.

Table 4.3.7 records the HIV and HSV-2 prevalence results in 2002 for the reduction in the frequency of unprotected sex for each target group at different levels. In the most extreme case in which the frequency of unprotected sex is reduced to zero for all

Table 4.3.6: HSV-2 transmission probability

	Discount scalar	HIV prevalence (in whole population)	HSV2 prevalence (in whole population)
Without intervention	N.A.	0.19	0.54
Target group			
100% population	0	0.15	0.19
	25	0.17	0.33
	50	0.17	0.41
	75	0.18	0.48
50% population	0	0.19	0.44
	25	0.18	0.45
	50	0.18	0.48
	75	0.18	0.50
IDU	0	0.17	0.36
(39 % population)	25	0.17	0.44
	50	0.18	0.48
	75	0.18	0.51
LGB	0	0.18	0.39
(29 % population)	25	0.18	0.45
	50	0.18	0.49
	75	0.18	0.51

nodes, HSV-2 prevalence rate drops from 54% to 19% and HIV prevalence rate drops from 19% to 4%. Targeting nodes with IDU type or with LGB type achieves much better prevalence results in both diseases than not targeting does. And we see once more that by targeting IDU or LGB nodes, both HIV and HSV-2 prevalence rates drop significantly.

Table 4.3.8 records the HIV and HSV-2 prevalence results in 2002 for the reduction in the frequency of needle sharing for nodes with IDU type at two levels. The middle levels were not tested because there was so little difference between the extreme levels. In the most extreme case in which the frequency of needle sharing is reduced to zero for every IDU node, HIV prevalence rate drops from 19% to 16%. It implies clearly that the injection subnetwork in our model network is not the main driver of HIV. The sexual subnetwork within our model is the key focus to decrease the incidence of

Table 4.3.7: The frequency of unprotected sex

	Discount scalar	HIV prevalence (in whole population)	HSV-2 prevalence (in whole population)
Without intervention	N.A.	0.19	0.54
Target group			
100% population	0	0.04	0.19
	25	0.07	0.34
	50	0.11	0.42
	75	0.14	0.48
50% population	0	0.11	0.42
	25	0.13	0.45
	50	0.14	0.48
	75	0.16	0.51
IDU	0	0.08	0.32
(39 % population)	25	0.11	0.42
	50	0.13	0.46
	75	0.16	0.50
LGB	0	0.09	0.33
(29 % population)	25	0.12	0.43
	50	0.14	0.47
	75	0.16	0.51

both diseases.

Table 4.3.8: The frequency of needle sharing

	Discount scalar	HIV prevalence (in whole population)	HSV-2 prevalence (in whole population)
Without intervention	N.A.	0.19	0.54
Target group			
IDU	0	0.16	0.54
(39 % population)	25	-	-
	50	-	-
	75	0.18	0.54

4.3.5 Conclusion

The scale invariant property is a valuable property of our model. It can help us to save computational time and generalize our model's findings from a small population

to a large population with similar attributes.

The results in this section surprisingly show that an overall HIV epidemic can be initiated by a relatively few number of contacts between the risky individuals (LGB or IDU nodes) and the less risky individuals (Straight and NIDU nodes), suggesting that in general intervention efforts to break down mixing between the two groups might be rewarding. The results also show that, in the presence of a moderate HSV-2 prevalence, the HIV prevalence within the less risky group can match up with that within the risky group if no intervention efforts are in place.

Finally, although IDU-node interventions can have disproportionally positive impact on HIV prevalence, the results suggest that it is the IDU nodes' sexual behavior, more than their IDU behavior, that drives the improvement. Thus, needle-sharing programs might be less impactful than, for example, condom distribution among the IDU population.

CHAPTER V

THESIS CONTRIBUTION AND FUTURE RESEARCH DIRECTIONS

In this thesis, we show that it is important to consider the human contact dynamics when a network model is used to represent a human contact network and to predict the disease spread within a population. We propose a new dynamic network model, a new model fitting framework with emphasis on both the network structures and the network sampling process in real life, and a disease spread model based on the former. Most of the proposed models' (the network model's and the disease spread model's) parameters can be directly estimated from the epidemiological data; the rest of them can be fitted from the data while approximating a large number of metrics, whose interactions are too complicated to manipulate. We initially find that neither the the the well-known static network models nor our dynamic network can closely fit the reported data from the target contact network observed in real life. After modeling the network sampling process, our model does successfully reproduce the target network, but none of the static network models do. Furthermore, compared with disease spread models on the static network models, our disease spread model outperforms them in fitting 12 year HIV prevalence estimates. The network structure and disease prevalence comparison results strongly support two hypotheses: (1) contact dynamics play an important role in forming the human contact network structures, and (2) consideration of contact dynamics is important to obtain accurate disease prevalence prediction.

We also show that the HIV prevalence in our model is likely a mild overestimate if

it is obtained from using the traditional single metric fitting method, but the imprecision can be corrected by increasing the required number of fitted network metrics in the model fitting process. In the results obtained from fitting our model with one of different metrics, on average, none can reproduce networks with 13 network metrics concurrently close to those of the target contact network, and most of them produce slightly higher HIV prevalence than the prevalence estimate in real life. In addition, the relationship between the number of fitted network metrics and the HIV prevalence in the fitted results shows that the larger the former is, the closer the latter is to the prevalence estimate in real life.

Before we use our model to inform HIV intervention policy making, a closer examination of our model finds that it possesses a scale invariant property for policy-relevant parameters. The property allows us to apply the policy analysis result to a population of different size from our simulation experiments but with similar characteristics. After analyzing different kinds of intervention approaches in our model, we conclude that, in the specific population, (1) HSV-2 intervention policies need to be implemented long or effectively enough to bring down the HSV-2 prevalence below 33% before achieving collateral and significant HIV prevalence reduction in the population, (2) HIV intervention policies targeting at IDU or LGB groups are much more effective than those implemented among randomly-chosen population, due to the risky groups' unique role in introducing the HIV virus to the non-risky group at the early stage of epidemic, (3) HIV intervention policies targeting at modifying IDU group's injection behaviors are less effective than those aiming at changing the group's sexual behaviors.

One significant limitation of our work, and opportunity for future analysis, is the unavailability of multiple data sets with sufficient information to understand the population dynamics. With our work on a dynamic model fitted to NNAHRAY data demonstrating potential value beyond that of static network models, it is our hope

that future data sets will become available and will be used to further inform and refine our dynamic network modeling framework.

APPENDIX A

COMPUTATIONAL MODELS

Data: Set of arc types in our model: AT
Set of node types in our model: NT
Set of disease types in our model: D
Set of genders for nodes: G
Set of disease A's infected stages in our model: S_A
Probability of a node being type A: $pnt[A]$
Probability of a type A node with a disease B: $pnd[A][B]$
Probability of an infected node of gender A with disease B to transmit: $p[A][B][S_B]$
Scaling effect of disease A on disease B's transmission: $s[S_A][S_B]$
Duration of disease A's stages: $d[A][S_A]$
Number of weeks to run our model: WEEK

Input: Degree of connectivity: α
Number of new nodes added at the same time: New
Time interval between adding new nodes: T_{Add}
Time interval between spreading diseases: T_{Spread}
Frequency of a type A arc: $F[A]$

Figure A.1: Preferential model - Part I

```
Function
preferential_model(AT, NT, D, G, S[ ], pnt[ ], pnd[ ], p[ ][ ], s[ ][ ], d[ ][ ])
begin
    add_counts = 0 ; change_counts = 0 ; spread_counts = 0 ;
    model_time = 0 ;
    N = 1; A = ∅ ;
    while add_counts ≤ (WEEK/T_Add) do
        for i = 1 to NEW do
            Choose node (|N| + i)'s type, node_type, ∈ NT using pnt[ ] as a
            probability mass function ;
            for d = 1 to |D| do
                With probability pnd[node_type][d]
                Choose (|N| + i)'s disease d's status from S_d uniformly at
                random ;
            end
        end
        preferential_attachment (N, A, New, CT, NT, pnc[ ][ ], pnn[ ][ ]) ;
        add_counts = add_counts + 1 ;
    end
    while model_time ≤ WEEK do
        if model_time = spread_counts × T_Spread then
            spread_disease (A, N, G, D, S[ ], p[ ][ ], s[ ][ ]) ;
            spread_counts = spread_counts + 1 ;
        end
        for i = 1 to |N| do
            for d_1 = 1 to |D| do
                if model_time = node i's disease d_1's status advancement time
                then
                    stage = node i's disease d_1's status ;
                    Advance node i's disease d_1's status ;
                    Change node i's disease d_1's status advancement time
                    based on d[d_1][stage] ;
                end
            end
        end
        model_time = model_time + 1 ;
    end
end
```

Figure A.2: Preferential model - Part II

Table A.1: Preferential parameter value - I

Parameter definition	Parameter notation	Value	Value source
Dynamic Network Model			
Set of contact types	AT	3 (SEX, IDU, SEX and IDU)	NNAHRAY [19]
Set of node types	NT	12 ($FBI, FHI, FSI, \ldots, MSN$)	NNAHRAY [19]
Set of diseases	D	2 (HIV and HSV-2)	NNAHRAY [19]
Age	Age	Uniformly distributed between 18 and 66	NNAHRAY [19]
Gender	$Gender$	57% male and 43% female	NNAHRAY [19]
Group type proportion within women:			
FBN	$pnt[FBN]$	22%	NNAHRAY [19]
FHN	$pnt[FHN]$	1%	NNAHRAY [19]
FSN	$pnt[FSN]$	43%	NNAHRAY [19]
FBI	$pnt[FBI]$	11%	NNAHRAY [19]
FHI	$pnt[FHI]$	0%	NNAHRAY [19]
FSI	$pnt[FSI]$	22%	NNAHRAY [19]
Group type proportion within men:			
MBN	$pnt[MBN]$	12%	NNAHRAY [19]
MHN	$pnt[MHN]$	3%	NNAHRAY [19]
MSN	$pnt[MSN]$	36%	NNAHRAY [19]
MBI	$pnt[MBI]$	9%	NNAHRAY [19]
MHI	$pnt[MHI]$	3%	NNAHRAY [19]
MSI	$pnt[MSI]$	37%	NNAHRAY [19]
Probability of a type A node having a type B contact	$pnc[A][B]$	See Table 3.2.2	NNAHRAY [19]
Probability of a type A node choosing a type B node	$pnn[A][B]$	See Table 3.2.3	NNAHRAY [19]
Number of weeks to run our model	$WEEK$	$(12 \times 52 = 624)$ weeks	Model setup
Time interval between adding new nodes	T_{Add}	4 weeks	Model setup
Time interval between spreading diseases	T_{Spread}	1 week	Model setup

Table A.2: Preferential parameter value - II

Parameter definition	Parameter notation	Value	Value source
HIV and HSV-2			
HIV prevalence rate of newcomers from 1990 to 2002:			
Set of HIV's infected stages in our model	S_{HIV}	4 (Primary, Asymptomatic, Symptomatic, AIDS)	[18, 26]
IDU:	$pnd[IDU][HIV]$	50%	[18, 26]
MSM:	$pnd[MSM][HIV]$	47%	[18, 26]
General population:	$pnd[General][HIV]$	9 %	[18, 26]
Probability of an infected node of gender A with HIV to transmit	$p[A][HIV][S_{HIV}]$	See Table 3.4.1	[36, 46]
Duration of HIVs stages	$d[HIV][S_{HIV}]$	See Table 3.4.1	[36, 46]
HSV-2 prevalence rate from 1990 to 2002:			
Set of HSV-2's infected stages in our model	S_{HSV-2}	5 (Primary, Early latent, Latent, Late latent, Recurrent Ulcers)	[36, 46]
General population:	$pnd[General][HSV_2]$	21% before 1996 and 17% after 1996	[48]
Probability of an infected node of gender A with HSV-2 to transmit	$p[A][HSV-2][S_{HSV-2}]$	See Table 3.4.2	[36, 46]
Duration of HIVs stages	$d[HSV-2][S_{HSV-2}]$	See Table 3.4.2	[36, 46]
Scaling effect of HSV-2 on HIV's transmission in Testing 1 and 2	$s[S_{HSV-2}][S_{HIV}]$	See Table 3.4.2	[36, 46]
Scaling effect of HSV-2 on HIV's transmission in Testing 3	0		
Number of new nodes added at the same time in Testing 1	New	10	Model setup
Number of new nodes added at the same time in Testing 2	New	4	Model setup

Data: Set of nodes in the current network: N
Set of arcs in the current network: A
Set of arc types in our model: AT
Probability of a type A node having a type B arc: $pnc[A][B]$
Probability of a type A node choosing a type B node: $pnn[A][B]$
Degree distribution: w

Function Grandom_AddArc($N, A, AT, w, pnc[\,][\,], pnn[\,][\,]$)
begin
  Generate $|N|$ node degree for each node using w as a probability mass function and store them in $n[\,]$;
  while $((max\{n[\,]\})^2 \geq \sum_{k \in N} n[k])$ **do**
    Generate $|N|$ node degree for each node using w as a probability mass function and store them in $n[\,]$;
  end
  while $((\sum_{k \in N} n[k])/2 \neq 0)$ **do**
    Generate $|N|$ node degree for each node using w as a probability mass function and store them in $n[\,]$;
  end
  Store the node type of each node in nt[ ] ;
  $arc\_type = SEX$;
  for $i = 1$ ***to*** $|N|$ **do**
    if $(nt[i] = IDU)$ **then**
      Choose an $arc\_type \in AT$ using $pnc[nt[i]][\,]$ as a probability mass function ;
    end
    for $j = 1$ ***to*** $|N|$ **do**
      $grandom\_weight[j] =$
      $(n[i] \times n[j]) / \sum_{k \in N} n[k] \times pnn[nt[i]][nt[j]] \times pnc[nt[j]][arc\_type]$;
      if $j = i$ **then**
        $grandom\_weight[j]$=0 ;
      end
    end
    for $j = 1$ ***to*** i **do**
      With probability $grandom\_weight[j]$ do
      Add type arc_type arc (i, j) to A ;
    end
  end
end

Figure A.3: Grandom_AddArc algorithm

Data: Set of arc types in our model: AT
Set of node types in our model: NT
Set of disease types in our model: D
Set of genders for nodes: G
Set of disease A's infected stages in our model: S_A
Probability of a node being type A: $pnt[A]$
Probability of a type A node with a disease B: $pnd[A][B]$
Probability of an infected node of gender A with disease B to transmit: $p[A][B][S_B]$
Scaling effect of disease A on disease B's transmission: $s[S_A][S_B]$
Duration of disease A's stages: $d[A][S_A]$
Number of weeks to run our model: WEEK
Degree distribution: w

Input: Time interval between adding new nodes: T_{Add}
Time interval between spreading diseases: T_{Spread}
Frequency of a type A arc: $F[A]$

Function
Grandom_model($AT, NT, D, G, S[\,], pnt[\,], pnd[\,], p[\,][\,], s[\,][\,], d[\,][\,]$)
begin
  $add\_counts = 0$; $change\_counts = 0$; $spread\_counts = 0$;
  $model\_time = 0$;
  $N = 1; A = \emptyset$;
  while $add\_counts \leq (WEEK/T_{Add})$ **do**
    for $i = 1$ ***to*** NEW **do**
      Choose node $(|N| + i)$'s type, $node\_type$, from NT using $pnt[\,]$ as a probability mass function ;
      for $d = 1$ ***to*** $|D|$ **do**
        With probability $pnd[node\_type][d]$ do
        Choose $(|N| + i)$'s disease d's status from S_d uniformly at random ;
      end
    end
    $add\_counts = add\_counts + 1$;
  end
  Grandom_AddArc ($N, A, AT, w, pnc[\,][\,], pnn[\,][\,]$) ;
end

Figure A.4: Grandom model - Part I

```
begin
    while model_time ≤ WEEK do
        if model_time = spread_counts × T_Spread then
            spread_disease (A, N, G, D, S[ ], p[ ][ ], s[ ][ ]) ;
            spread_counts = spread_counts + 1 ;
        end
        for i = 1 to |N| do
            for d_1 = 1 to |D| do
                if model_time = node i's disease d_1's status advancement time
                then
                    stage = node i's disease d_1's status ;
                    Advance node i's disease d_1's status ;
                    Change node i's disease d_1's status advancement time
                    based on d[d_1][stage] ;
                end
            end
        end
        model_time = model_time + 1 ;
    end
end
```

Figure A.5: Grandom model - Part II

Table A.3: Grandom parameter value -I

Definition of parameter	Parameter Notation	Value	Value source
Dynamic Network Model			
Set of contact types	AT	3 (SEX, IDU, SEX and IDU)	NNAHRAY [19]
Set of node types	NT	12 ($FBI, FHI, FSI, \ldots, MSN$)	NNAHRAY [19]
Set of diseases	D	2 (HIV and HSV-2)	NNAHRAY [19]
Age	Age	Uniformly distributed between 18 and 66	NNAHRAY [19]
Gender	$Gender$	57% male and 43% female	NNAHRAY [19]
Group type proportion within women:			
FBN	$pnt[FBN]$	22%	NNAHRAY [19]
FHN	$pnt[FHN]$	1%	NNAHRAY [19]
FSN	$pnt[FSN]$	43%	NNAHRAY [19]
FBI	$pnt[FBI]$	11%	NNAHRAY [19]
FHI	$pnt[FHI]$	0%	NNAHRAY [19]
FSI	$pnt[FSI]$	22%	NNAHRAY [19]
Group type proportion within men:			
MBN	$pnt[MBN]$	12%	NNAHRAY [19]
MHN	$pnt[MHN]$	3%	NNAHRAY [19]
MSN	$pnt[MSN]$	36%	NNAHRAY [19]
MBI	$pnt[MBI]$	9%	NNAHRAY [19]
MHI	$pnt[MHI]$	3%	NNAHRAY [19]
MSI	$pnt[MSI]$	37%	NNAHRAY [19]
Probability of a type A node having a type B contact	$pnc[A][B]$	See Table 3.2.2	NNAHRAY [19]
Probability of a type A node choosing a type B node	$pnn[A][B]$	See Table 3.2.3	NNAHRAY [19]
Probability of a type A node choosing a type B node	$pnn[A][B]$	See Table 3.2.3	NNAHRAY [19]
The degree distribution	w		NNAHRAY [19]
Number of weeks to run our model	$WEEK$	($12 \times 52 = 624$) weeks	Model setup
Time interval between adding new nodes	T_{Add}	4 weeks	Model setup
Time interval between spreading diseases	T_{Spread}	1 week	Model setup

Table A.4: Grandom parameter value -II

Definition of parameter	Parameter Notation	Value	Value source
HIV and HSV-2			
HIV prevalence rate of newcomers from 1990 to 2002:			
Set of HIV's infected stages in our model	S_{HIV}	4 (Primary, Asymptomatic, Symptomatic, AIDS)	[18, 26]
IDU:	$pnd[IDU][HIV]$	50%	[18, 26]
MSM:	$pnd[MSM][HIV]$	47%	[18, 26]
General population:	$pnd[General][HIV]$	9 %	[18, 26]
Probability of an infected node of gender A with HIV to transmit	$p[A][HIV][S_{HIV}]$	See Table 3.4.1	[36, 46]
Duration of HIVs stages	$d[HIV][S_{HIV}]$	See Table 3.4.1	[36, 46]
HSV-2 prevalence rate from 1990 to 2002:			
Set of HSV-2's infected stages in our model	S_{HSV-2}	5 (Primary, Early latent, Latent, Late latent, Recurrent Ulcers)	[36, 46]
General population:	$pnd[General][HSV_2]$	21% before 1996 and 17% after 1996	[48]
Probability of an infected node of gender A with HSV-2 to transmit	$p[A][HSV-2][S_{HSV-2}]$	See Table 3.4.2	[36, 46]
Duration of HIVs stages	$d[HSV-2][S_{HSV-2}]$	See Table 3.4.2	[36, 46]
Scaling effect of HSV-2 on HIV's transmission in Testing 1 and 2	$s[S_{HSV-2}][S_{HIV}]$	See Table 3.4.2	[36, 46]
Scaling effect of HSV-2 on HIV's transmission in Testing 3	0		
Number of new nodes added at the same time in Testing 1	New	10	Model setup
Number of new nodes added at the same time in Testing 2	New	4	Model setup

Data: Set of nodes in the current network: N
Set of arcs in the current network: A
Set of arc types in our model: AT
Probability of a type A node having a type B contact: $pnc[A][B]$
Probability of a type A node choosing a type B node: $pnn[A][B]$
Degree distribution: w

Function Configuration_AddArc($N, A, AT, w, pnc[\,][\,], pnn[\,][\,]$)
begin
  Generate $|N|$ node degree for each node using w as a probability mass function and store them in $n[\,]$ and $n_1[\,]$;
  while $((max\{n[\,]\})^2 \geq \sum_{k \in N} n[k])$ **do**
    Generate $|N|$ node degree for each node using w as a probability mass function and store them in $n[\,]$ and $n_1[\,]$;
  end
  while $((\sum_{k \in N} n[k])/2 \neq 0)$ **do**
    Generate $|N|$ node degree for each node using w as a probability mass function and store them in $n[\,]$ and $n_1[\,]$;
  end
  Store the node type of each node in nt[ ] ;
  $arc\_type = SEX$;
  for $i = 1$ ***to*** $|N|$ **do**
    $arc\_left = n_1[i]$;
    for $j = 1$ ***to*** arc_left **do**
      if $(nt[i] = IDU)$ **then**
        Choose an $arc\_type \in AT$ using $pnc[nt[i]][\,]$ as a probability mass function ;
      end
      for $k = i + 1$ ***to*** $|N|$ **do**
        $configuration\_weight[k] =$
        $n_1[k] / \sum_{k \in N} n_1[k] \times pnn[nt[i]][nt[k]] \times pnc[nt[k]][arc\_type]$;
      end
      Choose a node, k, from $\{i + 1, \ldots, N\}$ using $configuration\_weight[\,]$ as a probability mass function ;
      if $(i, k) \notin A$ **then**
        Add type art_type arc (i, k) to A ;
        $n_1[k] = n_1[k] - 1$;
      end
    end
  end
end

Figure A.6: Configuration_AddArc algorithm

Data: Set of arc types in our model: AT
Set of node types in our model: NT
Set of disease types in our model: D
Set of genders for nodes: G
Set of disease A's infected stages in our model: S_A
Probability of a node being type A: $pnt[A]$
Probability of a type A node with a disease B: $pnd[A][B]$
Probability of an infected node of gender A with disease B to transmit: $p[A][B][S_B]$
Scaling effect of disease A on disease B's transmission: $s[S_A][S_B]$
Duration of disease A's stages: $d[A][S_A]$
Number of weeks to run our model: WEEK
Degree distribution: w

Input: Time interval between adding new nodes: T_{Add}
Time interval between spreading diseases: T_{Spread}
Frequency of a type A arc: $F[A]$

Function
Configuration_model($AT, NT, D, G, S[\,], pnt[\,], pnd[\,], p[\,][\,], s[\,][\,], d[\,][\,]$)
begin
  $add\_counts = 0$; $change\_counts = 0$; $spread\_counts = 0$;
  $model\_time = 0$;
  $N = 1; A = \emptyset$;
  while $add\_counts \leq (WEEK/T_{Add})$ **do**
    for $i = 1$ ***to*** NEW **do**
      Choose node $(|N| + i)$'s type, $node\_type$, $\in NT$ using $pnt[\,]$ as a probability mass function ;
      for $d = 1$ ***to*** $|D|$ **do**
        With probability $pnd[node\_type][d]$ do
        Choose $(|N| + i)$'s disease d's status from S_d uniformly at random ;
      end
    end
    $add\_counts = add\_counts + 1$;
  end
  Configuration_AddArc $(N, A, AT, w, pnc[\,][\,], pnn[\,][\,])$;
end

Figure A.7: Configuration model - Part I

```
begin
    while model_time ≤ WEEK do
        if model_time = spread_counts × T_Spread then
            spread_disease (A, N, G, D, S[ ], p[ ][ ], s[ ][ ]) ;
            spread_counts = spread_counts + 1 ;
        end
        for i = 1 to |N| do
            for d_1 = 1 to |D| do
                if model_time = node i's disease d_1's status advancement time
                then
                    stage = node i's disease d_1's status ;
                    Advance node i's disease d_1's status ;
                    Change node i's disease d_1's status advancement time
                    based on d[d_1][stage] ;
                end
            end
        end
        model_time = model_time + 1 ;
    end
end
```

Figure A.8: Configuration model - Part II

Table A.5: Configuration parameter value - I

Definition of parameter	Parameter notation	Value	Value source
Dynamic Network Model			
Set of contact types	AT	3 (SEX, IDU, SEX and IDU)	NNAHRAY [19]
Set of node types	NT	12 ($FBI, FHI, FSI, \ldots, MSN$)	NNAHRAY [19]
Set of diseases	D	2 (HIV and HSV-2)	NNAHRAY [19]
Age	Age	Uniformly distributed between 18 and 66	NNAHRAY [19]
Gender	$Gender$	57% male and 43% female	NNAHRAY [19]
Group type proportion within women:			
FBN	$pnt[FBN]$	22%	NNAHRAY [19]
FHN	$pnt[FHN]$	1%	NNAHRAY [19]
FSN	$pnt[FSN]$	43%	NNAHRAY [19]
FBI	$pnt[FBI]$	11%	NNAHRAY [19]
FHI	$pnt[FHI]$	0%	NNAHRAY [19]
FSI	$pnt[FSI]$	22%	NNAHRAY [19]
Group type proportion within men:			
MBN	$pnt[MBN]$	12%	NNAHRAY [19]
MHN	$pnt[MHN]$	3%	NNAHRAY [19]
MSN	$pnt[MSN]$	36%	NNAHRAY [19]
MBI	$pnt[MBI]$	9%	NNAHRAY [19]
MHI	$pnt[MHI]$	3%	NNAHRAY [19]
MSI	$pnt[MSI]$	37%	NNAHRAY [19]
Probability of a type A node having a type B contact	$pnc[A][B]$	See Table 3.2.2	NNAHRAY [19]
Probability of a type A node choosing a type B node	$pnn[A][B]$	See Table 3.2.3	NNAHRAY [19]
Probability of a type A node choosing a type B node	$pnn[A][B]$	See Table 3.2.3	NNAHRAY [19]
The degree distribution	w		NNAHRAY [19]
Number of weeks to run our model	$WEEK$	$(12 \times 52 = 624)$ weeks	Model setup
Time interval between adding new nodes	T_{Add}	4 weeks	Model setup
Time interval between spreading diseases	T_{Spread}	1 week	Model setup

Table A.6: Configuration parameter value - II

Definition of parameter	Parameter notation	Value	Value source
HIV and HSV-2			
HIV prevalence rate of newcomers from 1990 to 2002:			
Set of HIV's infected stages in our model	S_{HIV}	4 (Primary, Asymptomatic, Symptomatic, AIDS)	[18, 26]
IDU:	$pnd[IDU][HIV]$	50%	[18, 26]
MSM:	$pnd[MSM][HIV]$	47%	[18, 26]
General population:	$pnd[General][HIV]$	9 %	[18, 26]
Probability of an infected node of gender A with HIV to transmit	$p[A][HIV][S_{HIV}]$	See Table 3.4.1	[36, 46]
Duration of HIVs stages	$d[HIV][S_{HIV}]$	See Table 3.4.1	[36, 46]
HSV-2 prevalence rate from 1990 to 2002:			
Set of HSV-2's infected stages in our model	S_{HSV-2}	5 (Primary, Early latent, Latent, Late latent, Recurrent Ulcers)	[36, 46]
General population:	$pnd[General][HSV_2]$	21% before 1996 and 17% after 1996	[48]
Probability of an infected node of gender A with HSV-2 to transmit	$p[A][HSV-2][S_{HSV-2}]$	See Table 3.4.2	[36, 46]
Duration of HIVs stages	$d[HSV-2][S_{HSV-2}]$	See Table 3.4.2	[36, 46]
Scaling effect of HSV-2 on HIV's transmission in Testing 1 and 2	$s[S_{HSV-2}][S_{HIV}]$	See Table 3.4.2	[36, 46]
Scaling effect of HSV-2 on HIV's transmission in Testing 3	0		
Number of new nodes added at the same time in Testing 1	New	10	Model setup
Number of new nodes added at the same time in Testing 2	New	4	Model setup

Data: Set of nodes in the current network: N
Set of arcs in the current network: A
Set of arc types in our model: AT
Probability of a type A node having a type B contact: $pnc[A][B]$
Probability of a type A node choosing a type B node: $pnn[A][B]$

Function Compartmental_AddArc($N, A, AT, pnc[\,][\,], pnn[\,][\,]$)
begin
  $arc\_type = SEX$;
  for $i = 1$ ***to*** $|N|$ **do**
    for $j = 1$ ***to*** i **do**
      if $(nt[i] = IDU)$ **then**
        Choose an $arc\_type \in AT$ using $pnc[nt[i]][\,]$ as a probability mass function ;
      end
      $compartmental\_weight = pnn[nt[i]][nt[j]] \times pnc[nt[j]][arc\_type]$;
      if $compartmental\_weight \neq 0$ **then**
        Add type art_type arc (i, j) to A ;
      end
    end
  end
end

Figure A.9: Compartmental_AddArc algorithm

Data: Set of arcs in current network: A
Set of nodes in current network: N
Set of genders for nodes: G
Set of diseases spreading in our model : D
Set of disease A's infected stages in our model: S_A
Probability of an infected node of gender A with disease B to transmit: $p[A][B][S_B]$
Scaling effect of disease A on disease B's transmission: $s[S_A][S_B]$
Spreading scaling parameter q

Input: Frequency of a type A arc: $F[A]$

Function spread_disease($A, N, G, D, S[\,], p[\,][\,], s[\,][\,], p$)
begin
  Store the stage of each disease of each node in $stage[\,][\,]$;
  Store the gender of each node in $gender[\,]$;
  With probability q do
  for $a = 1$ ***to*** $|A|$ **do**
    for $d_1 = 1$ ***to*** $|D|$ **do**
      if *arc a links an infected node* n_1 *with disease* d_1 *and an uninfected node* n_2 **then**
        arc_type stores a's type ;
        $infection\_probability = p[gender[n_1]][stage[n_1][d_1]]$;
        for $d_2 = 1$ ***to*** $|D|$ **do**
          $infection\_probability = infection\_probability \times s[stage[n_1][d_2]][stage[n_1][d_1]] \times s[stage[n_2][d_2]][stage[n_2][d_1]]$
        end
        With probability $infection\_probability)^{F[arc\_type]}$ do
        n_2 becomes infected with disease d_1 ;
      end
    end
  end
end

Figure A.10: Compartmental_SpreadDisease algorithm

Data: Set of arc types in our model: AT
Set of node types in our model: NT
Set of disease types in our model: D
Set of genders for nodes: G
Set of disease A's infected stages in our model: S_A
Probability of a node being type A: $pnt[A]$
Probability of a type A node with a disease B: $pnd[A][B]$
Probability of an infected node of gender A with disease B to transmit: $p[A][B][S_B]$
Scaling effect of disease A on disease B's transmission: $s[S_A][S_B]$
Duration of disease A's stages: $d[A][S_A]$
Number of weeks to run our model: WEEK
Degree distribution: w
Spreading scaling parameter q

Input: Time interval between adding new nodes: T_{Add}
Time interval between spreading diseases: T_{Spread}
Frequency of a type A arc: $F[A]$

Function
Compartmental_model($AT, NT, D, G, S[\,], pnt[\,], pnd[\,], p[\,][\,], s[\,][\,], d[\,][\,]$)
begin
  $add\_counts = 0$; $change\_counts = 0$; $spread\_counts = 0$;
  $model\_time = 0$;
  $N = 1; A = \emptyset$;
  while $add\_counts \leq (WEEK/T_{Add})$ **do**
    for $i = 1$ ***to*** NEW **do**
      Choose node ($|N| + i$)'s type, $node\_type$, $\in NT$ using $pnt[\,]$ as a probability mass function ;
      for $d = 1$ ***to*** $|D|$ **do**
        With probability $pnd[node\_type][d]$ do
        Choose ($|N| + i$)'s disease d's status from S_d uniformly at random ;
      end
    end
    $add\_counts = add\_counts + 1$;
  end
  Compartmental_AddArc ($N, A, AT, pnc[\,][\,], pnn[\,][\,]$) ;
end

Figure A.11: Compartmental model - Part I

```
begin
    while model_time ≤ WEEK do
        if model_time = spread_counts × T_Spread then
            Compartmental_SpreadDisease (A, N, G, D, S[ ], p[ ][ ], s[ ][ ], q) ;
            spread_counts = spread_counts + 1 ;
        end
        for i = 1 to |N| do
            for d_1 = 1 to |D| do
                if model_time = node i's disease d_1's status advancement time
                then
                    stage = node i's disease d_1's status ;
                    Advance node i's disease d_1's status ;
                    Change node i's disease d_1's status advancement time
                    based on d[d_1][stage] ;
                end
            end
        end
        model_time = model_time + 1 ;
    end
end
```

Figure A.12: Compartmental model - Part II

Table A.7: Compartmental parameter value - I

Definition of parameter	Parameter notation	Value	Value source
Dynamic Network Model			
Set of contact types	AT	3 (SEX, IDU, SEX and IDU)	NNAHRAY [19]
Set of node types	NT	12 ($FBI, FHI, FSI, \ldots, MSN$)	NNAHRAY [19]
Set of diseases	D	2 (HIV and HSV-2)	NNAHRAY [19]
Age	Age	Uniformly distributed between 18 and 66	NNAHRAY [19]
Gender	$Gender$	57% male and 43% female	NNAHRAY [19]
Group type proportion within women:			
FBN	$pnt[FBN]$	22%	NNAHRAY [19]
FHN	$pnt[FHN]$	1%	NNAHRAY [19]
FSN	$pnt[FSN]$	43%	NNAHRAY [19]
FBI	$pnt[FBI]$	11%	NNAHRAY [19]
FHI	$pnt[FHI]$	0%	NNAHRAY [19]
FSI	$pnt[FSI]$	22%	NNAHRAY [19]
Group type proportion within men:			
MBN	$pnt[MBN]$	12%	NNAHRAY [19]
MHN	$pnt[MHN]$	3%	NNAHRAY [19]
MSN	$pnt[MSN]$	36%	NNAHRAY [19]
MBI	$pnt[MBI]$	9%	NNAHRAY [19]
MHI	$pnt[MHI]$	3%	NNAHRAY [19]
MSI	$pnt[MSI]$	37%	NNAHRAY [19]
Probability of a type A node having a type B contact	$pnc[A][B]$	See Table 3.2.2	NNAHRAY [19]
Probability of a type A node choosing a type B node	$pnn[A][B]$	See Table 3.2.3	NNAHRAY [19]
Probability of a type A node choosing a type B node	$pnn[A][B]$	See Table 3.2.3	NNAHRAY [19]
Spreading scaling parameter	q	$386/(465 \times 464/2) = 0.0035$	NNAHRAY [19]
Number of weeks to run our model	$WEEK$	$(12 \times 52 = 624)$ weeks	Model setup
Time interval between adding new nodes	T_{Add}	4 weeks	Model setup
Time interval between spreading diseases	T_{Spread}	1 week	Model setup

Table A.8: Compartmental parameter value - II

Definition of parameter	Parameter notation	Value	Value source
HIV and HSV-2			
HIV prevalence rate of newcomers from 1990 to 2002:			
Set of HIV's infected stages in our model	S_{HIV}	4 (Primary, Asymptomatic, Symptomatic, AIDS)	[18, 26]
IDU:	$pnd[IDU][HIV]$	50%	[18, 26]
MSM:	$pnd[MSM][HIV]$	47%	[18, 26]
General population:	$pnd[General][HIV]$	9 %	[18, 26]
Probability of an infected node of gender A with HIV to transmit	$p[A][HIV][S_{HIV}]$	See Table 3.4.1	[36, 46]
Duration of HIVs stages	$d[HIV][S_{HIV}]$	See Table 3.4.1	[36, 46]
HSV-2 prevalence rate from 1990 to 2002:			
Set of HSV-2's infected stages in our model	S_{HSV-2}	5 (Primary, Early latent, Latent, Late latent, Recurrent Ulcers)	[36, 46]
General population:	$pnd[General][HSV_2]$	21% before 1996 and 17% after 1996	[48]
Probability of an infected node of gender A with HSV-2 to transmit	$p[A][HSV-2][S_{HSV-2}]$	See Table 3.4.2	[36, 46]
Duration of HIVs stages	$d[HSV-2][S_{HSV-2}]$	See Table 3.4.2	[36, 46]
Scaling effect of HSV-2 on HIV's transmission in Testing 1 and 2	$s[S_{HSV-2}][S_{HIV}]$	See Table 3.4.2	[36, 46]
Scaling effect of HSV-2 on HIV's transmission in Testing 3	0		
Number of new nodes added at the same time in Testing 1	New	10	Model setup
Number of new nodes added at the same time in Testing 2	New	4	Model setup

REFERENCES

[1] Albert, R., Jeong, H., and Barabási, A. L., "Internet: Diameter of the world-wide web," *Nature*, vol. 401, pp. 130–131, 1999.

[2] Anderson, R. M. and May, R. M., "Directly transmitted infectious disease: Control by vaccination," *Science*, vol. 215, pp. 1053–1060, 1982.

[3] Anderson, R. M. and May, R., "Vaccination against rubella and measles: quantitative investigations of different policies," *The Journal of Hygiene*, vol. 90, pp. 259–325, 1983.

[4] Anderson, R. and May, R., "Age-related changes in the rate of disease transmission: implications for the design of vaccination programmes," *The Journal of Hygiene*, vol. 94, pp. 365–436, 1985.

[5] Baccam, P. and Boechler, M., "Public health response to an anthrax attack: An evaluation of vaccination policy options," *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science*, vol. 5, no. 26-34, 2007.

[6] Barabási, A. and Albert, R., "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[7] Butts, C. T., *sna: Tools for Social Network Analysis*, 2013.

[8] Butts, C. T., Handcock, M. S., and Hunter, D. R., *network: Classes for Relational Data*, 2013.

[9] Christensen, C., Albert, I., Grenfell, B., and Albert, R., "Disease dynamics in a dynamic social network," *Physica A: Statistical Mechanics and its Applications*, vol. 389, pp. 2663–2674, 2010.

[10] Chung, F. and Lu, L., "Complex graphs and networks," in *CBMS Regional Conference Series in Mathematics*, no. 107, pp. 97–223, American Mathematical Society, 2006.

[11] Dasbach, E. J., Elbasha, E. H., and Insinga, R. P., "Mathematical models for predicting the epidemiologic and economic impact of vaccination against human papillomavirus infection and disease," *Epidemiolgoical Reviews*, vol. 28, pp. 88–100, 2006.

[12] Dumont, Y., Chiroleu, F., and Domerg, C., "On a temporal model for the Chikungunya disease: Modeling, theory and numerics," *Mathematical Biosciences*, vol. 213, pp. 80–91, 2008.

[13] Enns, E. and Brandeau, M., "Inferring model parameters in network-based disease simulation," *Health Care Management Science*, vol. 14, no. 2, pp. 174–188, 2011.

[14] Esoesie, E., Beckman, R., Shashaani, S., Nagaraj, K., and Marathe, M., "A simulation optimization approach to epidemic forecasting," *PloS one*, vol. 8, no. 6, p. e67164, 2013.

[15] Flaxman, A., Frieze, A., and Vera, J., "A geometric preferential attachment model of networks," *Internet Mathematics*, vol. 3, no. 2, pp. 187–205, 2006.

[16] Fleming, D. and Wasserheit, J., "From epidemiological synergy to public health policy and practice: The contribution of other sexually transmitted diseases to sexual transmission of HIV infection," *Sexually Transmitted Infections*, vol. 75, no. 1, pp. 3–17, 1999.

[17] Fordyce, E., Shum, R., Singh, T., and Forlenza, S., "Economic and geographic diversity in AIDS incidence among HIV exporsure groups in new york city: 1983 to 1985.," *AIDS and Public Policiy Journal*, vol. 4, pp. 103–114, 1998.

[18] Fordyce, J., Forlenza, S., Makki, H., Mojica, B., Nash, D., Peters, V., Sackoff, J., Schwartz, S., Singh, T., Thomas, P., Torian, L., Weisfuse, I., and Pauline, T., "25 years of HIV in New York City: Lessons from surveillance," *Journal of Urban Health*, vol. 78, pp. 669–678, 2001.

[19] Friedman, S. R. and Aral, S., "Social networks, risk-potential networks, health, and disease," *Journal of urban Health*, vol. 78, no. 411-418, 2001.

[20] Friedman, S. R., Bolyard, M., Mateu-Gelabert, P., Goltzman, P., Pawlowicz, M. P., Singh, D. Z., Touze, G., Rossi, D., Maslow, C., Sandoval, M., and Flom, P. L., "Some data-driven reflections on priorities in AIDS network research," *AIDs and Behavior*, vol. 11, pp. 641–651, 2007.

[21] Friedman, S. R., Bolyard, M., Sandoval, M., Mateu-Gelabert, P., Maslow, C., and Zenilman, J., "Relative prevalence of different sexually transmitted infections in HIV-discordant sexual partnerships: Data from a risk network study in a high-risk New York neighbourhood," *Sexually Transmitted Infections*, vol. 84, pp. 17–18, 2008.

[22] Gottlieb, M. S., Schanker, H. M., Fan, P. T., Saxon, A., Weisman, J. D., Pozalski, I., Oppong, C., Koenig, M. A., Foo, G., Hassaballa, H., and others, "Pneumocystis Pneumonia — Los Angeles," *Morbidity and Mortality Weekly Report*, vol. 30, no. 21, 1981.

[23] Hoare, A., Gray, R., and Wilson, D., "Could implementation of Australia's national gay men's Syphilis action plan have an indirect effect on the HIV epidemic?," *Sexual Health*, vol. 9, pp. 144–151, 2012.

[24] Hook, E., Robert, R. O., Nahmias, A. J., Lee, F. F., Campbell, C. H., Glasser, D., and Quinn, T., "Herpes simplex virus infection as a risk factor for human immunodeficiency virus infection in heterosexuals," *Journal of Infectious Diseases*, pp. 251–255, 1992.

[25] I. T. Vieira, R. C., Harper, P. R., and de Senna, V., "Small world network models of the dynamics of HIV infection," *Annals of Operations Research*, vol. 178, pp. 173–200, 2010.

[26] Jarlais, D., Perilis, T., Arasteh, K., Torian, L., Beatrice, S., Milliken, J., Mildvan, D., Yancovitz, S., and Friedman, S., "HIV incidence among injection drug users in New York City, 1990 to 2002: Use of serologic test algorithm to assess expansion of HIV prevention services.," *American Journal of Public Health*, vol. 95, pp. 1439–1444, 2005.

[27] Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., and Barabási, A.-L., "The large-scale organization of metabolic networks," *Nature*, vol. 407, pp. 651–654, 2000.

[28] Kermack, W. O. and Mckendrick, A. G., "A contribution to the mathematical theory of epidemics," *Proceedings of the Royal Society A*, vol. 115, pp. 700–721, 1927.

[29] Kohler, S. H. H.-P., "Sexual network structure and the spread of HIV in Africa: evidence from Likoma Island, Malawi," *AIDS*, vol. 21, pp. 2323–2332, 2007.

[30] Kuperman, M. and Abramson, G., "Small world effect in an epidemiological model," *Physical Review Letters*, vol. 86, pp. 2909–2912, 2001.

[31] Liljeros, F., Edling, C. R., L. A. Nunes Amaral, H. E. S., and Åberg, Y., "The web of human sexual contacts," *Nature*, vol. 411, pp. 907–908, 2001.

[32] Looker, K., Garnett, G., and Schmid, G., "An estimate of the global prevalence and incidence of herpes simplex virus type 2 infection," *Bulletin of the World Health Organization*, vol. 86, pp. 805–812, 2008.

[33] Molloy, M. and Reed, B., "The size of the giant component of a random graph with a given degree sequence," *Combinatorics Probability and Computing*, vol. 7, no. 3, pp. 295–305, 1998.

[34] Newman, M. E. J., "The spread of epidemic disease on networks," *Physical Review E*, vol. 66, p. 016128, 2002.

[35] Newman, N. E. J., "The structure of scientific collaboration networks," *Proceedings of the National Academy of Sciences*, vol. 98, pp. 404–409, 2001.

[36] Orroth, K., Freeman, E., Bakker, R., Buvé, A., Glynn, J. R., Boily, M., White, R., Habbema, J., and Hayes, R., "Understanding the differences between contrasting HIV epidemics in East and West Africa: Results from a simulation model of the four cities study," *Sexual Transmitted Infections*, vol. 83, pp. i5–i16, 2007.

[37] Pastor-Satorras, R. and Vespignani, A., "Epidemic spreading in scale-free networks," *Physical Review Letters*, vol. 86, pp. 3200–3203, 2001.

[38] P.Thomas, R.ODonnell, R.Williams, and M.A.Chiasson, "HIV infection in heterosexual female intravenous drug users in New York City, 1977-80," *The New England Jounal of Medicine*, vol. 319, p. 374, 1988.

[39] Riley, S., Fraser, C., Donnelly, C. A., Ghani, A. C., Abu-Raddad, L. J., Hedley, A. J., Leung, G. M., Ho, L., Lam, T., Thach, T., Chau, P., Chan, K., Lo, S., Leung, P., Tsang, T., Ho, W., Lee, K., Lau, E. M. C., Ferguson, N. M., and Anderson, R. M., "Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions," *Science*, vol. 300, pp. 1961–1966, 2003.

[40] Rocha, L. E. C., liljeros, F., and Holme, P., "Simulated epidemics in an empirical spatiotemporal network of 50185 sexual contacts," *PLoS computational biology*, vol. 7, p. e1001109, 2011.

[41] Volz, E. and Meyers, L. A., "Susceptible-infected-recovered epidemics in dynamic contact networks," *Proceedings of the Royal Soceiety B*, vol. 274, pp. 2925–2934, 2007.

[42] Volz, E., Miller, J., Galvani, A., and Meyers, L., "Effects of heterogeneous and clustered contact patterns on infectious disease dynamics," *PLoS Computational Biology*, vol. 7, no. 6, p. e1002042, 2011.

[43] Wang, W. and Zhao, X., "Threshold dynamics for compartmental epidemic models in periodic environments," *Journal of Dynamics and Differential Equations*, vol. 20, pp. 699–717, 2008.

[44] Watts, D. J. and Strogatz, S., "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[45] Weeks, M. R., Clair, S., Borgatti, S. P., Radda, K., and Schensul, J. J., "Social networks of drug users in high-risk sites: Finding the connections," *AIDS and Behavior*, vol. 6, pp. 193–206, 2002.

[46] White, R., Orroth, K., Korenromp, E., Bakker, R., Wambura, M., Sewankambo, N., R.H.Gray, Kamali, A., James, J., Grosskurth, H., Habbema, J., and Richard, R., "Can population differences explain the contrasting results of the Mwanza, Rakai, and Masaka HIV/Sexually Transmitted Disease intervention trials?: A modeling study," *Journal of Acquired Immune Deficiency Syndromes*, vol. 37, pp. 1500–1513, 2004.

[47] Williams, R. J. and Martinez, N. D., "Simple rules yield complex food webs," *Nature*, vol. 404, pp. 180–183, 2000.

[48] Xu, F., Sternberg, M. R., Kottiri, B. J., McQuillan, G. M., Lee, F. K., Nahmias, A. J., Berman, S. M., and Markowitz, L. E., "Trends in herpes simplex virus type 1 and type 2 seroprevalence in the United States," *The Journal of the American Medical Association*, vol. 296, pp. 964–973, 2006.