The Annals of Applied Statistics
2016, Vol. 10, No. 4, 2021–2046
DOI: 10.1214/16-AOAS963

MODELING CONCURRENCY AND SELECTIVE MIXING IN HETEROSEXUAL PARTNERSHIP NETWORKS WITH APPLICATIONS TO SEXUALLY TRANSMITTED DISEASES[1]

BY RYAN ADMIRAAL AND MARK S. HANDCOCK

Murdoch University and University of California, Los Angeles

Network-based models for sexually transmitted disease transmission rely on initial partnership networks incorporating structures that may be related to risk of infection. In particular, initial networks should reflect the level of concurrency and attribute-based selective mixing observed in the population of interest. We consider momentary degree distributions as measures of concurrency and propensities for people of certain types to form partnerships with each other as a measure of attribute-based selective mixing. Estimation of momentary degree distributions and mixing patterns typically relies on cross-sectional survey data, and, in the context of heterosexual networks, we describe how this results in two sets of reports that need not be consistent with each other. The reported momentary degree distributions and mixing totals are related through a series of constraints, however. We provide a method to incorporate those in jointly estimating momentary degree distributions and mixing totals. We develop a method to simulate heterosexual networks consistent with these momentary degree distributions and mixing totals, applying it to data obtained from the National Longitudinal Study of Adolescent Health. We first use the momentary degree distributions and mixing totals as mean value parameters to estimate the natural parameters for an exponential-family random graph model and then use a Markov chain Monte Carlo algorithm to simulate person-level heterosexual partnership networks.

1. Introduction. In its most recent estimates of human immunodeficiency virus (HIV) prevalence (or total number of infected individuals at a given point in time), the World Health Organization (WHO) estimated that the number of people worldwide infected with HIV had reached 35 million by the end of 2013, with approximately 70% of infected individuals living in Africa and the overwhelming majority of cases in sub-Saharan Africa [World Health Organization (2013)]. Despite the pervasiveness of infections in the Africa region, there is significant variability in HIV prevalence by country, as demonstrated by Demographic and Health Surveys (DHS) estimates by country presented in the Supplementary Ma-

Received November 2013; revised June 2016.

[1]Supported by R21HD075714 and the NSF Grant SES-1357619. No direct support was received from Grant P01-HD31921 for this analysis.

Key words and phrases. Heterosexual partnership networks, exponential-family random graph models, constrained maximum likelihood estimation, National Longitudinal Survey of Adolescent Health.

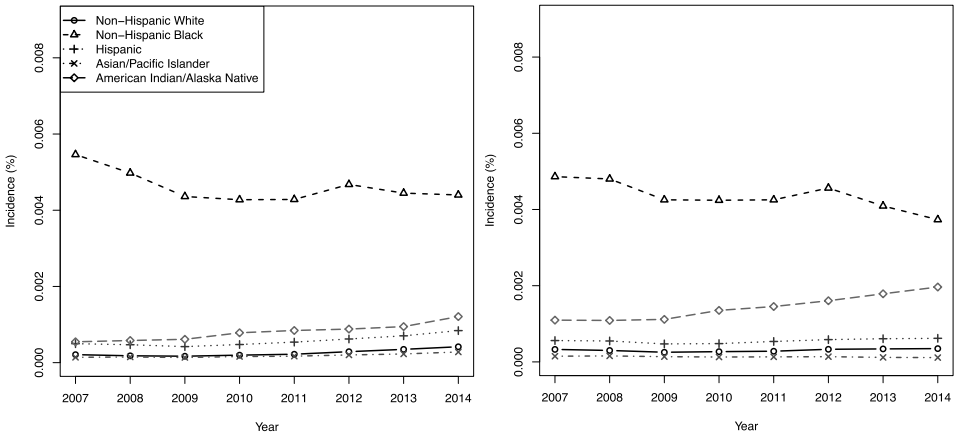

FIG. 1. *Gonorrhea incidence by year for males* (*left*) *and females* (*right*) *in the United States*. (*From* 2012–2014 *the "Asian/Pacific Islander" categorization was no longer used, so totals for that period strictly represent the "Asian" category.*) *Source*: *Centers for Disease Control and Prevention* (2015).

terial, Figure S1, for a subset of sub-Saharan Africa countries [Admiraal and Handcock (2016), Demographic and Health Surveys Program (2015)]. Similar disparities in the incidence (or new cases) of HIV and other sexually transmitted infections (STIs) exist across races represented in the United States. For example, Figure 1 presents yearly incidence of Gonorrhea by race from 2007–2014, as reported by the Centers for Disease Control and Prevention (2015). This shows significantly higher incidence for non-Hispanic blacks than other races.

The variability we observe in HIV prevalence across countries in sub-Saharan Africa and in STI incidence for different races in the United States may be related to biological factors, such as genetics, or behavioral factors, such as condom use. However, it may also be related to the structure of the network of partnerships in the population of interest. In particular, differing levels of concurrent partnerships (or partnerships overlapping in time) and selective mixing (or tendencies for specific subpopulations to come into contact with each other) may be important in explaining disparities.

In this paper, we first explain how selective mixing and concurrent partnerships (or "concurrency") may be important in producing differential disease prevalence according to key characteristics. (In the Supplementary Material, Section 1, we provide a review of common measures of concurrency, studies supporting the importance of concurrency in explaining STI prevalence, and critiques to commonly employed modeling approaches [Admiraal and Handcock (2016)].) We propose a method to simultaneously estimate measures of concurrency and selective mixing for heterosexual partnerships. This method is applicable to cross-sectional survey data, meaning that it can be readily applied to the most common forms of sexual partnership data that are increasingly being collected in surveys such as the

DHS. We explain how this method allows for greater flexibility in disease models, addressing many of the critiques to current modeling approaches, and we demonstrate how to generate heterosexual partnership networks consistent with measures of concurrency and selective mixing. These networks can serve as initial networks (or seeds) for network-based simulations for STI spread through heterosexual populations. Finally, we apply our method to data from the National Longitudinal Study of Adolescent Health.

2. Selective mixing and concurrency. One posited explanation for the higher prevalence of HIV in sub-Saharan Africa (and particularly Southern Africa) than other regions of the world is higher prevalence of concurrency [Morris, Epstein and Wawer (2010)]. This hypothesis, known as the "concurrency hypothesis," is predicated on the assumption that HIV is transmitted more quickly through concurrent partnerships than serial monogamy (or nonoverlapping sequential partnerships). Where concurrency relates to the rate of disease spread, selective mixing corresponds to transmission of the disease between subpopulations. Limited transmission between subpopulations can potentially allow a disease to propagate within subpopulations at different rates, producing different disease prevalence within subpopulations. Thus, both concurrency and selective mixing are important considerations when examining disparities in disease prevalence between subpopulations, whether these subpopulations be world regions, countries, sexes, races, etc.

2.1. *Selective mixing*. Selective mixing affects the level of connectivity among various subpopulations and can be instrumental in explaining how a disease may spread at different rates through subpopulations. In particular, strong assortative mixing (or preference for partnerships with those of similar characteristics) can effectively partition a population, limiting disease transmission from one subpopulation to another and allowing for different rates of disease spread within subpopulations when other factors influencing disease prevalence are different for these subpopulations. Conversely, if mixing is purely random (i.e., nonpreferential), then there will be numerous opportunities for disease transmission among subpopulations, leading to only minor differences in disease prevalence when other factors influencing prevalence are similar. Selective mixing is commonly represented by a mixing matrix (or contact matrix), where elements of the matrix give the number of partnerships between subpopulations. We refer to the elements of a mixing matrix as mixing totals.

A number of studies have considered the importance of selective mixing in explaining disparities in disease prevalence for subpopulations. Many of these studies focus on selective mixing based on level of sexual activity [Anderson (1992), Anderson, Gupta and Ng (1990), Busenberg and Castillo-Chavez (1989), Castillo-Chavez and Blythe (1989), Gupta, Anderson and May (1989), Jacquez, Simon and Koopman (1989), Garnett and Anderson (1993a, 1993b), Aral et al. (1999), Chick, Adams and Koopman (2000), Doherty et al. (2006), Garnett et al. (1996),

Ghani, Swinton and Garnett (1997), Morin et al. (2014)]. This is because much of the early incentive for considering selective mixing centered on the idea of a core group characterized by high levels of sexual activity (e.g., sex trade workers and their clients) being the primary agents of spread [Garnett and Anderson (1993a)]. Because the pervasiveness of the disease was then dependent on the frequency of partnerships between people with high levels of sexual activity and people with lower levels of sexual activity, selective mixing based on level of sexual activity was important in explaining disease spread.

Morris (1997) argues that the evidence supporting core groups as the primary mechanism for concentrated epidemics is strong, but this does not seem to be the primary mechanism for generalized epidemics (defined as greater than 1% prevalence in the entire population). Many sub-Saharan African countries have been experiencing generalized epidemics of HIV where new infections are no longer primarily the result of partnerships between people with high sexual activity levels and people with lower activity levels, and so the core group is no longer the driving force behind the epidemic. Consequently, a number of studies have considered attribute-based mixing (e.g., race, age, education, disease status) [Hyman and Stanley (1988), Busenberg and Castillo-Chavez (1989), Morris (1991, 1994, 1995), Aral et al. (1999), Morris et al. (2009), Morin et al. (2014), Hamilton and Morris (2015)]. Morris et al. (2009) and Hamilton and Morris (2015) assert that strong assortative mixing by race, coupled with different levels of concurrency, is a plausible explanation for the disparities we observe in HIV (and other STI) prevalence across different races in the United States. If we consider sub-Saharan Africa, few sexual partnerships occurring between individuals from different countries would allow for different rates of HIV spread within individual countries, possibly partially explaining the heterogeneity in HIV prevalence we observe.

2.2. *Concurrency.* Where selective mixing relates to connectivity of subpopulations and disease transmission between subpopulations, concurrency relates to connectivity of individuals and disease spread between individuals, and a number of studies have claimed that concurrency is important in explaining disparities in STI prevalence and the speed with which STIs propagate through a population [Watts and May (1992), Hudson (1993), Morris and Kretzschmar (1995, 1997, 2000), Kretzschmar and Morris (1996), Ghani, Swinton and Garnett (1997), Chick, Adams and Koopman (2000), Ghani and Garnett (2000), Koopman et al. (2000), Adimora and Schoenbach (2002, 2005), Adimora, Schoenbach and Doherty (2006, 2007), Doherty et al. (2006), Morris, Goodreau and Moody (2007), Morris et al. (2009), Johnson et al. (2009), Mah and Halperin (2010), Eaton, Hallett and Garnett (2011), Goodreau (2011), Goodreau et al. (2012), Kretzschmar and Caraël (2012), Hamilton and Morris (2015)]. If we consider a population under serial monogamy, current relationships must end before infected individuals place other noninfected individuals at risk. With concurrency, on the other hand, if an individual has concurrent sexual partners and becomes infected by one, the

other partners are placed at risk almost immediately. Consequently, greater connectivity of a population that is possible under higher levels of concurrency should theoretically provide a much more efficient mechanism for disease spread.

Historically, concurrency has been measured "directly" through individuals' self-reports of multiple partnerships being current and sexual (referred to as current prevalence of concurrency) or "indirectly" through individuals' reported first and last sexual encounters with each partner over a prescribed period of time [Helleringer, Mkandawire and Kohler (2014)]. In the Supplementary Material, Section 1, we review commonly used concurrency measures, including the point prevalence of concurrency (or percentage of the sexually active population with concurrent partnerships at a given point in time) and the κ-statistic, which was first proposed by Kretzschmar and Morris (1996) and incorporates the number of concurrent partnerships people have at a given point in time [Admiraal and Handcock (2016)].

More recent modeling approaches have moved away from point estimators of concurrency and instead incorporate full momentary degree distributions (from now on simply referred to as "degree distributions"), the distribution of numbers of sexual partnerships for individuals at a given point in time [e.g., Carnegie and Morris (2011), Goodreau et al. (2012), Hamilton, Handcock and Morris (2008), Morris et al. (2009)]. Degree distributions preserve more of the epidemiologically relevant information than a single point estimator while being able to recover both the κ-statistic and point prevalence of concurrency. In particular, the proportion of the sexually active population with a momentary degree greater than 1 exactly matches the point prevalence of concurrency. At the same time, the κ-statistic is a function of the mean and variance of the degree distribution, and so the degree distribution can also recover the κ-statistic. Consequently, in this paper we use degree distributions rather than more commonly reported point estimators of concurrency.

In the Supplementary Material, Section 2, we review the literature on concurrency and its relationship to disease spread [Admiraal and Handcock (2016)]. We also review the various critiques of the concurrency hypothesis as well as critiques of standard modeling approaches presented by Sawers (2013). The method we propose addresses many of these critiques and provides a significant improvement to commonly employed modeling approaches on a number of levels. First, concurrency can be stratified according to a variety of factors, including sex, race, age or other characteristics. Second, our approach provides a rigorous mechanism to reconcile inconsistencies in reports from cross-sectional surveys when it is not known if reports corresponding to one subpopulation (e.g., males, females) are biased. [Although not demonstrated here, when reports of partnerships for a subpopulation (or subpopulations) are believed to be unbiased whereas reports for those in other subpopulations are biased, the unbiased reports can be used as constraints to improve estimates for these other subpopulations using approaches similar to those considered by Handcock, Rendall and Cheadle (2005) and Rendall et al. (2008).] Third, the sizes of subpopulations need not be the same. Indeed, our application

considers a heterosexual partnership network with heterogeneous compositions of men and women of different races, and it estimates sex- and race-specific degree distributions. Finally, our method can incorporate population-based estimates of both concurrency and selective mixing, thereby avoiding many of the simplifying assumptions (such as random or symmetric assortative mixing) that are required for disease models that incorporate both. This signals a significant step forward in the capabilities of network-based disease models to accurately reflect partnership networks for underlying populations.

3. Simulating networks with degree distributions and mixing totals consistent with underlying populations. In order for disease models to be effective in measuring the anticipated impacts of various interventions on STI spread, they should incorporate potential risk factors for the population of interest. This includes concurrency (as measured through degree distributions) and selective mixing, and so initial networks should be consistent with observed degree distributions and selective mixing totals for the population under consideration. To construct initial networks that incorporate degree distributions and selective mixing for the population of interest, we use a three part process, shown in Figure 2. First, we use survey data to estimate degree distributions and mixing totals for a given population. Mixing totals can be based on attributes such as age, race, education level and even sexual activity level, and degree distributions can be stratified similarly. Next, these estimated degree distributions and mixing totals are specified as the mean value parameters (or expected values of the sufficient statistics) for an exponential-family random graph model (ERGM). (When full network data or a cross-sectional census is available, observed degree distributions and mixing totals can be specified as the mean value parameters.) These mean value parameters are used to estimate corresponding natural (or canonical) parameters for the ERGM. These natural parameters are interpretable as conditional log-odds ratios. Finally, once the natural parameters are specified, the ERGM can be used to simulate networks consistent with the desired degree distributions and mixing totals.

Although STI transmission through homosexual partnerships is significant for some populations [Grulich and Zablotska (2010) note that HIV incidence in North America, Australia and Western Europe is highest for homosexual men], in this paper we restrict our focus to disease transmission through heterosexual partnerships, in line with many of the previously cited studies. Heterosexual partnership

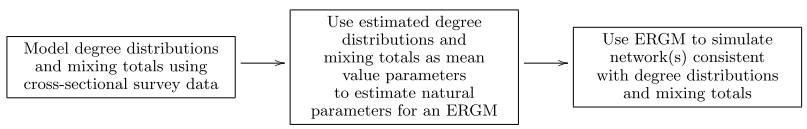

FIG. 2. *Procedure for simulating networks consistent with desired degree distributions and mixing totals.*

networks are bipartite in form, meaning that the population can be partitioned into two disjoint modes or types (in this case, males and females) with partnerships restricted to those including one individual of each type. At the same time, while dyad samples (or samples of pairs of individuals) have been used in some surveys of male and female reports on sexual activity [Julian et al. (1992), Kinsey, Pomeroy and Martin (1948), Ochs and Binik (1999), Seal (1997)], most surveys on sexual activity collect cross-sectional data of individuals and their reported partnerships (or egocentric network data, also referred to as index respondent data), and the methods we develop assume such data. When full network data is available, we note that the first step in Figure 2 is unnecessary, as degree distributions and mixing totals for the full network can be used as the mean value parameters.

3.1. *Modeling degree distributions and mixing totals.* Suppose a population consists of M males ($\mathcal{M}$) and F females ($\mathcal{F}$), the subpopulation of males consists of I different types of sizes $M_1, \ldots, M_I$ from which $m_1, \ldots, m_I$ males are sampled, and the subpopulation of females consists of J different types of sizes $F_1, \ldots, F_J$ from which $f_1, \ldots, f_J$ females are sampled.

3.1.1. *Degree distributions.* Let $D_{ik}^{\mathcal{M}}$ represent the number of males of type i with degree k, and let $D_{j\ell}^{\mathcal{F}}$ represents the number of females of type j with degree ℓ. If these population degree totals are a realization from some underlying stochastic process, then corresponding to these totals are expected (superpopulation) degree distributions. Denote the expected degree distribution for males of type i by $\pi_i^{\mathcal{M}} = \{\pi_{i0}^{\mathcal{M}}, \ldots, \pi_{iF}^{\mathcal{M}}\}$, $i = 1, \ldots, I$, where $\pi_{ik}^{\mathcal{M}}$ denotes the population proportion of males of type i with degree k; and denote the expected degree distribution for females of type j by $\pi_j^{\mathcal{F}} = \{\pi_{j0}^{\mathcal{F}}, \ldots, \pi_{jM}^{\mathcal{F}}\}$, $j = 1, \ldots, J$, where $\pi_{j\ell}^{\mathcal{F}}$ denotes the proportion of females of type j with degree ℓ. Then, under a cross-sectional study design and given M_i and F_j (the number of males of type i and the number of females of type j, respectively) and degree distributions $\pi_i^{\mathcal{M}}$ and $\pi_j^{\mathcal{F}}$ for males of type i and females of type j, $D_i^{\mathcal{M}} = (D_{i0}^{\mathcal{M}}, \ldots, D_{iF}^{\mathcal{M}})$ and $D_j^{\mathcal{F}} = (D_{j0}^{\mathcal{F}}, \ldots, D_{jM}^{\mathcal{F}})$ are modeled according to the multinomial distributions

$$\begin{aligned}
\big(D_{i0}^{\mathcal{M}}, \ldots, D_{iF}^{\mathcal{M}}\big)|\pi_i^{\mathcal{M}}, M_i &\sim \text{Multinomial}\big(M_i, \pi_{i0}^{\mathcal{M}}, \ldots, \pi_{iF}^{\mathcal{M}}\big),\\
\big(D_{j0}^{\mathcal{F}}, \ldots, D_{jM}^{\mathcal{F}}\big)|\pi_j^{\mathcal{F}}, F_j &\sim \text{Multinomial}\big(F_j, \pi_{j0}^{\mathcal{F}}, \ldots, \pi_{jM}^{\mathcal{F}}\big).
\end{aligned}$$

Unless a census is carried out, we do not obtain population degree totals $D_i^{\mathcal{M}}$ and $D_j^{\mathcal{F}}$ but instead obtain $d_i^{\mathcal{M}} = \{d_{i0}^{\mathcal{M}}, \ldots, d_{iF}^{\mathcal{M}}\}$ and $d_j^{\mathcal{F}} = \{d_{j0}^{\mathcal{F}}, \ldots, d_{jF}^{\mathcal{F}}\}$, the sample degree distributions for males of type i and females of type j. If these are obtained through a simple random sample of m_i males of type i and f_j females of type j, then these can be modeled as

$$\begin{aligned}
\big(d_{i0}^{\mathcal{M}}, \ldots, d_{iF}^{\mathcal{M}}\big)|\pi_i^{\mathcal{M}}, m_i &\sim \text{Multinomial}\big(m_i, \pi_{i0}^{\mathcal{M}}, \ldots, \pi_{iF}^{\mathcal{M}}\big),\\
\big(d_{j0}^{\mathcal{F}}, \ldots, d_{jM}^{\mathcal{F}}\big)|\pi_j^{\mathcal{F}}, f_j &\sim \text{Multinomial}\big(f_j, \pi_{j0}^{\mathcal{F}}, \ldots, \pi_{jM}^{\mathcal{F}}\big).
\end{aligned}$$

TABLE 1

Mixing matrix for males and females, stratified on I types for males and J types for females (left), and corresponding matrix of expected mixing totals (right)

		Females				
		1	2	$\cdots$	J	
Males	1	N_{11}	N_{12}	$\cdots$	N_{1J}	$N_{1\cdot}$
	2	N_{21}	N_{22}	$\cdots$	N_{2J}	$N_{2\cdot}$
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	I	N_{I1}	N_{I2}	$\cdots$	N_{IJ}	$N_{I\cdot}$
		$N_{\cdot 1}$	$N_{\cdot 2}$	$\cdots$	$N_{\cdot J}$	$N_{\cdot\cdot}$

		Females				
		1	2	$\cdots$	J	
Males	1	μ_{11}	μ_{12}	$\cdots$	μ_{1J}	$\mu_{1\cdot}$
	2	μ_{21}	μ_{22}	$\cdots$	μ_{2J}	$\mu_{2\cdot}$
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	I	μ_{I1}	μ_{I2}	$\cdots$	μ_{IJ}	$\mu_{I\cdot}$
		$\mu_{\cdot 1}$	$\mu_{\cdot 2}$	$\cdots$	$\mu_{\cdot J}$	$\mu_{\cdot\cdot}$

3.1.2. *Mixing totals*. Selective mixing is most commonly represented by a mixing matrix which breaks down the total number of partnerships between individuals of different types. For a heterosexual population, this matrix takes on the form of the matrix to the left in Table 1, where N_{ij} represents the number of partnerships between males of type i, $i = 1, \ldots, I$, and females of type j, $j = 1, \ldots, J$, and the "$\cdot$" notation represents summation over a particular index. This table is conditional on the presence of a tie, and so it does not provide information about the number of nonpartnerships between men and women of given types. If the observed population mixing totals are a realization from some underlying stochastic process, then corresponding to the mixing matrix is a matrix of expected mixing totals $\mu = (\mu_{ij})_{1\le i\le I, 1\le j\le J}$, represented by the matrix to the right in Table 1.

The expected mixing totals μ must be consistent with the expected degree distributions. In particular, the total expected number of partnerships for males of type i is given by

$$\mu_{i\cdot} = \sum_{j=1}^{J} \mu_{ij} = \sum_{k=0}^{F} k\pi_{ik}^{\mathcal{M}} M_i, \tag{1}$$

where $\sum_{k=0}^{F} k\pi_{ik}^{\mathcal{M}}$ gives the mean number of partners for males of type i. Similarly, for females of type j,

$$\mu_{\cdot j} = \sum_{\ell=0}^{M} \ell\pi_{j\ell}^{\mathcal{F}} F_j, \tag{2}$$

and the total expected number of partnerships must satisfy

$$\begin{aligned} \mu_{\cdot\cdot} &= \sum_{i=1}^{I}\left(\sum_{k=0}^{F} k\pi_{ik}^{\mathcal{M}} M_i\right) = \sum_{j=1}^{J}\left(\sum_{\ell=0}^{M} \ell\pi_{j\ell}^{\mathcal{F}} F_j\right) \\ &= \frac{1}{2}\left(\sum_{i=1}^{I}\sum_{k=0}^{F} k\pi_{ik}^{\mathcal{M}} M_i\right) + \frac{1}{2}\left(\sum_{j=1}^{J}\sum_{\ell=0}^{M} \ell\pi_{j\ell}^{\mathcal{F}} F_j\right). \end{aligned} \tag{3}$$

Thus, the expected mixing totals should be modeled in terms of the expected degree distributions to ensure consistency between the expected mixing totals and expected degree distributions.

There are a number of different ways that we could consider modeling expected mixing totals, but, given the constraints specified by (1), (2) and (3), we model μ_{ij} as

$$\begin{aligned}\mu_{ij} &= \frac{\mu_{i\cdot}\mu_{\cdot j}}{\mu_{\cdot\cdot}}\alpha_{ij} \\ &= \frac{2(\sum_{k=0}^{F} k\pi_{ik}^{\mathcal{M}} M_i)(\sum_{\ell=0}^{M} \ell\pi_{j\ell}^{\mathcal{F}} F_j)}{(\sum_{i=1}^{I}\sum_{k=0}^{F} k\pi_{ik}^{\mathcal{M}} M_i) + (\sum_{j=1}^{J}\sum_{\ell=0}^{M} \ell\pi_{j\ell}^{\mathcal{F}} F_j)}\alpha_{ij}.\end{aligned} \tag{4}$$

If $\alpha_{ij} = 1$ for all i and j, this is the standard independence model where cells of a contingency table are specified in terms of the marginals. Thus, α_{ij} represents the cell-specific dependence between males of type i and females of type j. In the context of mixing matrices, this relates to propensities for partnerships between males of type i and females of type j beyond what would be expected under random mixing, and so these are homophily (or assortative) effects if $i = j$ and heterophily (or disassortative) effects if $i \neq j$. The Supplementary Material, Section 3, describes several common models for these parameters [Admiraal and Handcock (2016)].

Note that the dependence parameters $\alpha = \{\alpha_{11}, \ldots, \alpha_{IJ}\}$ must satisfy certain constraints. In particular, $\mu_{i\cdot} = \sum_{j=1}^{J} \mu_{ij}$ and $\mu_{\cdot j} = \sum_{i=1}^{I} \mu_{ij}$, and so, using (1), (2) and (4), we obtain

$$\begin{aligned}\sum_{k=0}^{F} k\pi_{ik}^{\mathcal{M}} M_i &= 2\sum_{j=1}^{J} \frac{(\sum_{k=0}^{F} k\pi_{ik}^{\mathcal{M}} M_i)(\sum_{\ell=0}^{M} \ell\pi_{j\ell}^{\mathcal{F}} F_j)}{(\sum_{i=1}^{I}\sum_{k=0}^{F} k\pi_{ik}^{\mathcal{M}} M_i) + (\sum_{j=1}^{J}\sum_{\ell=0}^{M} \ell\pi_{j\ell}^{\mathcal{F}} F_j)}\alpha_{ij}, \\ \sum_{\ell=0}^{M} \ell\pi_{j\ell}^{\mathcal{F}} F_j &= 2\sum_{i=1}^{I} \frac{(\sum_{k=0}^{F} k\pi_{ik}^{\mathcal{M}} M_i)(\sum_{\ell=0}^{M} \ell\pi_{j\ell}^{\mathcal{F}} F_j)}{(\sum_{i=1}^{I}\sum_{k=0}^{F} k\pi_{ik}^{\mathcal{M}} M_i) + (\sum_{j=1}^{J}\sum_{\ell=0}^{M} \ell\pi_{j\ell}^{\mathcal{F}} F_j)}\alpha_{ij}.\end{aligned} \tag{5}$$

Now, with cross-sectional sampling, we do not observe $N = \{N_{11}, \ldots, N_{IJ}\}$. Nominated partners are likely to fall outside the sample, and so the mixing totals reported by males need not match those given by females, and this results in separate mixing matrices $n^{\mathcal{M}} = \{n_{11}^{\mathcal{M}}, \ldots, n_{IJ}^{\mathcal{M}}\}$ and $n^{\mathcal{F}} = \{n_{11}^{\mathcal{F}}, \ldots, n_{IJ}^{\mathcal{F}}\}$ from the male reports and female reports, respectively. Assume that m_i of the M_i males of type i and f_j of the F_j females of type j are randomly sampled. Then the expected mixing total for partnerships between males of type i and females of type j is given by

$$\mathbb{E}(N_{ij}) = \mu_{ij},$$

and the corresponding expected mixing totals from the samples of males and females are

$$\mathbb{E}\big(n_{ij}^{\mathcal{M}}\big) = \mathbb{E}\bigg(N_{ij}\frac{m_i}{M_i}\bigg) = \mu_{ij}\frac{m_i}{M_i}, \qquad \mathbb{E}\big(n_{ji}^{\mathcal{F}}\big) = \mu_{ij}\frac{f_j}{F_j}.$$

If we combine this with our model for expected mixing totals given by (4), we obtain

$$\mathbb{E}(n_{ij}^{\mathcal{M}}) = \frac{2(\sum_{k=0}^{F} k\pi_{ik}^{\mathcal{M}} m_i)(\sum_{\ell=0}^{M} \ell\pi_{j\ell}^{\mathcal{F}} F_j)}{(\sum_{i=1}^{I}\sum_{k=0}^{F} k\pi_{ik}^{\mathcal{M}} M_i) + (\sum_{j=1}^{J}\sum_{\ell=0}^{M} \ell\pi_{j\ell}^{\mathcal{F}} F_j)}\alpha_{ij},$$

$$\mathbb{E}(n_{ij}^{\mathcal{F}}) = \frac{2(\sum_{k=0}^{F} k\pi_{ik}^{\mathcal{M}} M_i)(\sum_{\ell=0}^{M} \ell\pi_{j\ell}^{\mathcal{F}} f_j)}{(\sum_{i=1}^{I}\sum_{k=0}^{F} k\pi_{ik}^{\mathcal{M}} M_i) + (\sum_{j=1}^{J}\sum_{\ell=0}^{M} \ell\pi_{j\ell}^{\mathcal{F}} F_j)}\alpha_{ij}.$$

Following the generative process considered by Morris (1991), we assume that the number of partnership opportunities for individuals is determined by a Poisson process, and the conditional distribution of the number of partnerships given a specific number of opportunities is binomial. Then the mixing totals can be shown to be distributed according to the Poisson distributions

$$n_{ij}^{\mathcal{M}} | \pi_i^{\mathcal{M}}, \pi_j^{\mathcal{F}}, \alpha_{ij}, m_i, M_i, f_j, F_j \sim \text{Poisson}\left(\frac{2(\sum_{k=0}^{F} k\pi_{ik}^{\mathcal{M}} m_i)(\sum_{\ell=0}^{M} \ell\pi_{j\ell}^{\mathcal{F}} F_j)}{(\sum_{i=1}^{I}\sum_{k=0}^{F} k\pi_{ik}^{\mathcal{M}} M_i) + (\sum_{j=1}^{J}\sum_{\ell=0}^{M} \ell\pi_{j\ell}^{\mathcal{F}} F_j)}\alpha_{ij}\right),$$

$$n_{ij}^{\mathcal{F}} | \pi_i^{\mathcal{M}}, \pi_j^{\mathcal{F}}, \alpha_{ij}, m_i, M_i, f_j, F_j \sim \text{Poisson}\left(\frac{2(\sum_{k=0}^{F} k\pi_{ik}^{\mathcal{M}} M_i)(\sum_{\ell=0}^{M} \ell\pi_{j\ell}^{\mathcal{F}} f_j)}{(\sum_{i=1}^{I}\sum_{k=0}^{F} k\pi_{ik}^{\mathcal{M}} M_i) + (\sum_{j=1}^{J}\sum_{\ell=0}^{M} \ell\pi_{j\ell}^{\mathcal{F}} F_j)}\alpha_{ij}\right).$$

3.2. *Maximum likelihood estimators for degree distribution and mixing parameters.* To jointly estimate the degree distributions $\pi^{\mathcal{M}} = \{\pi_1^{\mathcal{M}}, \ldots, \pi_I^{\mathcal{M}}\}$ and $\pi^{\mathcal{F}} = \{\pi_1^{\mathcal{F}}, \ldots, \pi_J^{\mathcal{F}}\}$ and mixing totals $\mu = (\mu_{ij})_{1\le i\le I, 1\le j\le J}$, we use the joint likelihood of $\pi^{\mathcal{M}}$, $\pi^{\mathcal{F}}$ and $\alpha = (\alpha_{ij})_{1\le i\le I, 1\le j\le J}$, which is shown in the Supplementary Material, Section S4 [Admiraal and Handcock (2016)]. Note that this likelihood must be maximized subject to the constraints

$$\sum_{k=0}^{F} \pi_{ik}^{\mathcal{M}} = 1, \qquad \sum_{\ell=0}^{M} \pi_{j\ell}^{\mathcal{F}} = 1 \tag{6}$$

for all i and j, the typical constraints on parameters for the multinomial distribution, as well as the constraints on α given by (5). Given that the likelihood must be maximized subject to these constraints, we use constrained maximum likelihood estimation. Specifically, we use the `Rsolnp` package [Ghalanos and Theussls (2012)], based on the SOLNP algorithm of Ye (1987), in `R` [R Core Team (2013)] to accomplish this. Once the joint distribution of $\pi^{\mathcal{M}}$, $\pi^{\mathcal{F}}$ and α is estimated, μ can be obtained using (4). To demonstrate that the estimation method is well behaved, we provide a simulation study in the Supplementary Material, Section 6.2 [Admiraal and Handcock (2016)].

3.3. *Estimating natural parameters from mean value parameters.* Once degree distributions and mixing totals are estimated from cross-sectional data, these can be specified as mean value parameters for the ERGM

$$P(Y = y|\theta) = \frac{\exp(\theta \cdot g(y))}{\kappa(\theta)}, \tag{7}$$

where y is a realization of some random network Y, $g(y)$ is a vector of sufficient statistics, θ is the vector of corresponding natural parameters, and $\kappa(\theta)$ is a normalizing constant. The method of estimating natural parameters θ for an ERGM of the form (7) using mean value parameters is detailed in the Supplementary Material, Section 5, and builds off the results of Barndorff-Nielsen (2014) [Admiraal and Handcock (2016)].

3.4. *Using ERGMs to simulate networks consistent with degree distributions and mixing totals.* With a method to estimate the natural parameters θ of an ERGM directly from an observed network using Markov chain Monte Carlo (MCMC) maximum likelihood estimation [Hunter and Handcock (2006)] or indirectly through the mean value parameters via Newton–Raphson [Handcock (2003)] as well as to simulate new networks for a set of natural parameters via MCMC [Hunter et al. (2008)], we turn our attention to simulation of networks with specific mixing totals and degree distributions for males and females. Here, we consider the specific ERGM given by

$$\begin{aligned}\log P(Y = y|\theta) = {} & \sum_{i=1}^{I}\sum_{k=1}^{F} \delta_{ik}^{\mathcal{M}} D_{ik}^{\mathcal{M}}(y) + \sum_{j=1}^{J}\sum_{\ell=1}^{M} \delta_{j\ell}^{\mathcal{F}} D_{j\ell}^{\mathcal{F}}(y) \\ & + \sum_{i=1}^{I}\sum_{j=1}^{J} \nu_{ij} N_{ij}(y) - \log \kappa\big(\delta^{\mathcal{M}}, \delta^{\mathcal{F}}, \nu\big),\end{aligned} \tag{8}$$

where $D_{ik}^{\mathcal{M}}(y)$ is the number of males of type i with momentary degree k, $D_{j\ell}^{\mathcal{F}}(y)$ is the number of females of type j with momentary degree ℓ, and $N_{ij}(y)$ is the number of partnerships between males of type i and females of type j. The natural parameters $\delta^{\mathcal{M}} = \{\delta_{11}^{\mathcal{M}}, \ldots, \delta_{IF}^{\mathcal{M}}\}$, $\delta^{\mathcal{F}} = \{\delta_{11}^{\mathcal{F}}, \ldots, \delta_{JM}^{\mathcal{F}}\}$ and $\nu = \{\nu_{11}, \ldots, \nu_{IJ}\}$ correspond to the sufficient statistics $D^{\mathcal{M}} = \{D_{11}^{\mathcal{M}}(y), \ldots, D_{IF}^{\mathcal{M}}(y)\}$, $D^{\mathcal{F}} = \{D_{11}^{\mathcal{F}}(y), \ldots, D_{JM}^{\mathcal{F}}(y)\}$ and $N = \{N_{11}(y), \ldots, N_{IJ}(y)\}$. These network statistics are subject to the constraints

$$\begin{aligned}\sum_{i=1}^{I}\sum_{k=1}^{F} k D_{ik}^{\mathcal{M}}(y) &= \sum_{i=1}^{I}\sum_{j=1}^{J} N_{ij}(y), \\ \sum_{j=1}^{J}\sum_{\ell=1}^{M} \ell D_{j\ell}^{\mathcal{F}}(y) &= \sum_{i=1}^{I}\sum_{j=1}^{J} N_{ij}(y),\end{aligned}$$

$$\sum_{k=1}^{F} k D_{ik}^{\mathcal{M}}(y) = \sum_{j=1}^{J} N_{ij}(y),$$

$$\sum_{\ell=1}^{M} \ell D_{j\ell}^{\mathcal{F}}(y) = \sum_{i=1}^{I} N_{ij}(y),$$

which ensure that the partnership totals produced by the degree distributions are in agreement with those from the mixing totals. The ERGM given by (8) is dyad-independent (i.e., the presence or absence of a tie for a given dyad does not influence the likelihood of a tie for another dyad), meaning that its likelihood can be expressed as a logistic regression model [Krivitsky (2012)]. This ensures that such a model is nondegenerate when parameters lie in the interior of the convex hull.

The desired network statistics corresponding to degree distributions ($\pi^{\mathcal{M}}$, $\pi^{\mathcal{F}}$) and mixing totals (μ) are in fact mean value parameters corresponding to $\delta^{\mathcal{M}}$, $\delta^{\mathcal{F}}$ and ν, respectively, and defined by

$$\pi_{ik}^{\mathcal{M}} = \mathbb{E}_{\delta^{\mathcal{M}},\delta^{\mathcal{F}},\nu}\left(\frac{D_{ik}^{\mathcal{M}}(y)}{M_i}\right),$$

$$\pi_{j\ell}^{\mathcal{F}} = \mathbb{E}_{\delta^{\mathcal{M}},\delta^{\mathcal{F}},\nu}\left(\frac{D_{j\ell}^{\mathcal{F}}(y)}{F_j}\right),$$

$$\mu_{ij} = \mathbb{E}_{\delta^{\mathcal{M}},\delta^{\mathcal{F}},\nu}(N_{ij}(y)).$$

Since we are able to estimate the natural parameters $\delta^{\mathcal{M}}$, $\delta^{\mathcal{F}}$ and ν from $\pi^{\mathcal{M}}$, $\pi^{\mathcal{F}}$ and μ, and since MCMC provides a means to simulate networks for a set of natural parameters, this model allows us to generate networks consistent with the desired degree distributions $\pi^{\mathcal{M}}$ and $\pi^{\mathcal{F}}$ and mixing totals μ.

We have implemented this procedure using the `ergm` package [Handcock et al. (2013), Hunter et al. (2008)], part of the `statnet` suite of `R` packages [Handcock et al. (2003)] available on CRAN [R Core Team (2013)], which has the capability to both determine the natural parameters corresponding to a vector of mean value parameters and simulate networks corresponding to this set of parameters. Thus, the conversion from mean value parameters to natural parameters is one that does not need to be done explicitly by the user but can be done automatically when mean value parameters are passed to the `ergm` function. To demonstrate that our method estimates natural parameters consistent with mean value parameters and, consequently, can simulate networks consistent with desired mean value parameters, we provide a simulation study in the Supplementary Material, Section 6.3 [Admiraal and Handcock (2016)].

4. Modeling race heterogeneity in heterosexual partner selection using the National Longitudinal Study of Adolescent Health. We apply our method to data from the National Longitudinal Study of Adolescent Health (Add Health)

[Harris et al. (2009)] to generate heterosexual partnership networks consistent with race- and sex-specific concurrency and race-based mixing for the population from which respondents were sampled. Add Health is a nationwide survey that began in the United States in 1994–1995. It has followed a group of adolescents periodically over their lifetimes to assess the impact of social and behavioral characteristics in adolescence on particular outcomes. Students from 80 representative high schools and 52 middle schools were included in the study, and a total of four waves of the study have been carried out so far, with the most recent wave being carried out in 2007–2009 and a fifth wave scheduled for 2016–2018.

Wave III of this study was carried out from August 2001 to April 2002 and included extensive questioning related to respondents' romantic relationship histories. At the time of the survey, respondents ranged in age from 18 to 28 years, and all were considered to be part of the sexually active population and included in our analysis. As part of the interview, information was collected on all relationships from June 1995 to the time of the interview. For each reported relationship, respondents were asked if the relationship was sexual and current. Further questions were asked for each partner, including the sex and race of the partner. Restricting our focus to heterosexual relationships, this allowed us to calculate concurrency by sex and race using a direct approach.

Although Eaton, McGrath and Newell (2012) suggest that using the direct approach in measuring concurrency leads to positive bias, their presented current prevalence of concurrency of 6.7% for KwaZulu-Natal compared to a point prevalence of concurrency of 4.7% six months prior to the interview which did not adjust for increasing levels of missing data for partnership data collected for earlier time points. Accounting for missing data leads to discrepancies between current and six month retrospective estimates for only men under the age of 30. Additionally, separate research by Glynn et al. (2012) using data from Malawi only found a 0.5% difference between prevalence of concurrency at the time of the interview and the UNAID Reference Group's recommended six month retrospective measurement. In contrast to the findings of Eaton, McGrath and Newell (2012), the estimate presented by Glynn et al. (2012) for the direct approach was lower than that produced by the indirect approach. Based on the minor differences between the direct and indirect approaches for these studies and conflicting directions of possible biases, Sawers (2013) claims that any bias from using the direct approach in measuring concurrency is minimal. As Wave III of Add Health predated 2009, UNAIDS Reference Group on Estimates, Modelling, and Projections (2009) recommendations were not implemented in the survey, and no questions were included regarding the timing of sexual events to allow for indirect approach measurements of concurrency. Thus, only a direct approach could be used in measuring concurrency at the time of the interview.

From respondents' answers to questions on their relationship histories and whether relationships were sexual and current, we were able to determine degree distributions by sex and race as well as mixing totals by race for respondents at

TABLE 2
Degree distributions for non-Hispanic black and white males and females, along with corresponding sample sizes and population totals

Degree	Black males	White males	Black females	White females
0	320	1855	289	1372
1	417	2254	600	3015
2	50	125	53	131
3	23	24	12	16
4	2	2		2
5	5			
6	2			
7				2
8	1			
Sample size	820	4260	954	4538
Population size	1,012,522	5,257,739	1,177,710	5,600,779

the time of the interview. We specifically consider degree distributions and mixing for non-Hispanic blacks and non-Hispanic whites. Degree distributions for males and females of both races, along with sample sizes and population totals, are presented in Table 2, and mixing totals from male and female reports are shown in Table 3. Note that the majority of partnerships fall along the diagonals of the mixing matrices, suggesting strong assortative mixing by race. Also, note that degree distributions for black males and white females include gaps between the maximal observed degree and lower degrees.

Maximizing the likelihood function in the Supplementary Material, equation (S1), subject to the constraints (6), we obtain the parameter estimates and standard errors shown in Table 4 [Admiraal and Handcock (2016)]. The estimated cell-specific dependence parameters and corresponding standard errors provide evidence of strong assortative mixing by race, as α_{BB} and α_{WW} (corresponding to mixing between black males and black females and white males and white females, respectively) are both significantly higher than 1, whereas α_{BW} and α_{WB} (corresponding to cross-race mixing) are both significantly lower than 1. They also suggest higher levels of concurrency for non-Hispanic black males and females than their non-Hispanic white counterparts. If we consider the point prevalence of concurrency, these (along with corresponding standard errors in parentheses) are estimated to be

	Non-Hispanice Black	Non-Hispanic White
Male	$\sum_{i=2}^{8} \pi_{Bi}^{\mathcal{M}} = 0.1451$ (0.0102)	$\sum_{i=2}^{4} \pi_{Wi}^{\mathcal{M}} = 0.0548$ (0.00344)
Female	$\sum_{i=2}^{3} \pi_{Bi}^{\mathcal{F}} = 0.056$ (0.0068)	$\sum_{i=2}^{7} \pi_{Wi}^{\mathcal{F}} = 0.0254$ (0.0020)

and suggest a point prevalence of concurrency for non-Hispanic black males and females that is roughly 2–2.5 times higher than their non-Hispanic white coun-

TABLE 3

Mixing totals between non-Hispanic black and non-Hispanic white males and females, as reported by males (left) and females (right)

		Females				Males	
		Black	White			Black	White
Males (Respondent)	Black	573	94	Females (Respondent)	Black	722	27
	White	53	2644		White	213	3221

terparts. This combination of strong assortative mixing coupled with higher rates of concurrency for non-Hispanic blacks than non-Hispanic whites would be consistent with the hypothesis that concurrency and selective mixing may partially explain higher prevalence of STIs observed in non-Hispanic blacks than whites, as illustrated for Gonorrhea in Figure 1.

To simulate networks consistent with the estimated degree distributions and mixing propensities, the estimates for the expected degree distributions $\pi_B^{\mathcal{M}}$, $\pi_W^{\mathcal{M}}$, $\pi_B^{\mathcal{F}}$ and $\pi_W^{\mathcal{F}}$ serve as the mean value parameters for the degree distributions for

TABLE 4

Parameter estimates and corresponding standard errors for $\pi^{\mathcal{M}}$ and $\pi^{\mathcal{F}}$, the degree distribution mean value parameters, and α, the cell-specific dependence parameters

		Estimate	**Std. err.**			**Estimate**	**Std. err.**
Degree distributions:							
Black males	$\pi^{\mathcal{M}}_{\mathrm{B0}}$	0.34578	0.01479	White males	$\pi^{\mathcal{M}}_{\mathrm{W0}}$	0.37719	0.00531
	$\pi^{\mathcal{M}}_{\mathrm{B1}}$	0.50909	0.01755		$\pi^{\mathcal{M}}_{\mathrm{W1}}$	0.56801	0.00729
	$\pi^{\mathcal{M}}_{\mathrm{B2}}$	0.07015	0.00957		$\pi^{\mathcal{M}}_{\mathrm{W2}}$	0.04141	0.00335
	$\pi^{\mathcal{M}}_{\mathrm{B3}}$	0.03793	0.00770		$\pi^{\mathcal{M}}_{\mathrm{W3}}$	0.01160	0.00213
	$\pi^{\mathcal{M}}_{\mathrm{B4}}$	0.00400	0.00282		$\pi^{\mathcal{M}}_{\mathrm{W4}}$	0.00179	0.00119
	$\pi^{\mathcal{M}}_{\mathrm{B5}}$	0.01270	0.00553				
	$\pi^{\mathcal{M}}_{\mathrm{B6}}$	0.00696	0.00478				
	$\pi^{\mathcal{M}}_{\mathrm{B8}}$	0.01339	0.00628				
Black females	$\pi^{\mathcal{F}}_{\mathrm{B0}}$	0.33990	0.01577	White females	$\pi^{\mathcal{F}}_{\mathrm{W0}}$	0.35221	0.00505
	$\pi^{\mathcal{F}}_{\mathrm{B1}}$	0.60408	0.01607		$\pi^{\mathcal{F}}_{\mathrm{W1}}$	0.62235	0.00570
	$\pi^{\mathcal{F}}_{\mathrm{B2}}$	0.04664	0.00628		$\pi^{\mathcal{F}}_{\mathrm{W2}}$	0.02261	0.00189
	$\pi^{\mathcal{F}}_{\mathrm{B3}}$	0.00938	0.00270		$\pi^{\mathcal{F}}_{\mathrm{W3}}$	0.00237	0.00059
					$\pi^{\mathcal{F}}_{\mathrm{W4}}$	0.00026	0.00018
					$\pi^{\mathcal{F}}_{\mathrm{W7}}$	0.00019	0.00013
Dependence parameters:							
	α_{BB}	4.37022	0.11983		α_{BW}	0.24061	0.01186
	α_{WB}	0.07106	0.00804		α_{WW}	1.20931	0.01074

TABLE 5
Mean value parameters $\mu = \{\mu_{BB}, \mu_{BW}, \mu_{WB}, \mu_{WW}\}$ *for mixing totals*

		Females	
		Black	White
Males	Black	806,841.8	197,152.4
	White	47,595.7	3,594,920.0

non-Hispanic black and white males and females. Although the expected mixing totals μ are not directly estimated, these can be recovered using (4) and the estimated degree distributions ($\pi_B^{\mathcal{M}}, \pi_W^{\mathcal{M}}, \pi_B^{\mathcal{F}}, \pi_W^{\mathcal{F}}$), cell-specific dependence parameters ($\alpha_{BB}, \alpha_{BW}, \alpha_{WB}, \alpha_{WW}$) and population sizes by sex and race. These expected mixing totals are presented in Table 5 and serve as the mean value parameters for mixing.

We simulated networks scaled to one ten-thousandth the size of the population considered as part of Add Health in order to produce networks that could readily be handled by `ergm` for simulation purposes. In applying this scalar, we preserved relative compositions by race, corresponding to reported subpopulation sizes in Table 2. This resulted in a simulated population of 1305 individuals consisting of 101 non-Hispanic black males, 526 non-Hispanic white males, 118 non-Hispanic black females and 560 non-Hispanic white females. By applying the same scalar to each sex–race category for the mean value parameters for mixing shown in Table 5, consistency with the mean value parameters for degree distributions is maintained.

Using the mean value parameters $\pi^{\mathcal{M}}$ and $\pi^{\mathcal{F}}$ for degree distributions and μ for mixing totals, we estimate the natural parameters $\delta^{\mathcal{M}}$, $\delta^{\mathcal{F}}$ and ν for the ERGM given by (8) using the procedure outlined in the Supplementary Material, Section 5 [Admiraal and Handcock (2016)]. We note, however, that we use truncated degree distributions rather than full degree distributions in the estimation procedure, and so only a subset of observed degrees is specified in the ERGM. This is for two reasons. First, specification of full degree distributions corresponding to observed degrees would result in the introduction of structural zeros for all unobserved degrees. This would mean that maximal degrees would be artificially capped at the maximum observed degree for a given subpopulation. Those with high degrees are less likely to be observed in cross-sectional surveys but are important for disease transmission, and so exclusion of such individuals by capping the maximum degree is potentially problematic. At the same time, if we consider the observed degree distributions for Add Health, we notice that there are gaps in the degree distributions for non-Hispanic black males and white females. These gaps are almost certainly due to sampling variability, and specifying structural zeros would artificially enforce an unrealistic degree distribution where, for instance, white females may have four partners or seven partners but not five or six.

One way to eliminate structural zeros is to artificially inflate cases with no observations, in which case we would obtain nonzero probabilities of observing individuals with degrees that were not actually observed in the sample. However,

we note that this would increase the population size (by introducing individuals corresponding to unobserved degrees) while leading to a significant increase in the total number of partnerships corresponding to those subpopulations with gaps in observed degree distributions. It would also require decisions in regard to maximum degree. Additionally, such an approach would impact estimates for degree distributions for other subpopulations and mixing totals because of the constraints that govern the relationships among degree distributions and mixing totals.

An alternative approach is to use truncated degree distributions. By truncating the degree distributions, the ERGM specifies a subset of degrees for which mean value parameters are used to estimate corresponding natural parameters. These degrees are informed by the data, and so the ERGM is able to target the specified mean value parameters, but this is not true for degrees that are omitted due to truncation. As long as truncation ensures that at least one observed degree is omitted from the model for each subpopulation, then the ERGM will attribute nonzero probability to all degrees omitted due to truncation. Ideally, then, truncation eliminates a minimal subset of observed degrees so that simulated networks are able to target observed degree distributions as closely as possible while ensuring that all degrees have nonzero probability of occurring. Truncation can achieve a similar result to that of artificially inflating the number of cases corresponding to unobserved degrees but without expanding the population, affecting the consistency of estimated degree distributions and mixing totals, or forcing decisions in regard to maximum degree. Consequently, we opt for truncated degree distributions.

Second, we use truncated degree distributions because, even if there were no gaps in observed degree distributions and there was a clearly specified maximal degree, inclusion of full observed degree distributions and mixing totals would lead to an over-specified model. This is due to both the dependence among degree distribution parameters within a given degree distribution, as reflected in the multinomial constraints given by (6), and the dependence between degree distributions and mixing totals, as specified in the constraints given by (1)–(4). Consequently, degree distributions would need to be truncated to ensure identifiability in our ERGM, and we truncate the degree distributions for males to include 0–3 partners (as the maximum observed degree for non-Hispanic white males is four) and degree distributions for females to include 0–2 partners (as the maximum observed degree for non-Hispanic black females is three).

Using the ERGM given by (8) and estimated natural parameters based on the specified selective mixing totals and truncated degree distributions, we simulate 10,000 partnership networks using a burn-in of 100,000 iterations and interval length of 10,000. This produces the simulated sex- and race-specific degree distributions shown in Figure 3 and simulated mixing totals presented in Figure 4. These simulated distributions, represented by boxplots, are shown along with gray diamonds for the mean value parameters (i.e., estimated degree distributions and mixing totals) used to generate the natural parameters of the ERGM. These mean value parameters represent the targeted values for each of these distributions.

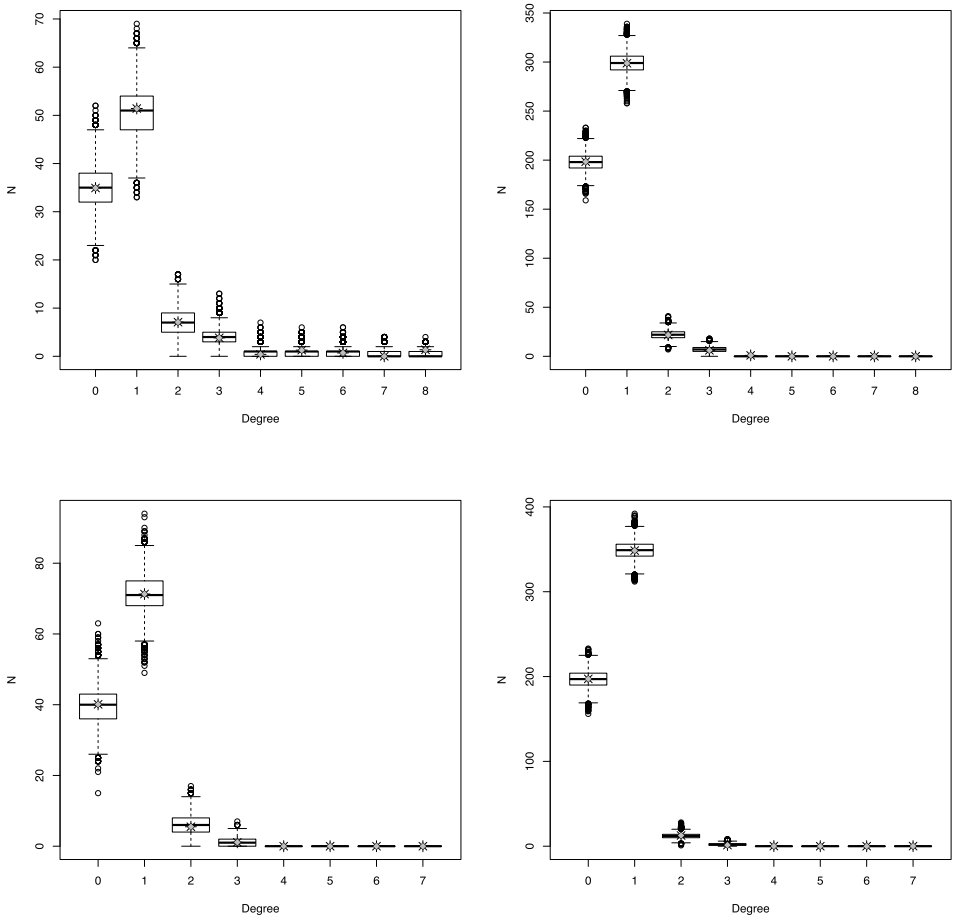

FIG. 3. *Simulated degree distributions for non-Hispanic black males* (*top left*), *non-Hispanic white males* (*top right*), *non-Hispanic black females* (*bottom left*) *and non-Hispanic white females* (*bottom right*). *Gray diamonds denote the mean value parameters for degree distributions.*

We note that the distributions of the statistics from the simulated networks are centered on the target mean values parameters, suggesting that our method produces networks consistent in expectation with the desired degree distributions and mixing totals for the population of interest. This is true even though degree distribution terms estimated for the ERGM incorporated only truncated degree distributions. We present simulated degree distributions and mean value parameters for degrees up to the maximum observed degree for each sex. Because truncated degree distributions were used, simulated degrees of 4–8 for males and 3–7 for females were done so without a specified target value. It should be evident that there is greater variability in simulated distributions for these degrees for non-Hispanic black males and females than their white counterparts. This can be ex-

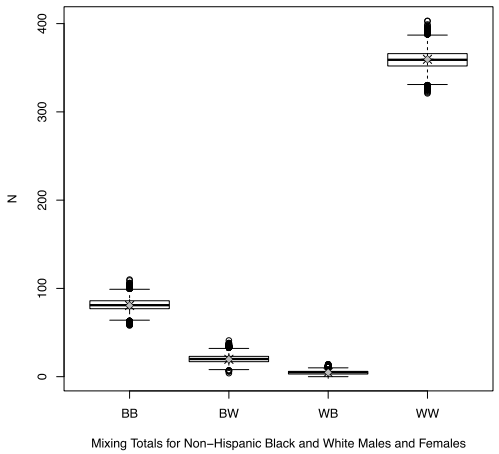

FIG. 4. *Simulated mixing totals by sex for non-Hispanic black* (*B*) *and non-Hispanic white* (*W*) *Americans. "BW" denotes partnerships between black males and white females, while "WB" denotes partnerships between white males and black females. Gray diamonds denote the mean value parameters for mixing totals.*

plained by the cumulative probabilities corresponding to these degrees without specified target values for each sex- and race-category, as based on estimated degree distributions shown in Table 4. In particular, this probability is 0.0371 and 0.0094 for non-Hispanic black males and females, respectively, while it is only 0.0018 and 0.0005 for their white male and female counterparts, respectively. It is important for modelers to recognize that variability in simulated distributions for higher degrees increases with greater levels of truncation, and so truncating a minimal number of observed degrees with low empirical probabilities will lead to greater precision in simulated network degree distributions.

5. Discussion. The method presented in this paper provides a means to use cross-sectional survey data to simulate initial networks consistent with two structures that may be important in STI transmission for heterosexual partnership networks—concurrency and selective mixing. Heterosexual networks generated using our approach can serve as initial networks for dynamic models, such as those of Snijders (2001) and Krivitsky and Handcock (2014). Such models can include a variety of subsequent stages or components, including random selection of an initial set of infected individuals, modeling of partnership duration and new partner selection (which can account for a variety of factors, including individual or pair characteristics, degrees of individuals, infection status, etc.), disease transmission (which depends on HIV stage, coital frequency, etc.) and population dynamics (where individuals enter or exit the population of sexually active individuals). Our method provides an important advancement in the first step of disease models, as it not only allows modelers to produce initial networks for a heterogeneous

population with different levels of concurrency (and, more specifically, degree distributions) according to various factors (e.g., sex and race), but also ensures that both selective mixing and degree distributions are consistent with that observed in the underlying population.

While the methods presented are explicitly for cross-sectional data, note that extensions to full network data are straightforward. In the rare instances where it may be possible to observe a complete network, degree distributions and selective mixing totals are guaranteed to be consistent with each other, meaning that the first step outlined in Figure 2 is unnecessary. Hunter and Handcock (2006) demonstrate how, in such cases, the natural parameters of an ERGM can be estimated directly from the observed network using MCMC maximum likelihood estimation, meaning that networks matching desired degree distributions and mixing totals can be simulated through a two-step process of first estimating the natural parameters from the network and then simulating networks according to these natural parameters. Developments by Handcock and Gile (2010) show how these methods can be extended to adaptive samples, such as data collected through link-tracing.

Although not specifically tied to concurrency, Aral et al. (1999) note that what is important for disease transmission is not only what happens locally but also what happens in the broader network. In other words, focus should not be restricted to just an individual and his or her partners (which is what concurrency does) but should also include a partner's partners. This suggests that dyad-dependent characteristics of a network such as 3-paths (i.e., a path of three edges connecting two males and two females for a heterosexual network), 4-cycles (the smallest connected component for a heterosexual network, consisting of two individuals of the same sex who share at least two partners) and the size of the largest connected component may be important to consider. (See Figure 5 for minimal network examples of both a 3-path and a 4-cycle.) Thus, focusing on concurrency in a strict sense may fail to capture other relevant network information.

For cross-sectional survey data, only dyad-independent network statistics such as concurrency and mixing can be measured, whereas dyad-dependent statistics such as 3-paths, 4-cycles and connected component size cannot. To accurately simulate networks consistent with these (as well as other dyad-dependent network statistics), full network data would likely be required, as Shalizi and Rinaldo

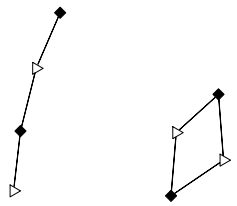

FIG. 5. *Examples of a* 3*-path* (*left*) *and a* 4*-cycle* (*right*) *for a heterosexual population of males* (*black square*) *and females* (*white triangle*).

(2013) demonstrate that models incorporating a variety of dyad-dependent terms are not projective. In other words, if model parameters are estimated for a subnetwork, then these are not necessarily true for the full network and vice versa, and so parameters are not consistent under scaling of the population. Carnegie and Morris (2011) provide a clear example of this, showing that, for a fixed level of concurrency and random mixing, increased network size leads to a smaller size of the largest connected component.

Models including only dyad-independent terms are projective, and so parameter estimates for the terms considered in our ERGM given by (8), which are dyad-independent, are consistent under scaling of the population. However, the results of Carnegie and Morris (2011) mean that the size of a sexual network may play a role in the level of connectedness (and, hence, efficacy for disease transmission) for a given level of concurrency, and so it is important to recognize the potential ramifications of a selected population size for simulations on a variety of dyad-dependent statistics. If appropriate targets for these dyad-dependent statistics are estimable from other sources of data or expert opinion, these can be incorporated in our estimation procedure by including terms for these in the specified ERGM and then using estimates for these statistics as mean value parameters. These can be paired with estimates for degree distributions and selective mixing totals in the second step of our procedure in Figure 2, producing networks that match not only degree distributions and mixing for the underlying population but also other desired network statistics. We demonstrate this in the case of both 3-paths and 4-cycles in the Supplementary Material, Section 6.4 [Admiraal and Handcock (2016)].

For the Add Health data, the distribution of 3-paths and 4-cycles for simulated networks under various scalings of the population is presented in Figure 6. For each

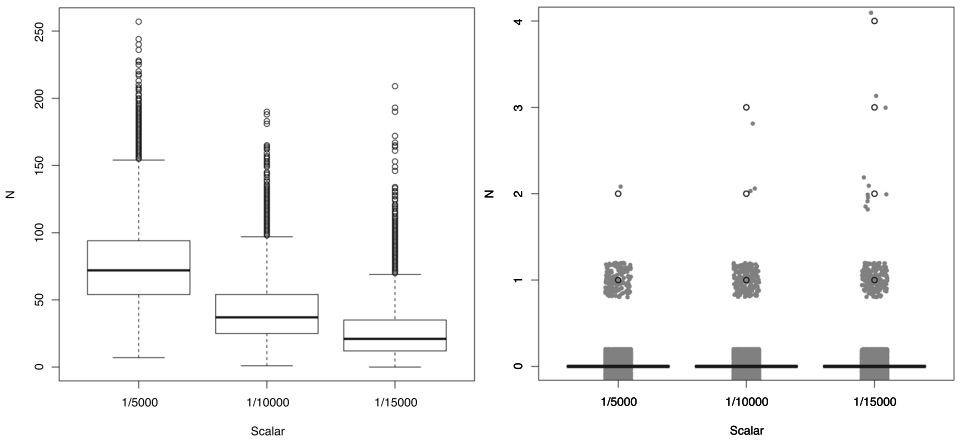

FIG. 6. *Simulated distribution of* 3*-paths* (*left*) *and* 4*-cycles* (*right*) *under various scalings of the population. Observed numbers of* 4*-cycles for the various scaling constants are presented as jittered gray circles.*

scalar considered, 10,000 networks were simulated with a burn-in of 100,000 iterations and interval length of 10,000. For 4-cycles, observed data are plotted over the boxplots and jittered to highlight the relative frequency of networks with at least one 4-cycle. This illustrates that smaller populations are likely to have fewer 3-paths and more 4-cycles. Additionally, neither the number of 4-cycles nor the number of 3-paths are linear functions of the scaling constant, indicating that the model is not projective for these network statistics. This highlights the importance of specifying target mean value parameters for such statistics that are appropriate for the population size under consideration. Although, ideally, simulations would consider a population size equivalent to that of the population under consideration, the computational burden associated with network-based models frequently prevents this. Consequently, modelers need to be cognizant of the impacts of selected scaling constants.

Acknowledgments. This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

SUPPLEMENTARY MATERIAL

Supplement to "Modeling concurrency and selective mixing in heterosexual partnership networks with applications to sexually transmitted diseases." (DOI: 10.1214/16-AOAS963SUPP; .pdf). We provide a full exposition of common measures of concurrency, studies providing evidence for the importance of concurrency in explaining disparities in the spread of sexually transmitted diseases and studies refuting this conclusion. We additionally present details for estimating natural parameters from mean value parameters for exponential-family random graph models, and we provide a full simulation study that demonstrates the usefulness of our method in generating networks consistent with populations having drastically different levels of concurrency and selective mixing patterns.

REFERENCES

ADIMORA, A. A. and SCHOENBACH, V. J. (2002). Contextual factors and the black–white disparity in heterosexual HIV transmission. *Epidemiology* **13** 707–712.

ADIMORA, A. A. and SCHOENBACH, V. J. (2005). Social context, sexual networks, and racial disparities in rates of sexually transmitted infections. *J. Infect. Dis.* **191** S115–S122.

ADIMORA, A. A., SCHOENBACH, V. J. and DOHERTY, I. A. (2006). HIV and African Americans in the southern United States: Sexual networks and social context. *Sex. Transm. Dis.* **33** S39–S45.

ADIMORA, A. A., SCHOENBACH, V. J. and DOHERTY, I. A. (2007). Concurrent sexual partnerships among men in the United States. *Am. J. Publ. Health* **97** 2230–2237.

ADMIRAAL, R. and HANDCOCK, M. S. (2016). Supplement to "Modeling concurrency and selective mixing in heterosexual partnership networks with applications to sexually transmitted diseases." DOI:10.1214/16-AOAS963SUPP.

ANDERSON, R. M. (1992). Some aspects of sexual behavior and the potential demographic impact of AIDS in developing countries. *Social Science and Medicine* **34** 271–280.

ANDERSON, R. M., GUPTA, S. and NG, W. (1990). The significance of sexual partner contact networks for the transmission dynamics of HIV. *J. Acquir. Immune Defic. Syndr.* **3** 417–429.

ARAL, S. O., HUGHES, J. P., STONER, B., WHITTINGTON, W., HANDSFIELD, H. H., ANDERSON, R. M. and HOLMES, K. K. (1999). Sexual mixing patterns in the spread of gonococcal and chlamydial infections. *Am. J. Publ. Health* **89** 825–833.

BARNDORFF-NIELSEN, O. (2014). *Information and Exponential Families in Statistical Theory*. Wiley, Chichester. MR3221776

BUSENBERG, S. and CASTILLO-CHAVEZ, C. (1989). Interaction, pair formation and force of infection terms in sexually transmitted diseases. In *Mathematical and Statistical Approaches to AIDS Epidemiology*. *Lecture Notes in Biomathematics* **83** 289–300. Springer, Berlin. MR1040595

CARNEGIE, N. B. and MORRIS, M. (2011). Size matters: Concurrency and the epidemic potential of HIV in small networks. *PLoS One* **7** e43048.

CASTILLO-CHAVEZ, C. and BLYTHE, S. P. (1989). Mixing framework for social/sexual behavior. In *Mathematical and Statistical Approaches to AIDS Epidemiology* (C. Castillo-Chavez, ed.). *Lecture Notes in Biomathematics* **83** 275–288. Springer, Berlin. MR1040594

CENTERS FOR DISEASE CONTROL AND PREVENTION (2015). Sexually Transmitted Diseases: Data & Statistics. Available at http://www.cdc.gov/std/stats/default.htm, 2015. Accessed: December 1, 2015.

CHICK, S. E., ADAMS, A. L. and KOOPMAN, J. S. (2000). Analysis and simulation of a stochastic, discrete-individual model of STD transmission with partnership concurrency. *Math. Biosci.* **166** 45–68.

DEMOGRAPHIC AND HEALTH SURVEYS PROGRAM (2015). HIV/AIDS Survey Indicators Database. Available at http://hivdata.dhsprogram.com, 2015. Accessed: December 1, 2015.

DOHERTY, I. A., SHIBOSKI, S., ELLEN, J. M., ADIMORA, A. A. and PADIAN, N. S. (2006). Sexual bridging socially and over time: A simulation model exploring the relative effects of mixing and concurrency on viral sexually transmitted infection transmission. *Sex. Transm. Dis.* **33** 368–373.

EATON, J. W., HALLETT, T. B. and GARNETT, G. P. (2011). Concurrent sexual partnerships and primary HIV infection: A critical interaction. *AIDS Behav.* **15** 687–692.

EATON, J. W., MCGRATH, N. and NEWELL, M.-L. (2012). Unpacking the recommended indicator for concurrent sexual partnerships. *AIDS* **26** 1037–1039.

GARNETT, G. and ANDERSON, R. M. (1993a). Contact tracing and the estimation of sexual mixing patterns: The epidemiology of Gonococcal infections. *Sex. Transm. Dis.* **20** 181–191.

GARNETT, G. P. and ANDERSON, R. M. (1993b). Factors controlling the spread of HIV in heterosexual communities in developing countries: Patterns of mixing between different age and sexual activity classes. *Philosophical Transactions*: *Biological Sciences* **342** 137–159.

GARNETT, G. P., HUGHES, J. P., ANDERSON, R. M., STONER, B. P., ARAL, S. O., WHITTINGTON, W. L., HANDSFIELD, H. H. and HOLMES, K. K. (1996). Sexual mixing patterns of patients attending sexually transmitted diseases clinics. *Sex. Transm. Dis.* **23** 248–257.

GHALANOS, A. and THEUSSLS, S. (2012). Rsolnp: General Non-linear Optimization Using Augmented Lagrange Multiplier Method. Available at CRAN.R-project.org/package=Rsolnp. Version 1.14.

GHANI, A. C. and GARNETT, G. P. (2000). Risks of acquiring and transmitting sexually transmitted diseases in sexual partner networks. *Sex. Transm. Dis.* **27** 579–587.

GHANI, A. C., SWINTON, J. and GARNETT, G. P. (1997). The role of sexual partnership networks in the epidemiology of gonorrhea. *Sex. Transm. Dis.* **24** 45–56.

GLYNN, J. R., DUBE, A., KAYUNI, N., FLOYD, S., MOLESWORTH, A., PARROTT, F., FRENCH, N. and CRAMPIN, A. C. (2012). Measuring concurrency: An empirical study of different methods in a large population-based survey in northern Malawi and evaluation of the UNAIDS guidelines. *AIDS* **26** 977–985.

GOODREAU, S. M. (2011). A decade of modelling research yields considerable evidence for the importance of concurrency: A response to Sawers and Stillwaggon. *Journal of the International AIDS Society* **14** 1–7.

GOODREAU, S. M., CASSELS, S., KASPRZYK, D., MONTAÑO, D. E., GREEK, A. and MORRIS, M. (2012). Concurrent partnerships, acute infection and HIV epidemic dynamics among young adults in Zimbabwe. *AIDS Behav.* **6** 312–322.

GRULICH, A. E. and ZABLOTSKA, I. (2010). Commentary: Probability of HIV transmission through anal intercourse. *Int. J. Epidemiol.* **39** 1064–1065.

GUPTA, S., ANDERSON, R. M. and MAY, R. M. (1989). Networks of sexual contacts: Implications for the pattern of spread of HIV. *AIDS* **3** 807–817.

HAMILTON, D. T., HANDCOCK, M. S. and MORRIS, M. (2008). Degree distributions in sexual networks: A framework for evaluating evidence. *Sex. Transm. Dis.* **35** 30–40.

HAMILTON, D. T. and MORRIS, M. (2015). The racial disparities in STI in the U.S.: Concurrency, STI prevalence, and heterogeneity in partner selection. *Epidemics* **11** 56–61.

HANDCOCK, M. S. (2003). Assessing degeneracy in statistical models of social networks. Working paper, Center for Statistics and the Social Sciences, Univ. of Washington.

HANDCOCK, M. S. and GILE, K. J. (2010). Modeling networks from sampled data. *Ann. Appl. Stat.* **40** 285–327.

HANDCOCK, M. S., RENDALL, M. S. and CHEADLE, J. E. (2005). Improved regression estimation of a multivariate relationship with population data on the bivariate relationship. *Sociol. Method.* **35** 291–334.

HANDCOCK, M. S., HUNTER, D. R., BUTTS, C. T., GOODREAU, S. M. and MORRIS, M. (2003). statnet: Software tools for the Statistical Modeling of Network Data. Seattle, WA, 2003. Available at http://statnetproject.org.

HANDCOCK, M. S., HUNTER, D. R., BUTTS, C. T., GOODREAU, S. M., KRIVITSKY, P. N. and MORRIS, M. (2013). ergm: Fit, Simulate and Diagnose Exponential-Family Models for Networks. The Statnet Project (http://www.statnet.org), 2013. Available at CRAN.R-project.org/package=ergm. R package version 3.1-0.

HARRIS, K. M., HALPERN, C. T., WHITSEL, E., HUSSEY, J., TABOR, J., ENTZEL, P. and UDRY, J. R. (2009). The National Longitudinal Study of Adolescent Health: Research Design [www document]. Technical report, Carolina Population Center, University of North Carolina at Chapel Hill, Available at: http://www.cpc.unc.edu/projects/addhealth/design.

HELLERINGER, S., MKANDAWIRE, J. and KOHLER, H.-P. (2014). A new approach to measuring partnership concurrency and its association with HIV risk in couples. *AIDS Behav.* **18** 2291–2301.

HUDSON, C. (1993). Concurrent partnerships could cause AIDS epidemics. *International Journal of STD and AIDS* **4** 349–353.

HUNTER, D. R. and HANDCOCK, M. S. (2006). Inference in curved exponential family models for networks. *J. Comput. Graph. Statist.* **15** 565–583. MR2291264

HUNTER, D. R., HANDCOCK, M. S., BUTTS, C. T., GOODREAU, S. M. and MORRIS, M. (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *J. Stat. Softw.* **24** 1–11.

HYMAN, J. M. and STANLEY, E. A. (1988). Using mathematical models to understand the AIDS epidemic. *Math. Biosci.* **90** 415–473. MR0958152

JACQUEZ, J. A., SIMON, C. P. and KOOPMAN, J. (1989). Structured mixing: Heterogeneous mixing by the definition of activity groups. In *Mathematical and Statistical Approaches to AIDS Epidemiology* (C. Castillo-Chavez, ed.). *Lecture Notes in Biomathematics* **83** 301–315. Springer, Berlin. MR1040596

JOHNSON, L. F., DORRINGTON, R. E., BRADSHAW, D., PILLAY-VAN WYK, V. and REHLE, T. M. (2009). Sexual behaviour patterns in South Africa and their association with the spread of HIV: Insights from a mathematical model. *Demogr. Res. Monogr.* **21** 289–339.

JULIAN, D., BOUCHARD, C., GAGNON, M. and POMERLEAU, A. (1992). Insider's views of marital sex: A dyadic analysis. *J. Sex Res.* **29** 343–360.

KINSEY, A. C., POMEROY, W. B. and MARTIN, C. E. (1948). *Sexual Behavior in the Human Male*. W. B. Saunders Company, Philadelphia.

KOOPMAN, J. S., CHICK, S. E., RIOLO, C. S., ADAMS, A. L., WILSON, M. L. and BECKER, M. P. (2000). Modeling contact networks and infection transmission in geographic and social space using GERMS. *Sex. Transm. Dis.* **27** 617–626.

KRETZSCHMAR, M. and CARAËL, M. (2012). Is concurrency driving HIV transmission in Sub-Saharan African sexual networks? The significance of sexual partnership typology. *AIDS Behav.* **16** 1746–1752.

KRETZSCHMAR, M. and MORRIS, M. (1996). Measures of concurrency in networks and the spread of infectious disease. *Math. Biosci.* **133** 165–195.

KRIVITSKY, P. N. (2012). Exponential-family random graph models for valued networks. *Electron. J. Stat.* **6** 1100–1128. MR2988440

KRIVITSKY, P. N. and HANDCOCK, M. S. (2014). A separable model for dynamic networks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 29–46. MR3153932

MAH, T. L. and HALPERIN, D. T. (2010). Concurrent sexual partnerships and the HIV epidemics in Africa: Evidence to move forward. *AIDS Behav.* **14** 11–16.

MORIN, B. R., PERRINGS, C., LEVIN, S. and KINZIG, A. (2014). Disease risk mitigation: The equivalence of two selective mixing strategies on aggregate contact patterns and resulting epidemic spread. *J. Theoret. Biol.* **363** 262–270. MR3278717

MORRIS, M. (1991). A log-linear modeling framework for selective mixing. *Math. Biosci.* **2** 349–377.

MORRIS, M. (1994). Epidemiology and social networks: Modeling structured diffusion. In *Advances in Social Network Analysis*: *Research in the Social and Behavioral Sciences* (S. Wasserman and J. Galaskiewicz, eds.) 26–52. Sage Publications, Thousand Oaks.

MORRIS, M. (1995). Data driven network models for the spread of infectious disease. In *Epidemic Models*: *Their Structure and Relation to Data* (D. Mollison, ed.) 302–322. Cambridge Univ. Press, Cambridge.

MORRIS, M. (1997). Sexual networks and HIV. *AIDS* **11** S209–S216.

MORRIS, M., EPSTEIN, H. and WAWER, M. (2010). Timing is everything: International variations in historical sexual partnership concurrency and HIV prevalence. *PLoS ONE* **5** 31–33.

MORRIS, M., GOODREAU, S. M. and MOODY, J. (2007). Sexual networks, concurrency, and STD/HIV. In *Sexually Transmitted Diseases*, 4th ed. (K. K. Holmes, P. F. Sparling, W. E. Stamm, P. Piot, J. N. Wasserheit, L. Corey and D. H. Watts, eds.) 109–125. McGraw-Hill, New York.

MORRIS, M. and KRETZSCHMAR, M. (1995). Concurrent partnerships and transmission dynamics in networks. *Social Networks* **17** 299–318.

MORRIS, M. and KRETZSCHMAR, M. (1997). Concurrent partnerships and the spread of HIV. *AIDS* **5** 641–648.

MORRIS, M. and KRETZSCHMAR, M. (2000). A microsimulation study of the effect of concurrent partnerships on the spread of HIV in Uganda. *Math. Popul. Stud.* **8** 109–133. The population dynamics of the HIV epidemic: projections. MR1806009

MORRIS, M., KURTH, A. E., HAMILTON, D. T., MOODY, J. and WAKEFIELD, S. (2009). Concurrent partnerships and HIV prevalence disparities by race: Linking science and public health practice. *Am. J. Publ. Health* **99** 1023–1031.

OCHS, E. P. and BINIK, Y. M. (1999). The use of couple data to determine the reliability of self-reported sexual behavior. *J. Sex Res.* **36** 374–384.

R CORE TEAM (2013). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2013. Available at http://www.R-project.org/.

RENDALL, M. S., ADMIRAAL, R., DEROSE, A., DIGIULIO, P., HANDCOCK, M. S. and RACIOPPI, F. (2008). Population constraints on pooled surveys in demographic hazard modeling. *Stat. Methods Appl.* **17** 519–539. MR2447573

SAWERS, L. (2013). Measuring and modelling concurrency. *Journal of the International AIDS Society* **16** 1–20.

SEAL, D. W. (1997). Interpartner concordance of self-reported sexual behavior among college dating couples. *The Journal of Sex Research* **34** 39–55.

SHALIZI, C. R. and RINALDO, A. (2013). Consistency under sampling of exponential random graph models. *Ann. Statist.* **41** 508–535. MR3099112

SNIJDERS, T. A. B. (2001). The statistical evaluation of social network dynamics. *Sociol. Method.* **31** 361–95.

UNAIDS REFERENCE GROUP ON ESTIMATES, MODELLING, AND PROJECTIONS (2009). Consultation on Concurrent Sexual Partnerships: Recommendations from a meeting of the UNAIDS Reference Group on Estimates, Modelling and Projections held in Nairobi, Kenya, April 20–21st 2009.

WATTS, C. H. and MAY, R. M. (1992). The influence of concurrent partnerships on the dynamics of HIV/AIDS. *Math. Biosci.* **108** 89–104.

WORLD HEALTH ORGANIZATION (2013). Number of people (all ages) living with HIV: Estimates by WHO region. Available at http://apps.who.int/gho/data/view.main.22100WHO?, 2013. Accessed: December 1, 2015.

YE, Y. (1987). Interior algorithms for linear, quadratic, and linearly constrained non-linear programming Ph.D. Thesis, Stanford Univ., Dept. of EES.

MATHEMATICS AND STATISTICS
MURDOCH UNIVERSITY
WESTERN AUSTRALIA
AUSTRALIA
E-MAIL: R.Admiraal@murdoch.edu.au

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
LOS ANGELES
USA
E-MAIL: handcock@stat.ucla.edu