Published in final edited form as:

Math Biosci. 2015 September ; 267: 124–133. doi:10.1016/j.mbs.2015.06.013.

Sampling Dynamic Networks with Application to Investigation of HIV Epidemic Drivers

Ravi Goyal^a and Victor De Gruttola^a

^aBiostatistics, Harvard School of Public Health, Boston, MA, USA

Abstract

We propose a method for randomly sampling dynamic networks that permits isolation of the impact of different network features on processes that propagate on networks. The new methods permit uniform sampling of dynamic networks in ways that ensure that they are consistent with both a given cumulative network and with specified values for constraints on the dynamic network properties. Development of such methods is challenging because modifying one network property will generally tend to modify others as well. Methods to sample constrained dynamic networks are particularly useful in the investigation of network-based interventions that target and modify specific dynamic network properties, especially in settings where the whole network is unobservable and therefore many network properties are unmeasurable. We illustrate this method by investigating the incremental impact of changes in networks properties that are relevant for the spread of infectious diseases, such as concurrency in sexual relationships. Development of the method is motivated by the challenges that arise in investigating the role of HIV epidemic drivers due to the often limited information available about contact networks. The proposed methods for randomly sampling dynamic networks facilitate investigation of the type of network data that can best contribute to an understanding of the HIV epidemic dynamics as well as of the limitations of conclusions drawn in the absence of such information. Hence, the methods are intended to aid in the design and interpretation of studies of network-based interventions.

Keywords

Dynamic Network; Constrained Dynamics; HIV; Network Process; Sampling; Concurrency

1. Introduction

This paper proposes a method for uniformly sampling random dynamic networks with particular temporal network constraints given a cumulative network. Such sampling facilitates investigation of the potential impact of varying some network properties while holding others fixed at particular values. We denote the cumulative network as $G(T_1, T_2)$, which represents all relationships over the period of time between T_1 and T_2 among a set of

ravi.goyal@mail.harvard.edu (Ravi Goyal).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

individuals, V . We further denote the set of relationships among individuals as E , and an edge that connects individuals i and j , as E_{ij} . A dynamic network consistent with the cumulative graph $G(T_1, T_2)$ can be represented as a set of tuples, where each tuple, $(E_{ij}, s(E_{ij}), e(E_{ij}))$, represents an edge $E_{ij} \in E$, and the start and end times of the relationship, $s(E_{ij})$ and $e(E_{ij})$, respectively, under the constraint that $s(E_{ij}) \leq T_2$ and $e(E_{ij}) \leq T_1$. Figure 1 shows three very simple dynamic networks, using the tuple notation. Each dynamic network has the same cumulative contact network, $G(T_1 = 0, T_2 = 6)$; only the timing of relationships differs (the bottom rows of the figure provide information about temporal constraints that will be described later in the paper). Below we refer to $G(T_1, T_2)$ as G when the time period of interest is unambiguous. Our focus is on dynamic networks in which relationships form and dissolve, but the population represented by the network is held constant during the evolution of the network (i.e., no individuals are added or removed).

Methods to sample constrained dynamic networks are particularly useful in the investigation of the impact of network-based interventions that target and modify specific dynamic network properties, especially in settings where the whole network is unobservable and therefore many network properties are unmeasurable. We denote the targeted network properties as $C_1, \dots, C_p$, and the measured and unmeasured cumulative network properties as $z_1, \dots, z_q$ and $f_1, \dots, f_q$, respectively. These methods will permit us to develop the two approaches, denoted as A1 and A2, to investigate network-based interventions:

- A1: Analyze the incremental impact of changes to specific network properties, $C_1, \dots, C_p$, while holding $z_1, \dots, z_q$ and $f_1, \dots, f_q$ fixed on outcome metrics of processes that propagate on networks
- A2: Analyze the incremental impact of changes to unmeasured network properties, $f_1, \dots, f_q$, while holding $C_1, \dots, C_p$ and $z_1, \dots, z_q$ fixed on outcome metrics.

Development of such dynamic network sampling methods is challenging because modifying one network property generally tends to modify others as well. For example, linear regression methods to estimate the the impact of the amount of concurrency in sexual relationships on epidemic spread have been described (Doherty et al., 2006; Ghani et al., 1997; Ghani and Garnett, 2000). However, the required underlying assumptions of linearity and independence may not hold as many network properties exhibit sharp thresholds and are highly correlated. We use concurrency in a general sense of individuals having overlapping sexual relationships here; see Materials and methods for the specific definition used in this paper.

The development of our method was motivated by the attempt to clarify the complexities inherent in making conclusions regarding concurrency based on observed data. Considerable efforts have been made to use mathematical models to investigate the potential of concurrent partnerships to increase the spread of HIV within communities (Morris et al., 2009; Morris and Kretzschmar, 1997; Kretzschmar and Morris, 1996; Eaton et al., 2011). For a good review of research on concurrency epidemic modeling see Goodreau (2011). More recently, investigators have used empirical evidence to evaluate this issue (Tanser et al., 2011; Sawers and Stillwaggon, 2010). However, many challenges arise in the investigation of the role of concurrency (Lurie and Rosenthal, 2010; Morris, 2010), such as the incompleteness of

information regarding network structure. Morris and Epstein (2011) noted that data on mixing patterns by geographic location were not sufficient to substantiate claims by Tanser et al. (2011). Given the evidence for variability in network properties like concurrency and mixing patterns (Morris et al., 2007), it is plausible that communities differ in other network properties that are not observed. For example, a broad range of parameter values that characterize mixing by sexual activity level have been used in mathematical models that simulate HIV spreading in a population (Eaton et al., 2012; Palombi et al., 2012; Morris et al., 2009). In this paper, we focus on development of methods to sample dynamic networks to aid in the interpretation of conclusions regarding the potential impact of intervening on network features like concurrency in the presence of unmeasured network properties.

We investigate implications of unmeasured sexual network structure by considering the impact of concurrency in sexual relationships on a particular network property, the size of the reachable path of infection (Morris et al., 2009), in a hypothetical population. The reachable path (RP) of infection includes the initial infected individual(s) and the subsequent time-ordered sequence of partnerships along which transmission is possible. In these investigations, we employ both approaches, A1 and A2, defined above. This is achieved by generating a collection of cumulative networks, $\{G_1, \dots, G_m\}$ with the following properties:

- $z_l(G_i) = z_l(G_j)$ for $l \in \{1, \dots, q\}$ and $i, j \in \{1, \dots, m\}$
- $f_k(G_i) = f_k(G_j)$ if $i = j$,

for each $k \in \{1, \dots, q\}$. There has been substantial amount of work devoted to generating collections of networks with particular network properties (Newman, 2002; Blitzstein and Diaconis, 2010). For example, to fix degree distribution and vary any arbitrary network property f , one can use the configuration model in combination with an edge switching algorithm—the procedure used to generate the network collection below (see Newman (2010) for additional details). With this collection of networks, the proposed method enables one to investigate the impact on an outcome metric of a network process using the approaches A1 and A2. For approach A1, one can sample, analyze, and compare dynamic networks consistent with G_i for two sets of values for $C_1, \dots, C_p$ for each $i \in \{1, \dots, m\}$. For approach A2, one can investigate trend the outcome metric of a network process as f_k increases but $z_1, \dots, z_q$ and $C_1, \dots, C_p$ are held constant, by sampling dynamic networks consistent with each network G_i in the collection of networks.

In this paper, we present evidence that data from ego-centric sexual history surveys—the most common sources of such data—may not be sufficient to address the role of concurrency on features like size of RP, which is related to potential epidemic size, noting that there are many factors that can influence the degree to which RP correlates with the size of the epidemic. This paper demonstrates the challenges in evaluating the impact of concurrency in the presence of missing information on network properties, thereby underscoring the uncertainty of evaluating the impact of concurrency on HIV incidence from observational data.

2. Material and methods

Assuming a given cumulative network G , the following section provides details on calculation of the expected size of the reachable path if dynamic networks are sampled from a uniform distribution in a way that assures that each set of start times is consistent with a particular set of temporal constraints. We assume unit length for relationship duration so only start times are necessary, noting that none of the theory presented below relies on this simplification, which decreases computation time dramatically. Although we focus on the largest reachable path (LRP), our methods are general and can be used for investigation of range of statistics associated with the reachable paths, including mean reachable paths, which was discussed in Morris et al. (2009).

We simultaneously impose two temporal constraints that represent particular aspects of concurrency. Each sampled dynamic network that will have exactly the pre-specified value for the following two metrics:

- C_1 : the number of people with sexual partnerships that overlap during the time period
- C_2 : the number of overlapping relationships during the time period

We further investigate how to estimate the RP when constraining an additional third concurrency metric, C_3 , denoting the total amount of time that relationships overlap during the time period. The three cumulative concurrency metrics measure different aspects of concurrency (size of the population that participates in concurrency, number of relationships that are concurrent, and the duration of the overlapping relationships). There is no single network statistic or set of statistics that fully capture all of the many notions of concurrency, whose variability has led to ambiguity debate (Lurie and Rosenthal, 2010; Morris, 2010). Refer to figure 1 for an illustration of the three concurrency metrics. The first scenario (second panel on the left) includes only sequential relationships (no concurrency); the second and third scenarios (third and fourth panels) fix three different concurrency metrics to specific values.

The algorithm to calculate the mean length of the largest reachable path of G is based on sampling subspaces of $\mathbb{R}^{|E|}$, where the coordinates for each point in the subspace represent start times for the set of relationships in G ; assuming unit length for relationship duration causes the end times to be determined by the start times. Figure 1 depicts a simple example of converting the tuple representation of a dynamic network to a point in $\mathbb{R}^{|E|}$. Denote $S(G) \subset \mathbb{R}^{|E|}$ as the subset where each point $s \in S(G)$ is consistent with the three values of concurrency, C_1 , C_2 and C_3 for a fixed cumulative contact network G . When there is no ambiguity, we denote $S(G)$ as S . $LRP(G)$ denotes the mean length of the largest reachable path for a contact network G and given concurrency values, and $LRP(S)$ denotes the mean length of the largest reachable path for the points in S when there is no ambiguity regarding the cumulative contact network G . Therefore $LRP(G)$ is equal to the following,

$$LRP(G) = LRP(S) = \frac{1}{Vol(S)} \int_S LRP(s) ds \quad (1)$$

where $Vol(S)$ denotes the volume of the subset S . As computing (1) directly is difficult, we approximate $LRP(S)$ using a Markov Chain Monte Carlo (MCMC) method. The MCMC method is based on sampling subsets $\{H_1, \dots, H_m\}$ where $H_i \subset \mathbb{R}^{|E|}$ with specific properties (outlined below); one requirement is that all points $h \in H$ have identical values for $LRP(h)$. There are computational and other advantages to sampling subsets of $\mathbb{R}^{|E|}$ consisting of points with identical LRP , e.g. computing the LRP for one point and provides the LRP for all the remaining points in the subset. Because sampling subsets of S with identical values for LRP is a complex process, we sample subsets for which all points in the subset have identical LRP from a larger space, $H_\Omega(G) \supseteq S$, where $H_\Omega(G)$ is the subset of $\mathbb{R}^{|E|}$ that is consistent with C_1 and C_2 but not necessarily C_3 for a graph G . Again, when there is no ambiguity, we denote $H_\Omega(G)$ as H_Ω . Partitioning H_Ω such that each element has a distinct value for the length of the largest reachable path is challenging because an element of the partition may not be connected in $\mathbb{R}^{|E|}$, and a connected space is necessary for our method. Therefore, we opt for a finer partition of H_Ω in which all reachable paths—not just the largest—are identical within an element of the partition; section 3.1 proves that each element of this partition is connected. Formally a valid partition $\{H_\beta\}_{\beta \in B}$ of H_Ω is defined by having the following property:

P1 If $p_1, p_2 \in K_\beta$ then $V^{p_1}(A) = V^{p_2}(A) \forall A \in V(G)$, where $V^p(A)$ denotes the set of nodes that are reachable from node A using the set of relationship start times defined by point $p \in \mathbb{R}^{|E|}$.

Section 3.1 describes a valid partition, $\{H_\beta\}_{\beta \in B}$, of H_Ω , i.e. a partition satisfying P1 by using linear constraints. The region $\cup_{\beta \in B} H_\beta$ does not fix the concurrency value C_3 , the total amount of time that relationships overlap; therefore, $S \subset \cup_{\beta \in B} H_\beta$. We can define a partition $\{X_\beta\}_{\beta \in B}$ of S , where $\{X_\beta\}_{\beta \in B} = \{H_\beta\}_{\beta \in B} \cap S$. Figure 2 provides an illustration of two valid partitions, one in which dynamic networks have no concurrent relationships and the other in which they do, for a simple contact network.

The subset S is equal to the following,

$$S = \cup_{\beta \in B} (H_\beta \cap X_\beta). \quad (2)$$

Using the property that $\{H_\beta\}_{\beta \in B}$ is a partition and by substituting (2) into equation (1),

$$LRP(S) = \frac{1}{Vol(S)} \sum_{\beta \in B} \int_{s \in H_\beta \cap X_\beta} LRP(s) ds. \quad (3)$$

Because the size of the largest reachable path is constant in H_β ,

$$LRP(S) = \frac{1}{Vol(S)} \sum_{\beta \in B} Vol(H_\beta \cap X_\beta) * LRP(H_\beta). \quad (4)$$

As we are interested in uniformly sampling points (i.e. sets of start times for the dynamic network simulation) that are consistent with the three metrics of concurrency, the probability of sampling a subspace is proportional to its volume. Therefore,

$$LRP(S) \propto \sum_{\beta \in B} P(H_\beta \cap X_\beta) * LRP(H_\beta) \quad (5)$$

$$= \sum_{\beta \in B} P(H_\beta) * P(X_\beta | H_\beta) * LRP(H_\beta). \quad (6)$$

Estimation of $LRP(S)$ using equation (6) requires subspaces of the form H_β to be sampled with probability proportional to $P(H_\beta) * P(X_\beta | H_\beta)$. Section 3.1 describes a valid partition of H_Ω while section 3.2 outlines an algorithm to generate a sample of H_β under a probability distribution, Q , which is different from a probability distribution defined by sampling H_β proportional to $P(H_\beta) * P(X_\beta | H_\beta)$. Therefore, appropriate adjustment of the procedure used in section 3.2 to sample subspaces requires importance weights; the procedure to re-weight the samples is shown in section 3.3.

2.1. Define a Valid Partition of H_Ω

We construct a valid element of the partition of H_Ω by designating a time-ordered relationship between specific pairs of edges in G . The computation cost increases sharply for each specified pair of edges, as will be shown in sections 3.2 and 3.3. Therefore, it is advantageous to limit the number of edge pairs in which these relationships are ordered in time. In this section, we first describe rules that are sufficient to ensure that $P1$ holds, i.e. a valid partition, and then demonstrate that some or all are necessary depending on the graph.

In order to exploit the fact that concurrency provides insight into the time order of pairs of sexual relationships, we utilize a concept from graph theory called the line graph of G , $L(G)$, which represents the adjacencies between edges of G (Harary, 1969; Morris and Kretschmar, 1997). $L(G) = (V_L, E_L)$ is an undirected simple graph where $V_L = E$ and $E_L = \{(E_{ij}, E_{ik})\}$, i.e. pairs of edges with a common endpoint in G . We use terminology from interval graph theory (Pe'er and Shamir, 1995) to represent associations between pairs of relationships. Let f be a mapping from pairs of edges in G , $E \times E$, to a set of designations, $\{\cap, <, >, \cup\}$. We represent an element in the domain of f either as an element from $E \times E$ or an edge in $K_{L(G)}$, as they both represent the same pairs of edges in G . To build our partition, certain edges in $K_{L(G)}$, i.e. pairs of edges in G , are mapped to one of the four designations $\{\cap, <, >, \cup\}$. The four designations describe a time-ordering of pairs of relationships in G . Before describing the designations, we introduce some notation. Let $s(E_{ij})$ and $e(E_{ij})$ denote the start and end times for the relationship $E_{ij} \in E$. As we assume relationships have unit length, $e(E_{ij}) = s(E_{ij}) + 1$. The four designations are defined as the following:

1. $f(E_{ij}, E_{kl}) = '\cap'$: Relationship E_{ij} overlaps with relationship E_{kl} , i.e. $s(E_{ij}) + 1 > e(E_{kl})$ and $s(E_{kl}) + 1 > e(E_{ij})$.
2. $f(E_{ij}, E_{kl}) = '<'$: Relationship E_{ij} starts and ends before E_{kl} , i.e. $s(E_{ij}) + 1 < s(E_{kl})$.

3. $f(E_{ij}, E_{kl}) = '>'$: Relationship E_{kl} starts and ends before E_{ij} , i.e. $s(E_{kl}) + 1 < s(E_{ij})$
4. $f(E_{ij}, E_{kl}) = '\cup'$: Relationship E_{ij} and E_{kl} can have any time order relationship, i.e. no restriction.

Figure 3 gives an illustration of each of the three restrictive designations. The fourth designation, ' $\cup$ ', does not place any restriction on the time order of relationships and hence is the union of all the other three.

From the designation descriptions and figure 3, we can see that each of the time order designations imposes a linear constraint on $\mathbb{R}^{|E|}$ see Fishburn (1985); Pe'er and Shamir (1995) for further details. To understand the connection between the time-order designations and H_Ω , we revisit the example in figure 2. In the example with concurrency, $f(E_{AB}, E_{AC}) = '\cap'$. For the example without concurrency, H_1 corresponds to $f(E_{AB}, E_{AC}) = '<'$ and H_2 corresponds to $f(E_{AB}, E_{AC}) = '>'$.

Each element in our partition $\{H_\beta\}_{\beta \in B}$ of H_Ω is defined by the restrictions on the time-ordering of relationships, i.e. defining an f for each partition element. Let f^{H_β} denote the mapping associated with element H_β in $\{H_\beta\}_{\beta \in B}$. The time-order restrictions need to ensure that (1) we are limited to H_Ω , i.e. only points in $\mathbb{R}^{|E|}$ consistent with C_1 and C_2 and (2) $P1$ is satisfied. First we define a coarser partition $\{H'_\beta\}_{\beta \in B'}$ of H_Ω that ensures (1) by defining functions $f^{H'_\beta}$. Once we define maps for each element in $\{H'_\beta\}_{\beta \in B'}$, we place additional restrictions to ensure (2).

To construct the partition $\{H'_\beta\}_{\beta \in B'}$ we define a set of maps, $\{f^{H'_\beta}\}_{\beta \in B'}$, such that each map corresponds to an element in $\{H'_\beta\}_{\beta \in B'}$. There are three properties, $O1$ – $O3$, that all maps in $\{f^{H'_\beta}\}_{\beta \in B'}$ must have to ensure each H'_β in the partition is consistent with C_1 and C_2 . First ($O1$), exactly C_2 edges, E_o , from E_L must be mapped to ' $\cap$ ', i.e. the relationships overlap in time. Second ($O2$), the edges must represent exactly $|V| * C_1$ distinct individuals. Third ($O3$), all edges $E_{no} := E_L \setminus E_o$ are mapped to $\{<, >\}$. The collection of all the maps satisfying these properties forms a partition of H_Ω that is consistent with C_1 and C_2 . We let $E_o^{H'_\beta}$ and $E_{no}^{H'_\beta}$ denote the edges in E_L for a particular map $f^{H'_\beta}$, which do and do not overlap. As mentioned above and demonstrated in sections 3.2 and 3.3, there are computational advantages to minimizing the number of restrictions; therefore, each function in $\{f^{H'_\beta}\}_{\beta \in B'}$ maps a pair of edges not represented in $L(G)$ to ' $\cup$ ', i.e. no time order restriction.

To construct the partition $\{H_\beta\}_{\beta \in B}$ we replace each map $f^{H'_\beta}$ with a set of mappings $\{f_1^{H'_\beta}, f_{l_\beta}^{H'_\beta}\}$. The set of mappings are constructed by remapping certain pairs of edges, $E_a^{H'_\beta}$, from ' $\cup$ ' to an element in $\{<, \cap, >\}$, while not changing mappings from pairs of edges in G

represented by $K_{L(G)} \setminus E_a^{H'_\beta}$. To identify $R_a^{H'_\beta}$ we need to introduce additional terminology.

Let $SL^{H'_\beta}(G) \subset L(G)$ be a subgraph of $L(G)$ defined by including only $R_o^{H'_\beta}$ edges in $L(G)$.

Let $SL_1^{H'_\beta}(G), \dots, SL_{l_\beta}^{H'_\beta}(G)$ denote the connected components of $SL^{H'_\beta}(G)$ and define

$R^{H'_\beta}$ as $K_{SL_1^{H'_\beta}(G)} \cup \dots \cup K_{SL_{l_\beta}^{H'_\beta}(G)} \cup L(G)$. Denote the edges in $R^{H'_\beta}$ as $E_R^{H'_\beta}$. $E_a^{H'_\beta} = E_R^{H'_\beta} \setminus E_L$.

We refer to R^{H_β} as the reachable path restriction graph of f^{H_β} . Theorem 1 states that these additional restrictions are sufficient to construct a valid partition, i.e.

$\{f^{H_\beta}\}_{\beta \in B} := \{f_i^{H'_\beta}\}_{i \in [1, l_\beta], \beta \in B'}$ forms a partition of H_Ω that is consistent with C_1 , C_2 and $P1$.

Theorem 1—Let $G = (V, E)$ be any contact graph. Let $L(G) = (V_L, E_L)$ represent the line graph of G . Let f be a mapping from $E \times E$ to $\{<, >, \cap, \cup\}$, with the following property:

- $f(E_1, E_2)$ maps to $\{\cap, <, >\}$ if and only if $(E_1, E_2) \in R$, where R is the reachable path restriction graph of f and $E_1, E_2 \in E$.

If $p_1, p_2 \in H$ then $V^{p_1}(A) = V^{p_2}(A) \forall A \in V$ where H is the unique subset defined by the mapping f .

The proof can be found in the Appendix. Figure 4 provides a simple example illustrating the

necessity for mapping edges in $K_{SL_1^{H'_\beta}(G)}, \dots, K_{SL_{l_\beta}^{H'_\beta}(G)}$. Next we describe a method for sampling a valid mapping function f in order to define a sample H_β .

2.2. Sampling Subspace H_β

There are $\tau = \sum_{i=1}^{|\Delta|} 2^{|E_{no}^i|} \times 3^{|E_H^i| - |E_L^i|}$ possible mapping functions consistent with C_1 , C_2 and $P1$, where τ represents that possible ways of selecting E_o that is consistent with $O1 - O2$. However, many possibilities do not lead to a valid realization (Pe'er and Shamir, 1995), i.e. pairs of real values representing the start and end times of each relationship such that the designations are preserved. Though τ is already a reduction in the number of mapping functions compared to the total number possible, it still may not be possible (or even if possible efficient—many maps do not lead to a valid realization) to sample functions f in a direct manner. Therefore, this section describes methods to increase efficiency of sampling mapping functions.

To sample a subspace H_β i.e. a map f , we first designate edges as E_o and E_{no} , which satisfy $O1$ and $O2$. Let $\{SL_1(G), \dots, SL_l(G)\}$ denote the connected components of $L(G)$ using only edges in E_o . Gains in efficiency are mainly achieved from the recognition that a valid realization for each $SL_i(G)$ implies the existence of a valid mapping function f consistent with the chosen overlapping edges, E_o . This can be shown by assigning a distinct block of time for each connected component, $SL_i(G)$. We proceed by first mapping edges in E_{no} contained in each connected component $SL_i(G)$. As $C_2 \ll |E|$, $SL(G)$ is extremely sparse, both the average degree and the expected second-order average degree are typically less than

one. Hence, all of the components are small (Chung and Lu, 2002); using methods in Pe'er and Shamir (1995), it can be easily verified by brute force whether a valid realization exists for each component $SL_i(G)$. Next, we map the remaining edges in E_{no} , i.e. edges connecting distinct components $SL_i(G)$ and $SL_j(G)$. This is done by starting with an empty set A and sequentially adding a single additional connected component to the set A . At each step of adding a component, $SL_i(G)$, to A , the subset of E_{no} edges connecting $SL_i(G)$ to any component in A are mapped to either ' $<$ ' or ' $>$ '.

Let H'_β be the subspace defined using the mapping from edges only in $L(G)$. To include the additional constraints, we sample a realization of H'_β using a method to sample points uniformly from a polytope (Smith, 1984) and map in the necessary edges, $E_a^{K'_\beta}$, to $\{\cap, <, >\}$ using the realization.

2.3. Re-weighting of the Samples of Subspace K_β

To evaluate equation (8), subspaces H proportional to $P(H) * (X|H)$ need to be sample, but the procedure in section 3.1 generates samples of H_β under a different distribution, Q . Therefore, we use importance weights to adjust for the discrepancy between the two probability distributions. To generate a sample H , we first generate an H' , and then sample K proportional to volume of H in H' . Therefore, $Q(K) = Y(H') * P(H/H')$, where $Y(H')$ is the

probability of sampling H' and $P(H|H') = \frac{Vol(H)}{Vol(H')}$.

The formula to calculate the estimated expected reachable path vector for a graph G uses the following weights for each sampled subspace, $H_1, \dots, H_m$.

$$w_i = \frac{h(H_i)/g(H_i)}{\sum_{j=1}^m h(H_j)/g(H_j)}$$

where,

$$h(H_i) := P(H_i) * P(X_i|H_i) \quad (7)$$

and,

$$g(H_i) := Q(H_i) = Y(H'_i) * P(H_i|H'_i) \quad (8)$$

Our procedure requires calculation of $P(H_i|H'_i)$, $P(H_i)$, $P(X_i|H_i)$, and $Y(H'_i)$.

To estimate $P(X/H)$, the proportion of H that is consistent with C_3 , we first sample points from H using the procedure in Pe'er and Shamir (1995). Once a set of points is sampled from H , calculation of C_3 for each point is straightforward. Since the joint distribution of C_3 and the starting times of relationships is smooth, kernel density estimation procedures may be used (Rosenblatt, 1956) to approximate $P(X/H)$.

Under the uniform distribution, $P(H_i) \propto \text{Vol}(H_i)$. To calculate the exact volume of a polytope is very difficult (Bárány and Füredi, 1987); approximating this volume is an active research area in computational geometry (Dyer et al., 1991). One approximation approach generates spaces $H_i = R_q \subseteq \dots \subseteq R_1$, where the volume of R_1 is easy to calculate. Once these nested subspaces are defined, the volume of H_i can be rewritten as:

$$\text{Vol}(H) = \text{Vol}(R_1) * \prod_{j=1}^{q-1} \frac{\text{Vol}(R_{j+1})}{\text{Vol}(R_j)} \quad (9)$$

The ratio of $\frac{\text{Vol}(R_{j+1})}{\text{Vol}(R_j)}$ for $j \in 1, \dots, q$ can be estimated using the proportion of sampled points in R_j that are also contained in R_{j+1} .

In our simulation R_1 is the space defined by distinct paths in $SL(G)$, $p_1, \dots, p_m$, that cover all the vertices. Thus,

$$\text{Vol}(R_1) = \prod_{j=1}^m \frac{1}{|p_j|} * (-(|p_j| - 1) * l + T)^{|p_j|}, \quad (10)$$

where l is the length of a relationship and T is the time period of interest for the study (in our case $l = 1$ month and $T = 12$ months). R_{j+1} is defined as R_j plus one additional edge designation from H . Since R_1 is a subset of designations in K'_i , $\exists$ a sequence of edge designations and a j such that $R_j = K'_i$. So, it is straightforward to calculate

$$P(K|H') = \frac{\text{Vol}(H)}{\text{Vol}(H')}. \text{ Finally, } Q(H'_i) \text{ can be computed using a brute force method.}$$

Using samples set of samples, $\{H_1, \dots, H_m\}$ and their associated weights, $\{w_1, \dots, w_m\}$, we approximate the LRP for the subspace that fixes C_1 , C_2 , and C_3 as:

$$\text{LRP}(S) = \frac{1}{m} \sum_{i=1}^m w_i * \text{LRP}(H_i) \quad (11)$$

3. Results

3.1. Simulation Input: Concurrency Metrics and Cumulative Degree Distribution

To illustrate our methods, we consider a hypothetical population for which the network is unimodal, such as would be the case if the all transmission occurred within people of the same gender. We use data from a pilot study of antiretroviral treatment for prevention of HIV transmission in the village of Mochudi, Botswana. The pilot study collected information on the start and end times of the respondents last three relationships in a household survey conducted in one sector of the village. This information is used in estimating values for the concurrency metrics. We note that as these data do not arise from people in a unimodal sexual network, the population we simulate is not a realistic description of the residents Botswana village, but serves only to illustrate how our network sampling methods work in practice. Consideration of a bipartite network is possible with our

methods but would be very computationally intensive. Although data are available on sexual behavior for men who have sex with men (a unimodal network), the high degree and rapid rates of partner change associated with this population make it more difficult to base simulations on such data. The degree distribution used in the simulations is shown in table 1 and represents only individuals who were sexually active in the last year. The distribution closely resembles that observed among older adolescents in Rakai district, Uganda (Konde-Lule et al., 1997).

In our simulation study, the population size, n , was set at 50. From the pilot study in Mochudi, Botswana, the observed percentage of individuals with overlapping relationships was 10.04%, $C_1 = \lceil .1004 * n \rceil$; each person surveyed in Mochudi had an average number of pairs of relationships that overlap equal to 0.123, so $C_2 = \lceil .123 * n \rceil$. For simplicity, we assume uniformity in relationship duration; without loss of generality this length can be a single unit. To accommodate this assumption, the third metric of concurrency, C_3 was set equal to $0.64 * C_2$, which is the mean fraction of time relationships overlap multiplied by the number of relationship pairs that are concurrent.

3.2. Simulation Output: Largest Reachable Path

Our interest lies in investigating the impact of concurrency on the largest reachable path (LRP). As mentioned above, to create reachable paths, we simulate propagation of epidemic models on the networks. We start by selecting an individual to be infected and set the probability of transmission to 1. The reachable path includes the initial infected individual and the subsequent time-ordered sequence of partnerships along which transmission was possible. The time-ordered sequence is generated by sampling start times for relationships in the cumulative network that are consistent with the concurrency metrics, C_1 , C_2 , and C_3 .

3.3. Simulation Setup

To investigate the incremental effect on the length of the largest reachable path of a cumulative network property f , beyond that of cumulative degree distribution and concurrency, we generated a collection of networks, $\{G_1, \dots, G_m\}$ with the property that $f(G_i) < f(G_k)$ if $i < k$. In particular, we investigate the trend of $LRP(G_i)$ as i increases, where $LRP(G_i)$ is the average length of the largest reachable path for contact network G_i . $LRP(G_i)$ was estimated by constructing dynamic networks where relationship start times that are consistent with given measures of concurrency were randomly sampled from a uniform distribution (see Methods). As the degree distribution and concurrency measures are identical in all networks, the variation in average potential epidemic sizes across networks characterizes the added value of knowledge of network property f . By using the same set of network collections, and fixing the three metrics for concurrency to zero, we estimate the size of the LRP in populations with identical set of sexual relationships, but for which concurrent relationships do not occur.

This paper investigates three important network properties: degree assortativity coefficient, clustering coefficient, and proportion of mixing between two equal sized subpopulations. The degree assortativity coefficient summarizes the propensity for individuals with similar number of cumulative partners to form relationships; the exact formula can be found in

Newman (2002). The clustering coefficient is the probability that two partners of the same individual form a partnership; this coefficient is formally defined as the fraction of connected triples that are three-cycles. The proportion of mixing is defined as the proportion of all relationships in the cumulative network that occur between individuals in two different subpopulations of equal size.

3.4. Simulation Results

Figure 5 displays the relationship between network property and percentage of individuals in the largest reachable path (pLRP) for communities with and without concurrent relationships. For the former C_1 , C_2 , and C_3 are set to the values specified in the simulation input section, while for the latter, $C_1 = C_2 = C_3 = 0$. Each black point represents the weighted mean value of pLRP among communities with concurrency; each red point represents the weighted mean pLRP among communities without concurrency. Each mean was taken over a collection of dynamic network simulations (median collection size is 370 with a range of 56-2130) that correspond to a set of five distinct cumulative networks. The variation of collection sizes are influenced by the proportion of valid dynamic networks for a particular cumulative network. The black and red lines represent the lowess curves for the scenarios with and without concurrency, respectively.

Figures 5a and 5b demonstrate a strong relationship between LRP and both degree mixing and clustering; the relationship is sufficiently strong to make it possible to depict two communities, labeled X and Y, for which the community with no concurrency has a higher potential epidemic size than does the community with concurrency. Figure 5c shows mixing between two groups has a weak relationship with the potential size of the epidemic.

Unlike the relationship between LRP and degree mixing, the relationship between LRP and clustering and between LRP and subpopulation mixing were strongly modified by the presence of concurrent partnerships. Higher values for the clustering coefficient sharply reduced the impact of concurrency. A high level of clustering creates multiple pathways for infection between individuals, thereby generating a similar mechanism used by concurrency to increase disease spread. Since communities with high clustering already contain many multiple pathways for infection, there is diminishing marginal impact due to concurrent partnerships creating additional pathways. Though in this paper we consider the clustering coefficient, the same logic applies for any subpopulation in a network with a high density of relationships among its members.

As the two curves are nearly parallel in figure 5a, the predicted increase in incidence due to concurrency appears to be independent of degree assortativity. Consequently, it is possible to estimate the average impact of concurrency reduction interventions on potential epidemic size without information on degree assortativity—a difficult quantity to measure—provided that other networks properties are held fixed. By contrast, the same does not hold for either clustering or mixing between two subpopulations. Figure 5 demonstrates that regardless of the value for the (A) degree assortativity coefficient, (B) clustering coefficient, or (C) proportion mixing between two groups, concurrency increases the average potential size of an epidemic.

4. Discussion

To illustrate the usefulness of our method, we applied it to a simplified version of a complex system that depends on both behavior and biological processes. We considered only reachable paths and did not attempt to include the biological characteristics of HIV transmission that are relevant in actual epidemics, such as the increased risk of transmission during the acute HIV infection or decreases in risk of transmission due to coital dilution (Sawers et al., 2011). The population size used was small due to computational complexity of the method; nonetheless some communities may have small sexual networks (Carnegie and Morris, 2012). However, we believe these limitations do not affect our fundamental point: sampling dynamic networks in ways that hold particular network properties fixed while others are varied allows investigation of network processes that have important implications for practice. In particular, we show that for our hypothetical population, network properties that are not estimable from ego-centric data are necessary to make conclusions about the impact of concurrency on LRP.

Collecting and analyzing ego-centric sexual network data can be difficult, time-consuming, and costly, and network centric data only adds to this challenge. It may not be possible to collect network centric data in settings of interest, though some studies have succeeded in this goal (Helleringer and Kohler, 2007). In any case, it is useful to investigate the way in which information regarding network centric data can contribute to an understanding of the HIV epidemic dynamics as well as the limitations of conclusions drawn in the absence of such information. Such research is predicated on the development of methods, like those described above, for randomly sampling dynamic networks.

Appendix: Proof of Theorem 1

Theorem 1

Let $G = (V, E)$ be any contact graph. Let $L(G) = (V_L, E_L)$ represent the line graph of G . Let f be a mapping from $E \times E$ to $\{\cap, <, >\cup\}$ with the following property:

- $f(E_1, E_2)$ maps to $\{\cap, <, >\cup\}$ if and only if $(E_1, E_2) \in R$, where R is the reachable path restriction graph of f and $E_1, E_2 \in E$.

If $p_1, p_2 \in H$ then $V^{p_1}(A) = V^{p_2}(A) \forall A \in V$ where H is the unique subset defined by the mapping f .

Proof by Induction

Let H be a subspace that is defined by f . Let $p_1 = \{t_1, \dots, t_{|E|}\}$ and $p_2 = \{r_1, \dots, r_{|E|}\}$ be two points in H . Therefore p_1 and p_2 each represent starting times for the set of relationships E . Let $s^{p_l}(A, B)$ represent the start time of the relationship between individuals A and B . Hence, $s^{p_l}(A, B)$ is equal to one of the coordinates in p_l where $l \in \{1, 2\}$. Let $e^{p_l}(A, B)$ represent the end time of the relationship between individuals A and B . Denote $V_i^{p_l}(A) \subseteq V^{p_l}(A)$ as the subset of nodes that are reachable from $A \in V$ using time set p_l via a path of exactly i steps where $l \in \{1, 2\}$. If $B \in V_i^{p_l}(A)$ then there exists a subset of distinct

nodes, $\{B_1, \dots, B_{i-1}\}$ such that the following is a valid reachable path under time set p_l : $A = B_0 \rightarrow B_1, B_1 \rightarrow B_2, \dots, B_{i-1} \rightarrow B_i = B$, i.e. A infects B_1 , B_1 infects B_2 , and finally B_{i-1} infects B . We define $I_A^{p_l}(B)$ as the infection time of B where only node A was initially infected using time set p_l . We want to show that $V^{p_1}(A) = V^{p_2}(A) \forall A \in V$. Since $\cup_i V_i^{p_l}(A) = V^{p_1}(A)$, if $V_i^{p_1}(A) = V_i^{p_2}(A) \forall A \in V$ and $\forall_i \in \mathbb{Z}^+$ then $V^{p_1}(A) = V^{p_2}(A) \forall A \in V$. Therefore, we will show that $V_i^{p_1}(A) = V_i^{p_2}(A) \forall A \in V$ and $\forall_i \in \mathbb{Z}^+$ by induction on i . We first present two base cases before the induction step.

Base Case: $i = 1$

Want to show that if $B \in V_1^{p_1}(A)$ then $B \in V_1^{p_2}(A)$. Let $B \in V_1^{p_1}(A)$, therefore there exists a path $A = B_0 \rightarrow B_1 = B$ in time set p_1 , i.e. A directly infects B . Thus, $(A, B) \in E$. So, regardless of the start time of the edge (A, B) , $B \in V_1^p(A)$ for any $p \in \mathbb{R}^{|E|}$ since the contact network is fixed. Thus, $A = B_0 \rightarrow B_1 = B$ is a valid path in time set p_2 , and therefore $B \in V_1^{p_2}(A)$.

Base Case: $i = 2$

Want to show that if $B \in V_1^{p_1}(A)$ then $B \in V_2^{p_2}(A)$. Let $B \in V_2^{p_1}(A)$, therefore there exists a path $A = B_0 \rightarrow B_1 \rightarrow B_2 = B$ in time set p_1 . There are four possible values, $\{\cap, <, >, \cup\}$ for $f((B_0, B_1), (B_1, B_2))$. Below we go through each of these possible mappings and figure 6 depicts the first three values, $\{\cap, <, >\}$, as the fourth is the union of the previous three.

I: $f((B_0, B_1), (B_1, B_2)) = '\cap'$. Regardless of the timing of overlap, $B \in V_2^{p_2}(A)$.

II: $f((B_0, B_1), (B_1, B_2)) = '<'$. Therefore $I_A^{p_2}(B_1) < s^{p_2}(B_1, B_2)$. So, $B_2 = B$ is able to be infected by B_1 at time $s^{p_2}(B_1, B_2)$.

III: $f((B_0, B_1), (B_1, B_2)) = '>'$. It is not possible for B_1 to infect B_2 because $e(B_1, B_2) < I_A^{p_l}(B_1)$ for $l \in \{1, 2\}$. So this case is not valid for p_1 or p_2 .

IV: $f((B_0, B_1), (B_1, B_2)) = '\cup'$. Since, edges (B_0, B_1) and (B_1, B_2) share a common individual, B_1 , $((B_0, B_1), (B_1, B_2)) \in E_L$. Therefore, $f((B_0, B_1), (B_1, B_2))$ maps to $'\cap'$, $'<'$, or $'>'$. Hence $f((B_0, B_1), (B_1, B_2)) = '\cup'$ and this assignment is not valid for p_1 nor p_2 .

Induction Step: Assume true for j where $i > 2$

Assume if $B \in V_j^{p_1}(A)$ then $B \in V_j^{p_2}(A)$ for $j = i$. We want to show if $B \in V_{i+1}^{p_1}(A)$ then $B \in V_{i+1}^{p_2}(A)$. Let $B \in V_{i+1}^{p_1}(A)$, thus there exists a path $A = B_0 \rightarrow B_1, B_1 \rightarrow B_2, \dots, B_i \rightarrow B_{i+1} = B$ using time set p_1 . We will show that the path is also valid for time set p_2 by first looking that each of two possible scenarios: (A) there exists a j such that the edge $((B_{j-1}, B_j), (B_j, B_{j+1})) \in E_{no}$ and (B) no such j exists, i.e. $((B_{j-1}, B_j), (B_j, B_{j+1})) \in E_o$ for all $j \in \{1, \dots, i\}$. We know $((B_{j-1}, B_j), (B_j, B_{j+1})) \in E_{no}$ or $E_o \forall j$ since each pair of relationships share a common individual. Below we go through each of these two possible scenarios and figure 7 depicts each one.

A: $\exists j$ such that the edge $((B_{j-1}, B_j), (B_j, B_{j+1})) \in E_{no}$.

In time set p_1 there exists a path $A = B_0 \rightarrow B_1, B_1 \rightarrow B_2, \dots, B_i \rightarrow B_{i+1} = B$, thus $B_j \in V_j^{p_1}(A)$ and $B_{i+1} \in V_{i-j}^{p_1}(B_{j+1})$. By the induction assumption $B_j \in V_j^{p_2}(A)$ and $B_{i+1} \in V_{i-j}^{p_2}(B_{j+1})$ since $j, i-j < i$. It only has to be shown that $I_A^{p_2}(B_j) \leq s^{p_2}(B_j, B_{j+1})$, because then the epidemic can be broken into two paths. The first path is where A is initially infected and then $B_1, \dots, B_j$ are sequentially infected. The second path is where B_j is initially infected and then $B_{j+1}, \dots, B_{i+1}$ are sequentially infected. Therefore, if $I_A^{p_2}(B_j) \leq s^{p_2}(B_j, B_{j+1})$ then these two paths, both which are valid for p_2 by induction assumption, can be combined as shown in panel a of figure 7. Since $(B_{j-1}, B_j), (B_j, B_{j+1}) \in E_{no}, f((B_{j-1}, B_j), (B_j, B_{j+1})) \in \{<, >\}$. For the path $A = B_0 \rightarrow B_1, B_1 \rightarrow B_2, \dots, B_i \rightarrow B_{i+1} = B$ to be valid for time set p_1 , it must be the case that $f((B_{j-1}, B_j), (B_j, B_{j+1})) = <$. Hence, $I_A^{p_l}(B_j) \leq s(B_j, B_{j+1})$ for $l \in \{1, 2\}$.

B: $\nexists$ a j such that the edges $((B_{j-1}, B_j), (B_j, B_{j+1})) \in E_{no}$.

$f(((B_{j-1}, B_j), (B_j, B_{j+1}))) = '\cap'$ $\forall j$ since $\nexists$ a j such that the edges $((B_{j-1}, B_j), (B_j, B_{j+1})) \in E_{no}$ and $((B_{j-1}, B_j)$ and $(B_j, B_{j+1}))$ have a common individual. Define the following sequence of pairs of pairs, $S = (((B_{i-1}, B_i), (B_i, B_{i+1})), ((B_{i-2}, B_{i-1}), (B_{i-1}, B_i)), ((B_{i-2}, B_{i-1}), (B_i, B_{i+1}))), \dots, ((B_0, B_1), (B_i, B_{i+1}))$. Let $E^* = ((B_j, B_{j+1}), (B_k, B_{k+1}))$ be the first element where $f((B_j, B_{j+1}), (B_k, B_{k+1})) = '\cap'$ if one exists, otherwise let $E^* = ((B_0, B_1), (B_i, B_{i+1}))$. Table 2 shows the order of the sequence of pairs of pairs, where the entry $((j, i), (k, l))$ in the table represents the location of $((B_j, B_i), (B_k, B_l))$ in S . Note that table 2 is symmetric, and therefore only the upper triangle entries are displayed.

Below we go through each of the possible mappings for $f(E^*)$ and figure 8 depicts three of them, $\{\cap, <, >\}$, as the fourth is the union of these three.

I: $f(E^*) = '\cap'$. This is only possible if $E^* = ((B_0, B_1), (B_i, B_{i+1}))$. Therefore, if

$$\begin{aligned} \cap_{j=0}^{j=i} [s(B_j, B_{j+1}), e(B_j, B_{j+1})] &\neq \emptyset \text{ then } B_{i+1} \in V_{i+1}^{p_2}(A), \text{ because any point} \\ z \in \cap_{j=0}^{j=i} [s^{p_2}(B_j, B_{j+1}), e^{p_2}(B_j, B_{j+1})] &\text{ allows for the instantaneous potential spread of} \\ \text{infection from } B_0 \text{ to } B_{i+1}. \text{ Since the intervals are continuous,} \\ \cap_{j=0}^{j=i} [s(B_j, B_{j+1}), e(B_j, B_{j+1})] &= [\max(s^{p_2}(B_j, B_{j+1}), \min(e^{p_2}(B_k, B_{k+1}))) \text{ for } j, k \in \\ \{0, i\}. \text{ Additionally, since all pairs of relationships overlap,} \\ s^{p_2}(B_j, B_{j+1}) < \min(e^{p_2}(B_k, B_{k+1})) &\text{ for all } j, k \in \{0, i\}. \text{ Therefore,} \\ \cap_{j=0}^{j=i} [s(B_j, B_{j+1}), e(B_j, B_{j+1})] &\neq \emptyset. \text{ Thus, } B_{i+1} \in V_{i+1}^{p_2}(A). \end{aligned}$$

II: $f(E^*) = '<'$. Let $E^* = ((B_j, B_{j+1}), (B_k, B_{k+1}))$. We follow similar logic as section A, i.e. break the epidemic into two paths. The two paths are $B_0 \rightarrow B_1, \dots, B_{k-1} \rightarrow B_k$ and

$B_k \rightarrow B_{k+1}, \dots, B_i \rightarrow B_{i+1}$. By induction assumption $B_k \in V_k^{p_2}(A)$ and

$B_{i+1} \in V_{i-k}^{p_2}(B_{k+1})$ since $k, i-k < i$. It only has to be shown that

$I_A^{p_2}(B_k) \leq s^{p_2}(B_k, B_{k+1})$. By similar logic as I (directly above), we get

$$\begin{aligned} \cap_{j=j}^{j=k-1} [s(B_j, B_{j+1}), e(B_j, B_{j+1})] &\neq \emptyset \text{ since all pairs of relationships in } \{((B_n, B_{n+1}), \\ (B_m, B_{m+1}))\}_{n,m \in \{j, \dots, k-1\}} &\text{ overlap. Therefore, there exists a time} \\ z \in \cap_{j=j}^{j=k-1} [s(B_j, B_{j+1}), e(B_j, B_{j+1})] &\text{ where instantaneously } B_0 \text{ through } B_k \text{ can be} \end{aligned}$$

infected. Since

$z \in \cap_{j=j}^{j=k-1} [s(B_j, B_{j+1}), e(B_j, B_{j+1})] \subset [s^{p_2}(B_j, B_{j+1}), e^{p_2}(B_j, B_{j+1})]$ implies that $I_A^{p_2}(B_k) \in [s^{p_2}(B_j, B_{j+1}), e^{p_2}(B_j, B_{j+1})]$. Since $e^{p_2}(B_j, B_{j+1}) < s^{p_2}(B_k, B_{k+1})$ by the assumption that $f(E^*) = '<'$, we get $I_A^{p_2}(B_k) < s^{p_2}(B_k, B_{k+1})$.

III: $f(E^*) = '>'$. Since $k > j$ we know

$I_A^{p_1}(B_{k+1}) > I_A^{p_1}(B_{j+1}) > s(B_j, B_{j+1}) > e(B_k, B_{k+1})$ which is not possible because $I_A^{p_1}(B_{k+1}) \in [s^{p_1}(B_k, B_{k+1}), e^{p_1}(B_k, B_{k+1})]$. Therefore, $f(E^*) = '>'$ is not a possibility for time-set p_1 .

IV: $f(E^*) = '\cup'$. Since the relationships $((B_{j-1}, B_j)$ and $(B_j, B_{j+1}))$ have a common individual and $((B_{j-1}, B_j), (B_j, B_{j+1})) \notin E_{no}$ we know that $f((B_{j-1}, B_j), (B_j, B_{j+1})) = \cap \forall j$. Therefore $(B_0, B_1), (B_1, B_2), \dots, (B_i, B_{i+1})$ exist in the same connected component of $SL^H(G)$; denote this component as $SL_0^H(G)$. Since $K_{SL_0^H(G)} \subset R$, all of the designations between the pair of edges exist in $\{<, >, \cap\}$. Therefore, $f(E^*) = '\cup'$.

This completes the induction step. Therefore, $V_i^{p_1}(A) = V_i^{p_2}(A) \forall A \in V$ and $\forall i \in \mathbb{Z}^+$. So, $V^{p_1}(A) = V^{p_2}(A) \forall A \in V$.

References

- Bárány I, Füredi Z. Computing the volume is difficult. *Discrete and Computational Geometry*. 1987; 2:319–326.
- Blitzstein J, Diaconis P. A sequential importance sampling algorithm for generating random graphs with prescribed degrees. *Internet Mathematics*. 2010; 6(4):487–520.
- Carnegie NB, Morris M. Size matters: Concurrency and the epidemic potential of hiv in small networks. *PloS one*. 2012; 7(8):e43048. [PubMed: 22937011]
- Chung F, Lu L. Connected components in random graphs with given expected degree sequences. *Annals of Combinatorics*. 2002; 6:125–145. URL <http://dx.doi.org/10.1007/PL00012580>. 10.1007/PL00012580
- Doherty IA, Shiboski S, Ellen JM, Adimora AA, Padian NS. Sexual bridging socially and over time: A simulation model exploring the relative effects of mixing and concurrency on viral sexually transmitted infection transmission. *Journal of the International AIDS Society*. 2006; 33(6):368–373.
- Dyer M, Frieze A, Kannan R. A random polynomial-time algorithm for approximating the volume of convex bodies. *J ACM*. Jan; 1991 38(1):1–17. URL <http://doi.acm.org/10.1145/102782.102783>.
- Eaton JW, Hallett TB, Garnett GP. Concurrent sexual partnerships and primary hiv infection: A critical interaction. *AIDS and Behavior*. 2011; 15(4):687–692. [PubMed: 20890654]
- Eaton JW, Johnson LF, Salomon JA, Bärnighausen T, Bendavid E, Bershteyn A, Bloom DE, Cambiano V, Fraser C, Hontelez JAC, Humair S, Klein DJ, Long EF, Phillips AN, Pretorius C, Stover J, Wenger EA, Williams BG, Hallett TB. Hiv treatment as prevention: Systematic comparison of mathematical models of the potential impact of antiretroviral therapy on hiv incidence in south africa. *PLoS Med*. Jul.2012 9(7):e1001245. [PubMed: 22802730]
- Fishburn, P. Interval orders and interval graphs: a study of partially ordered sets. Wiley; 1985. Wiley-Interscience series in discrete mathematics
- Ghani AC, Garnett G. Risks of acquiring and transmitting sexually transmitted diseases in sexual partner networks. *Sexually Transmitted Diseases*. 2000; 27(10):579–587. [PubMed: 11099073]
- Ghani AC, Swinton J, Garnett G. The role of sexual partnership networks in the epidemiology of gonorrhea. *Sexually Transmitted Diseases*. 1997; 24(1):45–56. [PubMed: 9018783]

- Goodreau S. A decade of modelling research yields considerable evidence for the importance of concurrency: a response to sawers and still-waggon. *Journal of the International AIDS Society*. 2011; 14:12. [PubMed: 21406079]
- Harary, F. *Graph Theory*. Addison-Wesley; Reading, MA: 1969.
- Helleringer S, Kohler HP. Sexual network structure and the spread of hiv in africa: evidence from likoma island, malawi. *AIDS*. 2007; 21(17):2323–2332. [PubMed: 18090281]
- Konde-Lule JK, Wawer MJ, Sewankambo NK, Serwadda D, Kelly R, Li C, Gray RH, Kigongo D. Adolescents, sexual behavior and hiv-1 in rural rakai district, uganda. *AIDS*. 1997; 11(6):791–799. [PubMed: 9143612]
- Kretzschmar M, Morris M. Measures of concurrency in networks and the spread of infectious disease. *Mathematical Biosciences*. 1996; 133:165–195. [PubMed: 8718707]
- Lurie M, Rosenthal S. Concurrent partnerships as a driver of the hiv epidemic in sub-saharan africa? the evidence is limited. *AIDS and Behavior*. 2010; 14:17–24. URL <http://dx.doi.org/10.1007/s10461-009-9583-5>. 10.1007/s10461-009-9583-5 [PubMed: 19488848]
- Morris M. Barking up the wrong evidence tree. comment on lurie and rosenthal, concurrent partnerships as a driver of the hiv epidemic in sub-saharan africa? the evidence is limited. *AIDS and Behavior*. 2010; 14:31–33. URL <http://dx.doi.org/10.1007/s10461-009-9639-6>. 10.1007/s10461-009-9639-6 [PubMed: 19997971]
- Morris M, Epstein H. Role of concurrency in generalised hiv epidemics. *Lancet*. 2011; 378(9806):1843. [PubMed: 22118438]
- Morris, M.; Goodreau, S.; Moody, J. Sexual networks, concurrency, and std/hiv. In: Holmes, K.; PF, S.; WE, S., editors. *Sexually Transmitted Diseases*. 4th. McGraw-Hill International Book Co; New York, NY, USA: 2007. p. 109-126.
- Morris M, Kretzschmar M. Concurrent partnerships and the spread of hiv. *AIDS*. 1997; 11:641–648. [PubMed: 9108946]
- Morris M, Kurth A, Hamilton D, Moody J, Wakefield S. Concurrent partnerships and hiv prevalence disparities by race: Linking science and public health practice. *American Journal of Public Health*. 2009; 99(6):1023–1031. [PubMed: 19372508]
- Newman M. Assortative mixing in networks. *Physical Review Letters*. 2002; 89(20):208701. [PubMed: 12443515]
- Newman, ME. *Networks An Introduction*. Oxford University Press; New York: 2010.
- Palombi L, Bernava GM, Nucita A, Giglio P, Liotta G, Nielsen-Saines K, Orlando S, Mancinelli S, Buonomo E, Scarcella P, Altan AMD, Guidotti G, Ceffa S, Haswell J, Zimba I, Magid NA, Marazzi MC. Predicting trends in hiv-1 sexual transmission in sub-saharan africa through the drug resource enhancement against aids and malnutrition model: Antiretrovirals for reduction of population infectivity, incidence and prevalence at the district level. *Clinical Infectious Diseases*. 2012
- Pe'er, I.; Shamir, R. Interval graphs with side (and size) constraints. *Proc of the Third Annual European Symp on Algorithms, (ESA 95) Corfu; Greece*. Springer; 1995. p. 142-154.
- Rosenblatt M. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*. 1956; 27(10):832–837.
- Sawers L, Isaac A, Stillwaggon E. Hiv and concurrent sexual partnerships: modelling the role of coital dilution. *Journal of the International AIDS Society*. 2011; 14(1) URL <http://www.jiasociety.org/index.php/jias/article/view/17472>.
- Sawers L, Stillwaggon E. Concurrent sexual partnerships do not explain the hiv epidemics in africa: a systematic review of the evidence. *Journal of the International AIDS Society*. 2010; 13:34. [PubMed: 20836882]
- Smith RL. Efficient monte carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research*. 1984; 32(6):1296–1308.
- Tanser F, Bärnighausen T, Hund L, Garnett GP, McGrath N, Newell ML. Effect of concurrent sexual partnerships on rate of new hiv infections in a high-prevalence, rural south african population: a cohort study. *The Lancet*. 2011; 378(9787):247–255.

Highlights

- Method to sample dynamic networks with constraints on temporal network properties
- Permits isolation of the incremental impact of network features on network processes
- Aid in the design and interpretation of studies of network-based interventions

Figure 1. Three dynamic networks: the first without concurrency and the second and third with concurrency. The left panel represents the cumulative contact network. The five relationships in the network are given start and end times in the other panels. In the second panel there are no concurrent partnerships. The last two dynamic networks have identical values for concurrency measurements C_1 , C_2 , and C_3 . Both of the last two dynamic networks have two individuals, A and D, participating in concurrent partnerships, two pairs of relationships that overlap, and a total amount of overlap of one time unit. In the third panel, relationships AC and AD overlap as do relationships AD and DF. In the fourth panel, relationships AB and AC overlap as do relationships DE and DF. The start times and largest reachable path is provided for each simulation.

Figure 2.

An illustration showing the relationship among G , $L(G)$, H_Ω , and S for two sets of concurrency values $C_1 = 1$, $C_2 = 1$, $C_3 = .6$ and $C_1 = C_2 = C_3 = 0$. Concurrency Present Example: H_Ω is the green area labeled H_1 and X_1 is the union of the two red lines ($y = x + .4$ and $y = x - .4$). Concurrency Absent Example: H_Ω is the union of H_1 (green area) and H_2 (blue area). In this example $X_1 = H_1$ and $X_2 = H_2$.

Figure 3.
An illustration showing the three restrictive time-order designations.

Figure 4.

In both dynamic networks, the relationships E_{AB} and E_{BC} overlap as do E_{BC} and E_{CD} . Therefore, edges (E_{AB}, E_{BC}) and (E_{BC}, E_{CD}) are contained in the same connected component of $SL(G)$. In the first panel, if A is initially infected then B , C and D can potentially be infected; however in the second panel, only B and C have the potential of being infected. Hence the need to designate the time order of relationships E_{AB} and E_{CD} .

Figure 5.

Relative importance of concurrency controlling for other network properties (x-axis) on the largest reachable path (y-axis). Network properties are (A) degree assortativity, (B) clustering coefficient, and (C) mixing between two subpopulations.

Figure 6.

An illustration of the first three possible mappings for $f((B_0, B_1), (B_1, B_2))$. Panel (a) depicts an example of $f((B_0, B_1), (B_1, B_2)) = '\cap'$; Panel (b) depicts an example of $f((B_0, B_1), (B_1, B_2)) = '<'$; Panel (c) depicts an example of $f((B_0, B_1), (B_1, B_2)) = '>'$.

Figure 7.

An illustration of the two possible scenarios. Panel (a) depicts an example where at least one adjacent pair of relationships do not overlap; Panel (b) depicts an example where all adjacent pairs of relationships overlap;.

Figure 8.

An illustration of the three possible sub-cases for the initial case that all adjacent pairs of edges in the path overlap. Panel (a) depicts an example of $f(E^*) = \cap$; Panel (b) depicts an example of $f(E^*) = <$; Panel (c) depicts an example of $f(E^*) = >$.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Distribution of Number of Partners

Degree Distribution						
Degree	1	2	3	4	5	6
Percentage of Individuals	66	18	8	4	2	2

Table 2

Sequence Order

Edges				
	$(0, 1)$	$(1, 2)$	$\dots$	$(i - 2, i - 1)$
$(1, 2)$	$-$	$\frac{i * (i + 1)}{2} - (i - 1)$	$\frac{i * (i + 1)}{2} - 2$	$\frac{i * (i + 1)}{2} - 1$
$(1, 2)$			$-$	$\frac{i * (i + 1)}{2} - (i + 1)$
$\vdots$				$\frac{i * (i + 1)}{2} - i$
$(i - 2, i - 1)$				3
$(i - 1, i)$				1
$(i, i + 1)$				$-$