DEPARTMENT OF STATISTICS
The Pennsylvania State University
University Park, PA 16802 U.S.A.

TECHNICAL REPORTS AND PREPRINTS

Number 12-04: October 2012

An Approximation Method for Improving Dynamic Network Model Fitting

Nicole Bohme Carnegie[1], Pavel N. Krivitsky[2], David R. Hunter[2], and Steven M. Goodreau[3]

[1]Harvard University
[2]The Pennsylvania State University
[3]University of Washington

An approximation method for improving dynamic network model fitting

Nicole Bohme Carnegie[1], Pavel N. Krivitsky[2], David R. Hunter[2], and Steven M. Goodreau[3]

[1]Harvard University
[2]Pennsylvania State University
[3]University of Washington

October 28, 2012

Abstract

There has been a great deal of interest recently in the modeling and simulation of dynamic networks, i.e., networks that change over time. One promising model is the separable temporal exponential-family random graph model (ERGM) of Krivitsky and Handcock, which treats the formation and dissolution of ties in parallel at each time step as independent ERGMs. However, the computational cost of fitting these models can be substantial, particularly for large, sparse networks. Fitting cross-sectional models for observations of a network at a single point in time, while still a non-negligible computational burden, is much easier. In this paper we show that a simple adjustment to the cross-sectional network parameters based on the mean duration of relationships is an adequate approximation to the dynamic parameters for sparse networks with relationships of moderate or long duration. In fact, the approximation method works best in precisely those cases where parameter estimation is most likely to fail—networks with very little change at each time step. We provide both empirical evidence of the applicability of the adjustment and a theoretical justification for certain cases. We consider a variety of cases: Bernoulli formation and dissolution of ties, independent-dyad formation and Bernoulli dissolution, independent-dyad formation and dissolution, and dependent-dyad formation models.

Key Words: Dynamic networks, model fitting, exponential random graph models (ERGMs), separable temporal exponential random graph models (STERGMs), Markov chain Monte Carlo

1 Introduction

Dynamic social networks—those in which relations form and break over time—have long been an area of research interest, both in terms of understanding the nature of the dynamics themselves (Holland and Leinhardt, 1977; Frank, 1991; Doreian and Stokman, 1997) and their implication for processes such as the flow of information or disease (Mercken et al., 2009; Sasovova et al., 2010; Jack et al., 2010; Weerman, 2011). This line of work remains very active, with a variety of models for dynamic networks being explored (Aral et al., 2009; Snijders, 2001; Snijders et al., 2007).

Recently, the exponential-family random graph modeling (ERGM) framework has become a popular approach for conducting inference on network structure, due to its generalizability and its basis in exponential family theory, a common and parsimonious statistical approach (Strauss and Ikeda, 1990; Wasserman and Pattison, 1996; Robins and Pattison, 2001; Snijders et al., 2006; Robins and Morris, 2007; Hunter et al., 2008a, for example). For many years, the application of the ERGM framework was limited to static, or cross-sectional, network datasets. However, recent developments have extended ERGMs to the modeling of dynamic networks, including the separable temporal ERGM (STERGM) approach of Krivitsky and Handcock (2010). The STERGM is a form of discrete temporal ERGM of Hanneke et al. (2010) that independently models the formation and dissolution of ties over discrete time steps. This allows for independent control of the incidence of tie formation and of relational duration. In the STERGM framework, it is possible to make use of independent data on cross-sectional network structure and the duration of relationships and to model the dynamic evolution of the network in a principled way.

While Krivitsky and Handcock (2010) and Hanneke et al. (2010) develop methods to fit dynamic ERGMs to panel data using conditional MLE (CMLE), network panel data are often not available about networks of interest, with sexual partnership networks be-

ing a major example. Krivitsky (2012a) fits STERGMs to cross-sectional and egocentric network data using a gradient descent to find the generalized methods of moments estimator (GMME) for the parameters, an approach that has been used in multiple applications (Morris et al., 2009; Goodreau et al., 2010; Krivitsky, 2009). In each of these cases, parameters for the relational dissolution process were assumed, and starting values for the estimation of the formation model were obtained from fitting a traditional static ERGM on the cross-sectional data. Although this approach produced stable parameter estimates for these specific cases, in general it suffers from two crucial limitations. First, it can be very time-consuming and memory-intensive, especially for networks that are large and sparse and have long relational durations. For example, fitting a STERGM in this way on a network of 10,000 nodes with a relational duration of about 1100 time steps and a mean degree of 0.4 takes several days on a high-performance UNIX cluster and is not amenable to parallelization. Moreover, as is often the case with ERGMs more generally, the model fitting process can be unstable if the starting values are far from the true parameters, and may fail to converge even when the model is good.

In this paper, we give technical justification for an approximation method using parameter estimates from a static ERGM and information on the duration of relationships to generate starting values for the existing algorithm. In practice, we have found that this simple adjustment results in substantially better performance than fitting a STERGM using the traditional approach. Not only does the algorithm generally converge to reasonable parameter estimates when started from the new values, and does so much more quickly, but in many cases the adjusted starting values themselves are an adequate estimate of the STERGM parameters, and it is unnecessary to run the fitting algorithm. The approach requires the same information used for the existing approach—a single cross-sectional network or target cross-sectional statistcs plus mean relational duration—and is found to work best in precisely those cases when STERGM estimation is slowest and most unstable.

The approximation procedure described in this paper was motivated by challenges encountered in two large network simulation projects. Both projects attempt to model the potential effect of interventions to prevent the spread of HIV over sexual networks. We use one of these projects, which models the HIV epidemic among men who have sex with men (MSM) in North and South America and the potential effect of a variety of prevention interventions on those epidemics (Goodreau et al., 2012) as a motivating example in this paper (see Sections 4.2 and 5.2).

Network dynamics can take many forms; in this paper, we concentrate on examples in which the set of actors and their attributes remains fixed while their pairwise relationships can stochastically form and dissolve. The processes underlying relational change itself remain constant, such that there is a stationary distribution for the expected values of network statistics. The approach described in this paper can be easily extended to simulating networks with changing sets of actors, using the approach of Krivitsky et al. (2011).

In Section 2, we give a short description of the separable temporal exponential random graph model. In Section 3, we show that a simple correction to the static ERGM parameters performs well as an approximation to STERGM parameters in the Bernoulli model case. In Section 4, we extend this to dyadic independence models for formation and dissolution, with an application to modeling the spread of HIV among men who have sex with men in the dyadic independence case. Section 5 discusses the utility of the approximation in the dyadic dependence case through simulations and an extension of the application.

2 Brief description of the STERGM

We first review the ERGM framework for *cross-sectional* or static networks, observed at a single point in time. Following the notation of Krivitsky and Handcock (2010), let $\mathbb{Y} \subseteq \{1, \ldots, n\}^2$ be the set of potential relations (dyads) among n nodes, ordered for

directed networks and unordered for undirected. We can represent a network $\mathbf{y}$ as a set of ties, with the set of possible sets of ties, $\mathcal{Y} \subseteq 2^{\mathbb{Y}}$, being the sample space: $\mathbf{y} \in \mathcal{Y}$. Let $\mathbf{y}_{ij}$ be 1 if $(i, j) \in \mathbf{y}$ — a relation of interest exists from i to j — and 0 otherwise.

The network also has an associated covariate array $\mathbf{X}$ containing attributes of the nodes, the dyads, or both. An exponential-family random graph model (ERGM) represents the pmf of $\mathbf{Y}$ as a function of a p-vector of network statistics $g(\mathbf{Y}, \mathbf{X})$, with parameters $\theta \in \mathbb{R}^p$, as follows:

$$\Pr_\theta(\mathbf{Y} = \mathbf{y} \mid \mathbf{X}) = \frac{\exp\{\theta \cdot g(\mathbf{y}, \mathbf{X})\}}{c(\theta, \mathbf{X}, \mathcal{Y})}, \tag{2.1}$$

where the normalizing constant $c(\theta, \mathbf{X}, \mathcal{Y}) = \sum_{\mathbf{y}' \in \mathcal{Y}} \exp\{\theta \cdot g(\mathbf{y}', \mathbf{X})\}$ is a summation over the space of possible networks on n nodes, $\mathcal{Y}$. Where $\mathcal{Y}$ and $\mathbf{X}$ are held constant, as in a typical cross-sectional model, they may be suppressed in the notation. Here, on the other hand, the dependence on $\mathcal{Y}$ and $\mathbf{X}$ is made explicit.

In modeling the transition from a network $\mathbf{Y}^{t-1}$ at time $t-1$ to a network $\mathbf{Y}^t$ at time t, the separable temporal ERGM assumes the formation and dissolution of ties to occur independently from each other within each time step, with each half of the process modeled as an ERGM. For two networks (sets of ties) $\mathbf{y}, \mathbf{y}' \in \mathcal{Y}$, let $\mathbf{y} \supseteq \mathbf{y}'$ if any tie present in $\mathbf{y}'$ is also present in $\mathbf{y}$. Define $\mathcal{Y}^+(\mathbf{y}) = \{\mathbf{y}' \in \mathcal{Y} : \mathbf{y}' \supseteq \mathbf{y}\}$, the networks that can be constructed by forming ties in $\mathbf{y}$; and $\mathcal{Y}^-(\mathbf{y}) = \{\mathbf{y}' \in \mathcal{Y} : \mathbf{y}' \subseteq \mathbf{y}\}$, the networks that can be constructed dissolving ties in $\mathbf{y}$.

Given $\mathbf{y}^{t-1}$, a *formation network* $\mathbf{Y}^+$ is generated from an ERGM controlled by a p-vector of formation parameters θ^+ and formation statistics $g^+(\mathbf{y}^+, \mathbf{X})$, conditional on only adding ties:

$$\Pr\left(\mathbf{Y}^+ = \mathbf{y}^+ \mid \mathbf{Y}^{t-1}; \theta^+\right) = \frac{\exp\{\theta^+ \cdot g^+(\mathbf{y}^+, \mathbf{X})\}}{c\left(\theta^+, \mathbf{X}, \mathcal{Y}^+(\mathbf{Y}^{t-1})\right)}, \quad \mathbf{y}^+ \in \mathcal{Y}^+(\mathbf{y}^{t-1}). \tag{2.2}$$

A *dissolution network* $\mathbf{Y}^-$ is simultaneously generated from an ERGM controlled by a (pos-

sibly different) q-vector of dissolution parameters θ^- and corresponding statistics $g^-(\mathbf{y}^-, \mathbf{X})$, conditional on only dissolving ties from $\mathbf{y}^{t-1}$:

$$\Pr\left(\mathbf{Y}^- = \mathbf{y}^- \mid \mathbf{Y}^{t-1}; \theta^-\right) = \frac{\exp\left\{\theta^- \cdot g^-(\mathbf{y}^-, \mathbf{X})\right\}}{c\left(\theta^-, \mathbf{X}, \mathcal{Y}^-(\mathbf{Y}^{t-1})\right)}, \quad \mathbf{y}^- \in \mathcal{Y}^-(\mathbf{y}^{t-1}). \tag{2.3}$$

The cross-sectional network at time t is then constructed by applying the changes in $\mathbf{Y}^+$ and $\mathbf{Y}^-$ to $\mathbf{y}^{t-1}$: $\mathbf{Y}^t = \mathbf{Y}^{t-1} \cup (\mathbf{Y}^+ \backslash \mathbf{Y}^{t-1}) \setminus (\mathbf{Y}^{t-1} \backslash \mathbf{Y}^-)$.

This transition process is ergodic: a series of networks generated from this process converges to an equilibrium distribution of networks $\Pr(\mathbf{Y}^t = \mathbf{y}; \theta^+, \theta^-)$. To the extent that the observed network $\mathbf{y}$ is a consequence of the long-run evolution of the social process being modeled, it may be modeled as a draw from this equilibrium distribution.

Note that the underlying, network-to-network transition process does not change over time, but rather the set of existing relationships evolves in a manner consistent with the stochastic process as specified by θ^+, θ^-, and the particular statistics that make up $g^+(\mathbf{y}^+, \mathbf{X})$ and $g^-(\mathbf{y}^-, \mathbf{X})$.

Implicit in this model is a Markovian assumption that the survival of a relationship from time t to $t+1$ is, conditional on its existence at t, independent of what happened before t. Also implicit under a dyadic independence dissolution model is a geometric distribution for the relational duration for each tie, with mean given by the reciprocal of the probability of survival for that tie. The probability of survival may vary across dyads; however, under a strict Bernoulli dissolution model, in which all ties have an equal probability $(1-p)$ of survival, relational durations will follow a geometric distribution with mean duration d and probability of dissolution per time step $p = 1/d$. This process can be captured by including a single statistic for the edge count in $g^-(\mathbf{y}^-, \mathbf{X})$; its parameter θ^- should equal the log-odds of a relation surviving the time step (Krivitsky, 2012b), i.e., $\theta^- = \log[(1-p)/p] = \log(d-1)$.

We assume from now on that the average duration(s) are specified and thus that θ^-

is known. This situation would arise, for instance, if we wished to simulate a dynamic network from a cross-sectional observation of a network together with information on the duration of ties, as with Morris et al. (2009).

Recall that we are fitting a dynamic model only to information about a single cross-sectional network, with some tie duration information. In that setting, to the extent that an assumption can be made that the observed network is a draw from the stationary distribution of the STERGM process (which is ergodic), parameter values θ^+ may be estimated using the generalized method-of-moments approach. The generalized method-of-moments estimator (GMME) in the static ERGM is the solution to the equation $E_\theta \{g(\mathbf{Y}, \mathbf{X})\} = g(\mathbf{y}, \mathbf{X})$, where $\mathbf{y}$ is the observed network and $E_\theta \{g(\mathbf{Y}, \mathbf{X})\}$ is the expected value of the network statistics of interest under the stationary distribution induced by θ.

Parameter estimation is accomplished by the following steps. First, a set of target network statistics and initial parameter estimates θ_0^+ are obtained, either by direct specification or through simple estimation methods, such as maximum pseudo-likelihood. The GMME is then found using the procedure of Krivitsky (2012a).

At the core of the algorithm is the estimation of moments of statistics of interest $E_\theta (g(\mathbf{Y}, \mathbf{X}))$ by simulation. If the network evolves slowly, successive networks drawn from the model are likely to be highly autocorrelated. Thus, to obtain a sufficiently precise estimate of $E_\theta (g(\mathbf{Y}, \mathbf{X}))$ we must simulate a very long series of such networks — a time–consuming process.

As is common in ERG modeling, if the starting values θ_0^+ are far from the GMME, the sampled networks are likely to be degenerate, making it next to impossible to estimate the necessary moments. Consequently, the algorithm fails to converge in a feasible timeframe. Thus, we require a value for θ_0^+ that produces networks that are reasonably similar to the target. We introduce one such starting point below in Section 3.

Throughout this paper, we use the **statnet** suite of packages (Handcock et al., 2008) in **R** (R Development Core Team, 2011) for simulation and model fitting.

3 Bernoulli models

We first consider the simplest case, a STERGM with a Bernoulli model for both formation and dissolution of edges. In this case, each process is controlled by a single parameter, giving an equal probability of formation (or dissolution) for all non-edges (or extant edges). We can derive the value and corresponding GMoME ($\hat{\theta}^+$) of the formation parameter θ^+ as a function of network density and relational duration. We wish to justify the following approximation for a particular fixed θ^-:

$$\tilde{\theta}^+ \stackrel{\text{def}}{=} \log\left(\frac{m_1}{1-m_1}\right) - \theta^- = \text{logit}\,(m_1) - \theta^- \approx \hat{\theta}^+, \tag{3.1}$$

where m_1 is the observed cross-sectional density or the density we want to achieve at equilibrium.

The development below depends on $\theta^- \to \infty$, which is reasonable here because θ^- determines average duration of relationships in MCMC time, so we can in practice change the time scale so that this average is long (e.g., equate an MCMC step to a day instead of a month).

Since parameters in the STERGM control either formation or dissolution independently, we can interpret $\text{logit}^{-1}(\theta^+) = e^{\theta^+}(1+e^{\theta^+})^{-1}$ as the expected fraction of null edges that will become ties, while $\text{logit}^{-1}(\theta^-)$ is the expected fraction of existing edges that will survive the time step. Take M to be the (fixed) number of possible edges, and let $m_1(t)$ be the (random) count of extant edges at time t divided by M, or the cross-sectional density at t. Furthermore, $m_0(t) = 1 - m_1(t)$ is the density of non-edges at time t. At equilibrium, the

average number of ties created and destroyed are the same, so

$$E[(1 - \text{logit}^{-1}(\theta^-)) \cdot m_1(t) \cdot M] = E[\text{logit}^{-1}(\theta^+) \cdot m_0(t) \cdot M].$$

Hence,

$$\frac{\text{logit}^{-1}(\theta^+)}{1 - \text{logit}^{-1}(\theta^-)} = \frac{E[m_1(t)]}{E[m_0(t)]} = \frac{E[m_1(t)]}{1 - E[m_1(t)]} = \frac{\mu_1}{1 - \mu_1},$$

where μ_1 is the expected cross-sectional density of the network. Thus, the formation parameter we wish to estimate and its corresponding GMoME are given by

$$\theta^+ = -\log\left(\frac{1 + e^{\theta^-}}{\frac{\mu_1}{1-\mu_1}} - 1\right) \quad \text{and} \quad \hat{\theta}^+ = -\log\left(\frac{1 + e^{\theta^-}}{\frac{m_1}{1-m_1}} - 1\right), \tag{3.2}$$

respectively, where m_1 is defined as in Equation (3.1). Recall from Section 2 that the dissolution parameter θ^-, as a simple transformation of the average relational duration, is assumed known.

3.1 Derivation of the approximation

To verify Approximation (3.1), we assume that as the network size M increases, the density goes to zero, which means that $\theta^- \to \infty$. This assumption is justified by the fact that for most social phenomena, mean degree is likely to stay roughly constant even as the network grows (Krivitsky et al., 2011). For example, people living in a town of 10,000 should have roughly the same numbers of relationships per person as those in a city of 10 million, and not three orders of magnitude fewer. Combining Equations (3.1) and (3.2), we find that

$$\tilde{\theta}^+ - \hat{\theta}^+ = \log\left(1 + e^{-\theta^-}\left[\frac{1 - 2m_1}{1 - m_1}\right]\right), \tag{3.3}$$

which clearly tends to zero as $\theta^- \to \infty$ as long as m_1 stays bounded away from 1. Since $m_1 = 1$ with probability tending to zero as $M \to \infty$, we obtain the following result:

Proposition 1. *For a Bernoulli formation and Bernoulli dissolution model,* $\left(\tilde{\theta}^+ - \hat{\theta}^+\right) \xrightarrow{\text{P}} 0$ *as* $M \to \infty$ *and* $\theta^- \to \infty$.

3.2 Accuracy of approximation

Figure 1 shows the difference between the approximation $\tilde{\theta}^+$ and the GMoME $\hat{\theta}^+$ for densities ranging from 0 to 0.5 and durations ranging from 2 to 25 time steps. We can see that the accuracy of the approximation improves extremely rapidly with duration and verify that when $\mu_1 = 0.5$, approximation (3.1) is exact. For any network density the approximation error is less than 0.2 for any duration longer than 6 time units, under 0.1 for durations over 11 time units and under 0.05 for durations over approximately 20 time units. Recall that time units here are not tied to calendar time, so a sufficiently long mean relational duration can be achieved by changing the scale on which relational duration is measured. If we want relationships to last on average one year, we could use the analytic approximation with confidence if we parameterized durations in terms of months for model fitting and simulation. This is limited only by computing resources: as each discrete time step represents a smaller and smaller amount of time, the number of time steps that need to be simulated to advance a given time period increases.

Recall that in the proof we require both that duration be sufficently long and network density be sufficiently small. In fact, as can be seen in Figure 1, it appears that for lower density a slightly longer duration is required to achieve a good approximation using this method. This could make the method less useful in very low-density situations, unless the average relational duration is long. While this result may seem counter-intuitive at first, it is understandable when we recall that there are two opposing trends with regard

to the density of the network. On the one hand, the closer the density is to 0.5, the closer the approximation is to the analytic result regardless of duration. On the other hand, we have an asymptotic argument that the approximation approaches the analytic result as the duration increases and density decreases. The figure reveals the reduced need for asymptotics as the density increases, balanced against the increasing validity of the asymptotic argument as the density decreases.

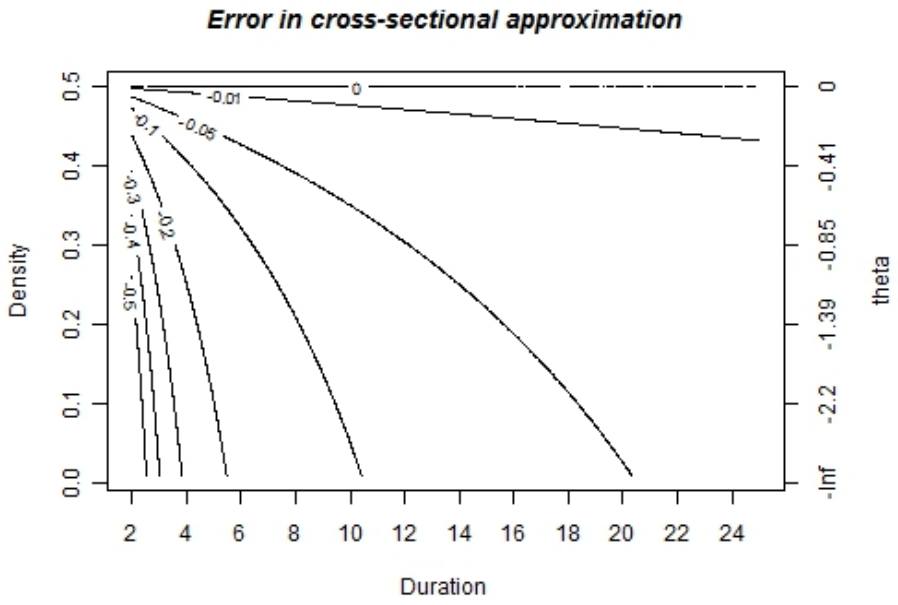

Figure 1: Contour plot of approximation error $|\hat{\theta}^{+} - \tilde{\theta}|$ in a Bernoulli formation and dissolution model with network density less than 0.5.

4 Dyadic independence models

In this section, we extend the above results to models with dyadic independence terms in the formation and dissolution models. In a cross-sectional ERGM, dyadic independence means that the states of all dyads are stochastically independent (Hunter et al., 2008b), but we do not assume independence across time steps. Note that we are assuming the

network consists of undirected relations. We can extend this easily to directed networks, in which case we are not assuming true dyadic independence, but rather independence of the random variables for each directed edge.

4.1 Derivation of the approximation

We now consider STERGMs with dyadic independence models for formation and dissolution. Under dyadic independence, the vectors of network statistics are

$$g^+(\mathbf{y}^+, \mathbf{X}) = \sum_{(i,j)\in\mathbb{Y}} y_{ij}^+ W_{ij}^+ \quad \text{and} \quad g^-(\mathbf{y}^-, \mathbf{X}) = \sum_{(i,j)\in\mathbb{Y}} y_{ij}^- W_{ij}^-, \tag{4.1}$$

where $W_{ij}^+ \equiv W_{ij}^+(\mathbf{X})$ and $W_{ij}^- \equiv W_{ij}^-(\mathbf{X})$ are arbitrary vectors of attributes for the potential edge (i, j) of length p and q, respectively, and y_{ij}^+ (y_{ij}^-) is the indicator that this edge is contained in $\mathbf{y}^+$ ($\mathbf{y}^-$). It is common that the W_{ij}^* may be written as some function of nodal covariates defined for i and j. For instance, the Euclidean distance between nodes i and j may be expressed as a function of their location vectors. In general, however, W_{ij}^* may include entries that cannot be expressed in this way (e.g., some non-Euclidean "distances" such as shortest road distance or average driving time between locations i and j).

We make the additional assumptions that $q \leq p$ and the first q elements of W_{ij}^+ are simply W_{ij}^- for all i and j. This assumption may be justified theoretically because in many processes we are interested in modeling, the formation of relationships is more complicated than their dissolution. For example, the creation of spousal relationships through marriage is a complex matching problem, but their dissolution is simplified because there is only one possible tie to dissolve. In the examples used in this paper, the distribution of relationship lengths can be reasonably modeled by a simple geometric distribution—i.e., a Bernoulli dissolution model. Krivitsky and Handcock (2010) discuss this issue further.

Now we consider the case in which W_{ij}^+ is categorical, taking a total of K unique values

as (i, j) ranges over all M possible edges, where we assume that $K \ll M$. Let M_k be the number of possible edges in category k, $k = 1, \ldots, K$, and let $D_k \subset \mathbb{Y}$ denote the set of all such edges. Then if we define $\lambda_k^+ = \theta^+ \cdot W_{ij}^+$ for $(i, j) \in D_k$, we obtain

$$\theta^+ \cdot g^+(\mathbf{y}^+, \mathbf{X}) = \sum_{k=1}^{K} \lambda_k^+ \sum_{(i,j) \in D_k} y_{ij}^+,$$

and so Model (2.2) may be factored into K independent sub-models:

$$\Pr\left(\mathbf{Y}^+ = \mathbf{y}^+ \mid \mathbf{Y}^{t-1}; \theta^+\right) = \frac{1}{c^+} \prod_{k=1}^{K} \exp\left\{ \lambda_k^+ \sum_{(i,j) \in D_k} y_{ij}^+ \right\}, \quad \mathbf{y}^+ \in \mathcal{Y}^+(\mathbf{y}^{t-1}). \tag{4.2}$$

Since each of these sub-models is merely a Bernoulli model with parameter λ_k^+, we may follow the same logic of Section 3 to prove a result similar to Proposition 1: For $k = 1, \ldots, K$, let us define m_{1k} to be the observed cross-sectional density among potential edges in the kth category, i.e., the count of extant edges in D_k divided by M_k, or alternatively the density we wish to achieve at equilibrium. Also define $\lambda_k^- = \theta_k^- \cdot W_{ij}^-$, where $(i, j) \in D_k$, for the (known) dissolution parameter in the kth category. (Not all of the λ_k^- values must be different, but since the categories are determined by the formation covariates, which include the dissolution covariates, we do know that all $(i, j) \in D_k$ will give the same value of $\theta_k^- \cdot W_{ij}^-$.) Furthermore, let $\tilde{\lambda}_k^+ = \text{logit}(m_{1k}) - \lambda_k^-$ and

$$\hat{\lambda}_k^+ = -\log\left(\frac{1 + e^{\lambda_k^-}}{\frac{m_{1k}}{1 - m_{1k}}} - 1 \right). \tag{4.3}$$

Then the following proposition justifies the approximation of $\hat{\lambda}_k^+$ by $\tilde{\lambda}_k^+$. Recall from Section 3 that we require $M_k \to \infty$ in order to ensure that m_{1k} is bounded away from 1.

Proposition 2. *Suppose that the formation and dissolution models have p- and q-dimensional statistics given by Equation (4.1) where $q \leq p$ and the first q elements of W_{ij}^+ are equal to*

W_{ij}^- *for all* $(i,j) \in \mathbb{Y}$. *For* $k = 1, \ldots, K$, *if* $M_k \to \infty$ *and* $\lambda_k^- \to \infty$, *then* $\left(\tilde{\lambda}_k - \hat{\lambda}_k^+\right) \xrightarrow{\mathrm{P}} 0$.

Since $\text{logit}(m_{1k})$ is undefined when $m_{1k} = 0$ or $m_{1k} = 1$, the preceding theoretical development suggests that the subsets D_k of constant covariate values should be large enough so that these two extreme values of the density are essentially never observed in practice. Yet the elegance of this approach lies in the fact that it may be implemented without actually calculating $\text{logit}(m_{1k})$ for every k, as follows: Since we make the assumption that the first q statistics in the W_{ij}^+ vector are equal to W_{ij}^-, regardless of the values of i and j, then we find for $(i,j) \in D_k$ that

$$\lambda_k^+ - \lambda_k^- = \left[\theta^+ \cdot W_{ij}^+\right] - \left[\theta^- \cdot W_{ij}^-\right] = \left[\theta^+ - (\theta^-, 0, \ldots, 0)\right] \cdot W_{ij}^+. \tag{4.4}$$

(Recall that θ^- is a sub-vector of θ^+ in this scenario.) In other words, when $W_{ij}^+ \supseteq W_{ij}^-$ our approach may be implemented simply by following these steps:

1. Calculate the maximum likelihood estimator, say $\hat{\theta}$, for a cross-sectional ERGM for the observed network using sufficient statistics W_{ij}^+;
2. subtract θ^- from the components of $\hat{\theta}$ corresponding to the dissolution statistics;
3. call the resulting vector $\tilde{\theta}^+$ and take $\tilde{\theta}^+$ as the approximation to the GMoME $\hat{\theta}^+$.

Note that a special case of this is a STERGM with a dyadic independence model for formation and Bernoulli dissolution. In this case, θ^- is a scalar, and if we assume that W_{ij}^+ includes the constant statistic equal to unity we can implement the approximation simply be subtracting θ^- from the component of $\hat{\theta}$ corresponding to the constant 1 statistic. This statistic controls for the overall edge count of the network, which means it is extremely commonly used in ERGMs, for much the same reason that a standard regression model generally includes an intercept term.

Although the theoretical development leading to Proposition 2 relies on W_{ij}^+ being categorical, the implementation just outlined does not. Indeed, we find that in practice, this method works well for arbitrary W_{ij}^+. We explore an application in which the W_{ij}^+ are continuous in Section 4.2. A further open question, which we explore in Section 5, is whether a similar technique applies even when the assumption of dyadic independence is violated.

4.2 Application: Modeling HIV spread among MSM in the US

To demonstrate the use of this approximation method in practice, we discuss a large modeling project examining the potential impact of a variety of behavioral and biomedical interventions that are now or are soon to be available for men who have sex men (MSM) in the United States.

In the US and other developed nations, HIV remains concentrated among MSM, with well over half of all new HIV diagnoses within this community (Hall et al., 2008). Rates of HIV incidence are on the rise among young MSM, and especially among young Black and Latino MSM (Prejean et al., 2011). The work reported here is a piece of the baseline work for a larger NIH-funded project, called PUMA (Prevention Umbrella for MSM in the Americas). Part of the MP3 (Methods for Prevention Packages Program), the goal of this work is to better assess the ways in which existing and imminent interventions can be combined, packaged and tailored for greater efficacy. As a first step in this work, the project is modeling a wide variety of possible tailored combination interventions; here we discuss only the baseline model.

To estimate and simulate models of HIV transmission among MSM we use a separable temporal ERGM to model main partnerships, in order to represent realistic partnership dynamics. A main partnership here is defined as any relationship in which men feel a close emotional connection and thus includes many short, dating-type relationships and a smaller

number of long-term committed partnerships. Casual contacts are modeled using a cross-sectional ERGM not discussed here. To demonstrate the performance of the approximation for dyadic independence models, we begin by specifying the partnership formation part of the ERGM with an edge count term to control density, an individual attribute of race to allow for differential rates of partnership formation across races, and partnership-level matching on race, age and preferred sexual role (strictly insertive, strictly receptive, or versatile). The formation model is thus

$$\log\left\{\Pr\left(\mathbf{Y}^{+}=\mathbf{y}^{+}\mid\mathbf{Y}^{t-1};\theta^{+}\right)\right\}\propto\theta^{+}(e)\cdot e+\sum_{r}\theta^{+}(u_{r})\cdot u_{r}+\sum_{r}\theta^{+}(m_{r})\cdot m_{r}$$
$$+\theta^{+}(a)\sum_{k<l}|(\sqrt{\text{age}(k)}-\sqrt{\text{age}(l)})|+\sum_{c=r,i}\theta^{+}(m_{c})\cdot m_{c},$$

where y_{ij} = the pair of persons i and j, and $y_{ij}=1$ indicates they are partners; $\mathbf{Y}^{c}_{ij}$ = the rest of the pairs in the network, excluding the y_{ij} pair; e = total number of partnerships of all types in the network; u_r = the number of partnerships of persons of race r; m_r = the number of partnerships with both partners of race r; m_c = the number of partnerships with both partners of role class c (which can take values i = strictly insertive or r = strictly receptive); k and l represent the actors in each main partnership. The parameters for matching on sexual role are fixed at negative infinity to enforce a prohibition on sexual encounters between either two strictly insertive men or two strictly receptive men.

The model for dissolution of partnerships is a simple Bernoulli model with parameter determined by the average partnership duration. The mean duration is approximately 3.1 years for all main partnerships. This, together with the fact that the model is stepped forward day by day, yields a dissolution parameter $\theta^{-}=\log(3.1\cdot 365-1)=7.0$.

Estimating the formation parameters for this model on a network of 10,000 nodes using the traditional approach outlined in Section 1 is extremely computationally intensive—on the order of days to weeks. We use instead the adjustment to the static ERGM fit discussed

above as parameter estimates. We simulate a network forward in time using the parameters generated and compare the distribution of simulated network statistics to the target values taken from the observed network.

Figure 2 shows the standardized deviations of the simulated cross-sectional network statistics from their target values over the evolution of the network for 10,000 time steps. Counts of the total number of partnerships of nodes of a certain type are referred to as "node factor" terms, while counts of edges with both nodes of the same type are referred to as "node match" terms. Note that the node match terms for role class are excluded from the plot since they are coerced to be zero at all times. Since the simulation follows all steps of the MCMC chain without thinning, we expect to see some autocorrelation and "walks" away from the target over time, however, the observed statistics at each cross-section are largely within two standard deviations of the target. This indicates that the approximation has yielded reasonable estimates of the model parameters.

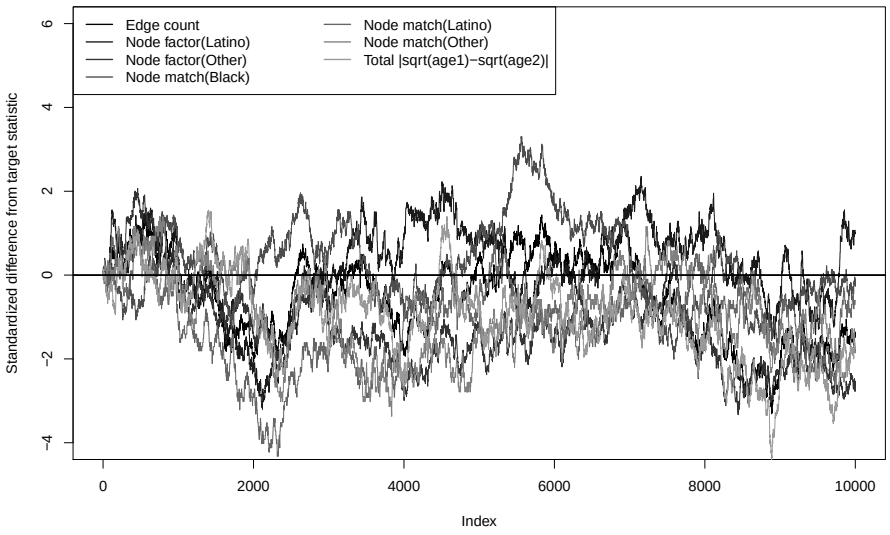

Figure 2: Standardized deviations of cross-sectional network statistics from target statistics over 10,000 time steps under dyad independence model.

We expect, given the long duration and the analytic results above, that the approximate

parameter estimates would be adequate in this case, where the models for formation and dissolution are both dyadic independence models.

5 Dyadic dependence formation models

Of course, in practice we are more often interested in models that do not fit the dyadic independence framework. An explicit derivation of the performance of the approximation in this case is not available, so instead we present here a simulation study examining the use of the approximation method developed above in a case where the formation model includes dyad dependent terms and an application of such a model to real data. The simulation has two parts: first, we compare the values of the parameter estimates from full STERGM fits to those generated using the approximation. We then simulate networks from the approximate values and examine how closely they match the network statistics from the observed network over time. In section 5.2 we re-examine the application to modeling the spread of HIV among MSM using a dyadic dependence model.

5.1 Simulation study

We began by simulating undirected networks of 1000 nodes with mean degree of 1.2, but with varying degree distributions. A mean degree of 1.2 allows for a wide range of degree distributions with the majority of the weight on 0, 1, or 2 ties. For all simulated networks, 30% of the nodes had exactly one incident edge, and the percent with two edges was set for different networks at 10, 15, 20, 25, 30, 35, 40 and 45%. In a strictly Bernoulli network with mean degree 1.2, we would expect approximately 22% of the nodes to have degree 2 and about 36% to have degree 1. We consider networks with percentages further from 22% to be in a sense "more dependent" as the model departs further from the independent Bernoulli case, and expect that these networks will be less well fit by the approximation.

Note that density in this case is constant over all simulations, and there is no constraint on degrees other than 1 and 2. We examine the effect of duration by taking the average duration of ties to be d =5, 10, 15 or 20 time steps.

For each of the 32 scenarios, we simulated 50 network datasets and fit a cross-sectional ERGM to each with terms for the edge count, number of nodes with degree 1, and number of nodes with degree 2. Note that the latter two terms are dyad dependent. We next calculated the approximation estimate of the STERGM parameters by subtracting $\gamma = \log(d-1)$ from the edges coefficient, as discussed in Section 4.1 for dyadic independence models. Using these values as a starting point for the Robbins-Monro algorithm, we then ran a dynamic model fit with the same terms as the cross-sectional model, and compared the resulting estimates to the approximated values.

Figure 3 gives the mean of the differences between the approximated and estimated STERGM parameters. We see that the differences are largest for short duration and very high percentage of nodes with degree 2. The former is to be expected given the asymptotic nature of our results, and the latter indicates that models that depart more from dyadic independence are indeed less well fit by the approximation methods. In general, the magnitudes of the differences for degree terms are under 0.1 for models with at least moderate duration and degree distributions not too extreme. The edge count term exhibits differences of a slightly larger magnitude, but generally under 0.3.

The question remains, however, whether the observed differences are large enough to impact the characteristics of networks simulated from the model, which can be used either as a measure of goodness-of-fit of the model or for simulation studies. To examine this, we simulated networks for 5000 time steps from a randomly selected vector of the approximated parameter estimates. Figure 4 shows the percent difference between the mean network statistics from the last 4000 time steps (the first 1000 were discarded as burn-in) and the target statistics for each network.

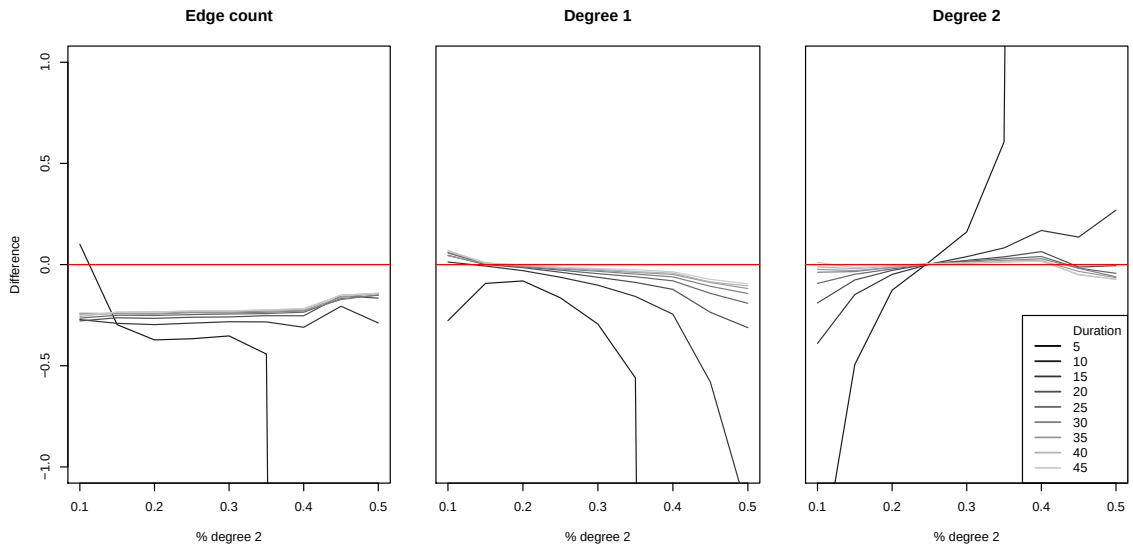

Figure 3: Average differences between approximated model coefficients and fitted coefficients over 50 simulations under dyad dependence model.

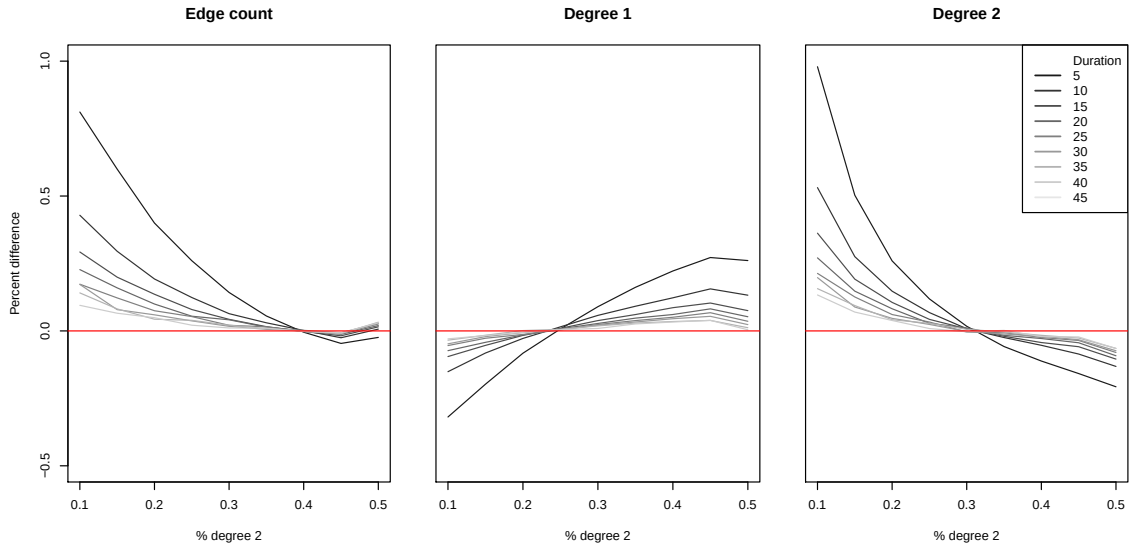

Figure 4: Percent deviations of cross-sectional network statistics from target statistics over 4,000 time steps under dyad dependence model.

We can see that the bias in the cross-sectional network statistics decreases consistently with longer duration, regardless of the percent of nodes that are of degree 2, as we would expect given the asymptotic results for dyad independence models. On the other dimension, we see the bias tending to increase as the percent of nodes of degree 2 moves further from what we would expect under a Bernoulli model. Thus, the approximation can be quite bad for short durations or strong dependence induced by very non-Bernoulli degree distributions, but for the majority of the scenarios the differences are rather small. It also seems likely that in practice few applications would have durations even as short as those we use in this simulation, and often much longer, making it more likely that the use of the approximation method will be satisfactory.

In a case where the approximation is not statisfactory, however, using the estimates from the approximation results in radically improved fitting time for the STERGM. For example, in the case from these simulations with a mean duration of 5 time steps and percent degree 2 of 35%, we see that the simulated network statistics using the approximated values are off by 20–30%, so we would likely not want to use the approximation itself as our parameter estimates. If we run a traditional STERGM estimation approach for this case, it takes over five hours to complete the fitting process. If, however, we use the approximated values as the starting point for the fitting algorithm, the fit completes in 1.7 minutes.

5.2 Application: Modeling the spread of HIV among MSM in the US, revisited

In Section 4.2, we demonstrated the use of the adjustment to the static ERGM fit as model parameters in a dyadic independence model. In reality, the degree distribution is of vital importance when modeling sexual networks, particularly when modeling the spread of disease over said networks. We would like to add a term for the number of men in two

simultaneous relationships to control for the tendency for or against forming concurrent partnerships.

The new formation model is then given by

$$\log\left\{\Pr\left(\mathbf{Y}^{+}=\mathbf{y}^{+}\mid\mathbf{Y}^{t-1};\theta^{+}\right)\right\}\propto\theta^{+}(e)\cdot e+\sum_{r}\theta^{+}(u_r)\cdot u_r+\sum_{r}\theta^{+}(m_r)\cdot m_r+\theta^{+}(d_2)\cdot d_2$$
$$+\theta^{+}(a)\sum_{k<l}|(\sqrt{\text{age}(k)}-\sqrt{\text{age}(l)})|+\sum_{c=r,i}\theta^{+}(m_c)\cdot m_c,$$

where we have added a term d_2 = the number of actors in exactly two main partnerships at a given time. Note that this is a dyad dependent term. The model also enforces a constraint of no more than two main partnerships at a time, another form of dependence.

Figure 5 shows the standardized deviations of the simulated cross-sectional network statistics from their target values over the evolution of the network for 10,000 time steps using the dyadic dependence model. The parameter values used for network simulation here are those given by the approximation to the full dynamic model fit given in Section 4.1. In this case, the simulated network including the dyad dependent degree 2 term has fewer extreme deviations from the target statistics than were observed in the dyad independence case. This should not be over-interpreted, since each simulation is only one realization of a stochastic process. We can say, however, that the network evolution over time is consistent with the cross-sectional data used to generate the target network statistics when using the approximation in place of a full dynamic model fit.

6 Conclusion

This paper introduces a useful approximation method to generate estimates of the STERGM dynamic model parameters, which is much less computationally intensive than the full algorithm used for estimation in the dynamic model. We assume that dissolution parameters are

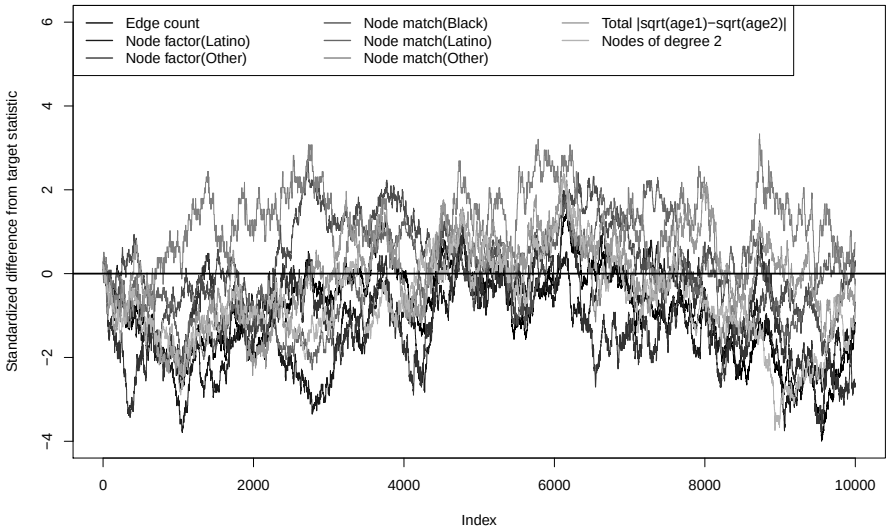

Figure 5: Standardized deviations of cross-sectional network statistics from target statistics over 10,000 time steps under dyad dependence model.

fixed as a function of the average durations of relationships, and approximate the STERGM formation parameters by subtracting these dissolution parameters from the corresponding elements of the vector of parameter estimates obtained by estimating the formation model with a cross-sectional ERGM. Even in cases where the approximation itself is inadequate as an estimator, using the approximation estimates as a starting point for the algorithm greatly increases the likelihood of convergence in the estimation procedure. The proofs given in this paper assume dyadic independence models and that the model terms are categorical, but as we see in the examples and in practice, the approximation works well even for non-categorical terms and dependence models in many cases. This work might therefore be a starting point for further theoretical and empirical exploration of this method.

The asymptotic nature of the results suggests that they are most useful for duration sufficiently long and density sufficiently small. In one sense, it is quite easy to deal with the issue of duration; since duration here refers to MCMC steps, we could in theory rescale the relational duration so that one MCMC step equals an arbitrarily small unit of real time.

An additional advantage of equating an MCMC step to a small unit of time is improved plausibility of the assumtion of separability of the formation and dissolution processes within MCMC step. This poses computational difficulties, however, if the model is to be used to simulate dynamically evolving networks, as it increases the number of simulation steps needed to produce a particular duration proportionately. In some cases, such as the example used in the application, there may be an external reason why equating an MCMC step to a particular time unit is necessary. Given the likelihood of computational or other constraints, further work is needed to ascertain in exactly which cases it is appropriate to use the approximation method, and when it is advisable to fit the STERGM.

As we demonstrate in the HIV modeling examples, it is important when using this approximation to perform model diagnostics by simulating one or more dynamic networks from the model parameters and checking to see that the cross-sectional network statistics are varying stochastically about the target statistics. This is particularly true if one wishes to use the approximation as a final parameter estimate instead of merely a starting point for a full dynamic model fit.

The initial development of ERGMs offered enormous promise as a generalized framework for the statistical modeling of cross-sectional social networks. However, years of research and development were needed to identify and overcome a variety of issues to ensure that their practical application lived up to this promise. The recent proposed STERGM class of models offers similar promise for dynamic social networks. Since STERGMs build upon ERGMs, there is reason to hope that most of the issues that will arise in their application will parallel those that have already been investigated, and thus will be quick to resolve. At least one new issue has already arisen, however—the particularly high computational burden needed to fit these models under certain conditions—which this paper explains and then identifies a solution to. We hope that resolving this issue will position STERGMs to be of as great general applicability and usefulness as ERGMs have proven. Our concurrent

and successful application of STERGMs in a number of ongoing research projects, made possible by the work laid out in the paper, suggest to us that this will indeed be the case.

Acknowledgments

This work is supported by NIH grants R01–GM083603, R01–HD068395, and R01–AI083060. We are grateful to Martina Morris, Skye Bender-deMoll, and the rest of the University of Washington network modeling team for helpful discussions.

References

Aral, S., Muchnik, L., and Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 106(51):21544–21549.

Doreian, P. and Stokman, F. N., editors (1997). *Evolution of social networks*. Routledge.

Frank, O. (1991). Statistical analysis of change in networks. *Statistica Neerlandica*, 45(3):283–293.

Goodreau, S., Cassels, S., Kasprzyk, D., Montao, D., Greek, A., and Morris, M. (2010). Concurrent partnerships, acute infection and hiv epidemic dynamics among young adults in zimbabwe. *AIDS and Behavior*, pages 1–11. 10.1007/s10461-010-9858-x.

Goodreau, S. M., Carnegie, N. B., Vittinghoff, E., Lama, J. R., Sanchez, J., Grinsztejn, B., Koblin, B. A., Mayer, K. H., and Buchbinder, S. P. (2012). What drives the US and Peruvian HIV epidemics in men who have sex with men (MSM)? In *Poster #MOPE143 presented at the XIX International AIDS Conference, Washington DC.*

Hall, H. I., Song, R., Rhodes, P., Prejean, J., An, Q., Lee, L. M., Karon, J., Brookmeyer, R., Kaplan, E. H., McKenna, M. T., and Janssen, R. S. (2008). Estimation of HIV incidence in the United States. *JAMA*, 300(5):520–529.

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., and Morris, M. (2008). statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software*, 24(1):1–11.

Hanneke, S., Fu, W., and Xing, E. P. (2010). Discrete temporal models of social networks. *Electroninc Journal of Statistics*, 4:585–605.

Holland, P. W. and Leinhardt, S. (1977). A dynamic model for social networks. *Journal of Mathematical Sociology*, 5(1):5–20.

Hunter, D. R., Goodreau, S. M., and Handcock, M. S. (2008a). Goodness of fit of social network models. *Journal of the American Statistical Association*, 103(481):248–?258.

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., and Morris, M. (2008b). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, 24(3):1–29.

Jack, S., Moult, S., and Anderson, A. R. (2010). An entrepreneurial network evolving: Patterns of change. *International Small Business Journal*, 28(4):315–337.

Krivitsky, P. N. (2009). *Statistical models for social network data and processes*. PhD thesis, University of Washington.

Krivitsky, P. N. (2012a). Modeling of dynamic networks based on egocentric data with durational information. Technical Report 2012-01, Pennsylvania State University Department of Statistics.

Krivitsky, P. N. (2012b). Modeling tie duration in ergm-based dynamic network models. Technical Report 2012-02, Pennsylvania State University Department of Statistics.

Krivitsky, P. N. and Handcock, M. S. (2010). A separable model for dynamic networks. *arXiv:1011.1937v1*.

Krivitsky, P. N., Handcock, M. S., and Morris, M. (2011). Adjusting for network size and composition effects in exponential family random graph models. *Statistical Methodology*, 8(4):319–339.

Mercken, L., Snijders, T. A., Steglich, C., and de Vries, H. (2009). Dynamics of adolescent friendship networks and smoking behavior: Social network analyses in six european countries. *Social Science and Medicine*, 69(10):1506–1514.

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. *American Journal of Public Health*, 99:1023–1031.

Prejean, J., Song, R., Hernandez, A., Ziebell, R., Green, T., Walker, F., Lin, L. S., An, Q., Mermin, J., Lansky, A., Hall, H. I., and HIV Incidence Surveillance Group, . (2011). Estimated HIV incidence in the United States, 2006-?2009. *PLoS ONE*, 6(8):e17502.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Robins, G. and Morris, M. (2007). Advances in exponential random graph (p*) models. *Social Networks*, 29(2):169–?172.

Robins, G. and Pattison, P. (2001). Random graph models for temporal processes in social networks. *Journal of Mathematical Sociology*, 25(1):5–41.

Sasovova, Z., Mehra, A., Borgatti, S. R., and Schippers, M. C. (2010). Network churn: The effects of self-monitoring personality on brokerage dynamics. *Administrative Science Quarterly*, 55(4):639–670.

Snijders, T. A. (2001). The statistical evaluation of social network dynamics. *Sociological Methodology*, 31(1):361–?395.

Snijders, T. A., Pattison, P. E., Robins, G. L., and Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology*, 36(1):99–?153.

Snijders, T. A., Steglich, C., and Schweinberger, M. (2007). Modeling the co-evolution of networks and behavior. In Van Montfort, K., Oud, H., and Satorra, A., editors, *Longitudinal models in the behavioral and related sciences*, pages 41?–71. Lawrence Erlbaum, Mahwah, NJ.

Strauss, D. and Ikeda, M. (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, 85(409):204?–212.

Wasserman, S. and Pattison, P. (1996). Logit models and logistic regressions for social networks. I. An introduction to Markov graphs and p*. *Psychometrika*, 60:401–?425.

Weerman, F. M. (2011). Delinquent peers in context: A longitudinal network analysis of selection and influence effects. *Criminology*, 49(1):253–286.