A Separable Model for Dynamic Networks

Pavel N. Krivitsky and
Pennsylvania State University, University Park, USA

Mark S. Handcock
University of California at Los Angeles, Los Angeles, USA

Summary

Models of dynamic networks — networks that evolve over time — have manifold applications. We develop a discrete-time generative model for social network evolution that inherits the richness and flexibility of the class of exponential-family random graph models. The model — a Separable Temporal ERGM (STERGM) — facilitates separable modeling of the tie duration distributions and the structural dynamics of tie formation. We develop likelihood-based inference for the model, and provide computational algorithms for maximum likelihood estimation. We illustrate the interpretability of the model in analyzing a longitudinal network of friendship ties within a school.

Keywords

1. Introduction

Relational phenomena occur in many fields and are increasingly being represented by networks. There is a need for realistic and tractable statistical models for these networks, especially when the phenomena evolves over time. For example, in epidemiology there is a need for data-driven modeling of human sexual relationship networks for the purpose of modeling and simulation of the spread of sexually transmitted disease. As Morris and Kretzschmar (1997) show, spread of such disease is affected not just by the momentary number of partnerships, but their timing. To that end, the models used must have realistic temporal structure as well as cross-sectional structure.

Holland and Leinhardt (1977), Frank (1991), and others describe continuous-time Markov models for evolution of social networks. (See Doreian and Stokman (1997) for a review.) The most popular parametrisation is the *actor-oriented* model described by Snijders (2005) and Snijders et al. (2010), which can be viewed in terms of actors making decisions to make and withdraw ties to other actors. This model was then extended by Snijders et al. (2007) to jointly model actors' network-related choices ("selection") and the effects of neighboring actors on each other's attributes ("inuence").

Exponential-family random graph models (ERGMs) for social networks are a natural way to represent dependencies in cross-sectional graphs and dependencies between graphs over time, particularly in a discrete context. Robins and Pattison (2001) first described this

Address for correspondence: Pavel N. Krivitsky, Department of Statistics, Penn State University, University Park, PA 16802, U.S.A., krivitsky@stat.psu.edu.

approach. Hanneke and Xing (2007) and Hanneke et al. (2010) also define and describe a *Temporal ERGM (TERGM)* ("Discrete Temporal ERGM" in the 2007 publication), postulating an exponential family model for the transition probability from a network at time t to a network at time $t + 1$.

Most of the attention in modeling of dynamic networks has focused on fitting the model to a network series (Snijders, 2001; Hanneke and Xing, 2007; Hanneke et al., 2010) or an enumeration of instantaneous events between actors in the network (Butts, 2008). In the former case, the dyad census of the network of interest is observed at multiple time points. In the latter case, each event of interest and its exact time of occurrence is observed.

A primary issue in modeling dynamic networks that has received limited attention is that of attribution of prevalence. A snapshot of a network at a single time point provides information about *prevalence* of the network properties of interest — such as the total number of ties — as opposed to properties of a dynamic network process that has produced it: *incidence* — the rate at which new ties are formed — and *duration* — how long they tend to last once they do. Multiple snapshots over the same set of actors (panel data) contain information about incidence and duration, but, as we show below, the model parametrisations presently in use do not allow convenient control over this attribution of prevalence.

In Section 2, we review discrete-time ERGM-based network models, and in Section 3, we extend these network models to provide a more interpretable and convenient parametrisation that separates incidence from duration. In Section 4, we develop conditional maximum likelihood estimators (CMLE) based on regularly-spaced network series data by extending the approach of Hunter and Handcock (2006). In Section 5, we illustrate the methodology with application to a longitudinal network of friendship ties within a school. In Section 6, we consider some extensions that the model framework suggests and allows.

2. Discrete-Time ERGM-Based Models for Network Evolution

We first consider a discrete-time dynamic network model in which the network at time t is a single draw from an ERGM conditional on the network at time $t - 1$ (and possibly time $t - 2$, etc.), extending the Temporal ERGM (TERGM) of Hanneke and Xing (2007) and Hanneke et al. (2010). In this section we specify the model and discuss its fundamental properties.

2.1. Model Definition

Suppose that N is the set of $n = |N|$ actors of interest, labeled $1, …, n$, and let $\mathbb{Y} \subseteq N \times N$ be the set of potential ties among them — with pairs $(i, j) \in \mathbb{Y}$ ordered for directed and unordered for undirected networks — and let $\mathcal{Y} \subseteq 2^{\mathbb{Y}}$ be the set of possible networks of interest formed among these actors. For a network realization $\mathbf{y} \in \mathcal{Y}$ define $\mathbf{y}_{i,j}$ to be an indicator of a tie from actor i to actor j, and further let $\mathbf{y}_{i,\cdot}$ be the set of actors to whom i has a tie, $\mathbf{y}_{\cdot,j}$ the set of actors who have ties to j, and $\mathbf{y}_i$ the set of actors with undirected ties with i. Let $\mathbf{Y}^t \in \mathcal{Y}$ be a random variable representing the state of the network at the discrete time point t and $\mathbf{y}^t \in \mathcal{Y}$ be its realization.

Following Hunter and Handcock (2006), let $\theta \in \mathbb{R}^q$ be a vector of q model parameters, and let $\eta(\theta) : \mathbb{R}^q \to \mathbb{R}^p$ be a mapping from θ to natural parameters $\eta \in \mathbb{R}^p$, with $q \leq p$. Let $\mathbf{g} : \mathcal{Y}^2 \to \mathbb{R}^p$ be the sufficient statistic for the transition from network $\mathbf{y}^{t-1}$ at time $t - 1$ to network $\mathbf{y}^t$ at time t. The one-step transition probability from $\mathbf{y}^{t-1}$ to $\mathbf{y}^t$ is then defined to be

$$\Pr_{\eta,\boldsymbol{g}}(\boldsymbol{Y}^t{=}\boldsymbol{y}^t \,|\, \boldsymbol{Y}^{t-1}{=}\boldsymbol{y}^{t-1};\theta){=}\frac{\exp\left(\eta(\theta)\cdot\boldsymbol{g}(\boldsymbol{y}^t,\boldsymbol{y}^{t-1})\right)}{c_{\eta,\boldsymbol{g}}(\theta,\boldsymbol{y}^{t-1})}, \boldsymbol{y}^t,\boldsymbol{y}^{t-1}\in\mathcal{Y}, \quad (1)$$

or, with a *k*-order Markov assumption, and letting $\boldsymbol{g}:\mathcal{Y}^{k+1}\rightarrow\mathbb{R}^p$,

$$\Pr_{\eta,\boldsymbol{g}}(\boldsymbol{Y}^t{=}\boldsymbol{y}^t \,|\, \boldsymbol{Y}^{t-1}{=}\boldsymbol{y}^{t-1},\ldots,\boldsymbol{Y}^{t-k}{=}\boldsymbol{y}^{t-k};\theta){=}\frac{\exp\left(\eta(\theta)\cdot\boldsymbol{g}(\boldsymbol{y}^t,\boldsymbol{y}^{t-1},\ldots,\boldsymbol{y}^{t-k})\right)}{c_{\eta,\boldsymbol{g}}(\theta,\boldsymbol{y}^{t-1},\ldots,\boldsymbol{y}^{t-k})}, \boldsymbol{y}^t,\boldsymbol{y}^{t-1},\ldots,\boldsymbol{y}^{t-k}\in\mathcal{Y}, \quad (2)$$

and

$$c_{\eta,\boldsymbol{g}}(\theta,\boldsymbol{y}^{t-1},\ldots,\boldsymbol{y}^{t-k}){=}\sum_{\boldsymbol{y}'\in\mathcal{Y}}\exp\left(\eta(\theta)\cdot\boldsymbol{g}(\boldsymbol{y}',\boldsymbol{y}^{t-1},\ldots,\boldsymbol{y}^{t-k})\right),$$

the normalizing constant.

TERGMs are a natural elaboration of the traditional ERGM framework. They are essentially stepwise ERGM in time. Note that the definitions of Robins and Pattison (2001) and Hanneke and Xing (2007) used linear ERGMs only, where $\eta(\theta)\equiv\theta$ and $p\equiv q$. To simplify notation, from this point on we suppress reference to η and $\boldsymbol{g}$.

2.2. Model Specification and Interpretation

The class of models specified by (1) is very broad and a key component of model specification is the selection of $\boldsymbol{g}$. Natural candidates are those developed for cross-sectional networks, such as those enumerated by Morris et al. (2008). However, the choices in this dynamic situation are richer and can be any valid network statistics evaluated on $\boldsymbol{y}^t$ especially those that depend on $\boldsymbol{y}^{t-1}$. Hanneke and Xing (2007) focused on a choice of $\boldsymbol{g}$ that had the property of *conditional dyadic independence* — that

$$\Pr(\boldsymbol{Y}^t{=}\boldsymbol{y}^t \,|\, \boldsymbol{Y}^{t-1}{=}\boldsymbol{y}^{t-1};\theta){=}\prod_{(i,j)\in\mathbb{Y}}\Pr(\boldsymbol{Y}^t_{i,j}{=}\boldsymbol{y}^t_{i,j}\,|\,\boldsymbol{Y}^{t-1}{=}\boldsymbol{y}^{t-1};\theta), \quad (3)$$

the distribution of $\boldsymbol{Y}^t$ in which tie states are independent, but only conditional on *the whole of $\boldsymbol{Y}^{t-1}$*.

However, caution must be used in interpreting their parameters. Consider the simplest such statistic, the edge count:

$$\boldsymbol{g}(\boldsymbol{y}^t,\boldsymbol{y}^{t-1}){=}|\boldsymbol{y}^t|.$$

A higher coefficient on $\boldsymbol{g}$ will, for any $\boldsymbol{y}^{t-1}$, produce a $\boldsymbol{Y}^t$ distribution in which networks with more ties have higher probability. But, note that this term would accomplish it in two ways simultaneously: it would both increase the weight of those networks in which more ties were formed on previously empty dyads and increase the weight of those networks in which more extant ties were preserved (fewer dissolved). That is, it would both increase the incidence and increase the duration.

Hanneke and Xing (2007) gave an example of a statistic that controls the rate of evolution of the network: a measure of *stability*. This statistic counts the number of tie variables whose

states did not change between time steps, which is then divided by the maximum number of ties an actor could have (a constant):

$$g(\boldsymbol{y}_{i,j}^{t}, \boldsymbol{y}_{i,j}^{t-1}) = \frac{1}{n-1} \sum_{(i,j) \in \mathbb{Y}} \left(\boldsymbol{y}_{i,j}^{t} \boldsymbol{y}_{i,j}^{t-1} + (1 - \boldsymbol{y}_{i,j}^{t})(1 - \boldsymbol{y}_{i,j}^{t-1}) \right).$$

A higher coefficient on it will slow the evolution of the network down and a lower coefficient will speed it up. From the point of view of incidence and duration, however, it will do so in two ways: a higher coefficient will result in networks that have fewer new ties formed and fewer extant ties dissolved — incidence will be decreased and duration will be increased.

The two-sided nature of these effects tends to muddle parameter interpretation, but a more substantial issue arises if selective mixing statistics, like those described by Koehly et al. (2004), are used. Consider a concrete example, with actors partitioned into K known groups, with $\mathbb{K} \subseteq \{1, \ldots, K\}^2$ being the set of pairs of groups between whose actors there may be ties. (For example, in a directed network, $\mathbb{K} = \{1, \ldots, K\}^2$.) Let P_k be the set of actors who belong to group k and $P(i)$ be the partition to which actor i belongs. The model with transition probability

$$\Pr(\boldsymbol{Y}^t = \boldsymbol{y}^t | \boldsymbol{Y}^{t-1} = \boldsymbol{y}^{t-1}; \theta) \propto \exp \left(\theta_0 \sum_{(i,j) \in \mathbb{Y}} \left(\boldsymbol{y}_{i,j}^t \boldsymbol{y}_{i,j}^{t-1} + (1 - \boldsymbol{y}_{i,j}^t)(1 - \boldsymbol{y}_{i,j}^{t-1}) \right) + \sum_{(k_1,k_2) \in \mathbb{K}} \theta_{k_1,k_2} |\boldsymbol{y}_{P_{k_1},P_{k_2}}| \right), \quad (4)$$

models stability, controlled by θ_0, and mixing among the groups, controlled by θ_{k_1,k_2}. (Here, $|\boldsymbol{y}_{Pk_1,Pk_2}|$ is defined as the number of ties from actors in group k_1 to actors in group k_2 for directed networks, and ties between actors in those groups for undirected networks.)

Given $\boldsymbol{y}^{t-1}$, the probability that a given non-tied directed pair (i, j) will gain a tie in a given time step is

$$\Pr(\boldsymbol{Y}_{i,j}^t = 1 | \boldsymbol{Y}_{i,j}^{t-1} = 0; \theta) = \text{logit}^{-1}(-\theta_0 + \theta_{P(i),P(j)}),$$

and the probability that an extant tie (i, j) will be removed is

$$\Pr(\boldsymbol{Y}_{i,j}^t = 0 | \boldsymbol{Y}_{i,j}^{t-1} = 1; \theta) = \text{logit}^{-1}(-\theta_0 + \theta_{P(i),P(j)}),$$

the latter leading to a duration distribution which is geometric with support $\mathbb{N}$ and expected value (Casella and Berger, 2002, pp. 621–622)

$$\left(\text{logit}^{-1}(-\theta_0 - \theta_{P(i),P(j)}) \right)^{-1} = 1 + \exp(\theta_0 + \theta_{P(i),P(j)}).$$

Thus, a higher value of coefficient θ_{k_1,k_2} simultaneously increases the incidence of ties between actors in group k_1 and actors in group k_2 and their duration.

This coupling between the incidence of ties and their duration not only makes such terms problematic to interpret, but has a direct impact on modeling. Consider a sexual partnership network, possessing strong ethnic homophily, with ties within each ethnic category being more prevalent (relative to the potential number of ties) than ties between ethnic categories.

(A real-world illustration of this effect was given by Krivitsky et al. (2011).) This structure could be a consequence of the within-ethnic ties being formed more frequently than between-ethnic ties, of the within-ethnic ties lasting, on average, longer than between-ethnic ties, or some combination of the two. With cross-sectional data alone, it is impossible to tell these apart and a model like (4) implies a dynamic process in which cross-ethnic ties toggle unnaturally frequently, or "churn". We refer to a model with this dynamic pathology as a "churning model" as this stochastic property is unlikely to be seen in real phenomena. Churning is related to the degeneracy properties of ERGM (Handcock, 2003).

3. Separable Parametrisation

We now motivate and describe the concept of *separability* of formation and dissolution in a dynamic network model, and describe the *Separable Temporal ERGM* (STERGM).

3.1. Motivation

Intuitively, those social processes and factors that result in ties being formed are not the same as those that result in ties being dissolved. For example, in the above-mentioned sexual partnership network, the relative lack of cross-ethnic ties may be a result of racial segregation, language and cultural barriers, racism, and population-level differences in socioeconomic status, all of which have a strong effect on the chances of a relationship forming. Once an interracial relationship has been formed, however, either because these factors either did not apply in that case or were overcome, the duration of such a relationship would likely not be substantially lower. Even if it were lower, the differences in the probability of such a relationship ending during a particular time interval would not, in general, be a perfect reflection the differences in the probability of it forming during such a time interval.

Furthermore, it is often the case in practice that information about cross-sectional properties of a network (i.e. *prevalence*) has a different source from that of the information about its longitudinal properties (i.e. *duration*), and it may be useful to be able to consider them separately (Krivitsky and Handcock, 2008; Krivitsky, 2009).

Thus, it is useful for the parametrisation of a model to allow separate control over incidence and duration of ties and separate interpretation, at least over the short run. (For any nontrivial process, formation and dissolution would likely interact with each other in the long run.)

3.2. Model Specification

In this section, we introduce a class of discrete-time models for network evolution, which assumes that these processes are separable from each other within a time-step. We consider a sub-class of models based on the ERGM family, which inherits the interpretability and flexibility of those processes.

3.2.1. General Separable Models—We represent networks as sets of ties, so given y, y' $\in \mathcal{Y}$ the network $y \bigcup y'$ has the tie (i, j) if, and only if, (i, j) exists in y or y' or both; the network $y \bigcap y'$ has (i, j) if, and only if, (i, j) exists in both y and y'; and the network $y \backslash y'$ has tie (i, j) if, and only if, (i, j) exists in y but not in y'. The relation $y \supseteq y'$ holds, if, and only if, y has all of the ties that y' does (and, possibly, other ties as well), and conversely for $y \subseteq y'$.

Consider the evolution of a random network at time $t - 1$ to time t, and define two intermediate networks, the *formation network* Y^+, consisting of the initial network Y^{t-1} with ties formed during the time step added and the *dissolution network* Y^-, consisting of the

initial network Y^{t-1} with ties dissolved during the time step removed (with y^+ and y^- being their respective realized counterparts). Then, given y^{t-1}, y^+, and y^-, the network y^t may be evaluated via a set operation, as

$$y^t = y^+ \setminus (y^{t-1} \setminus y^-) = y^- \cup (y^+ \setminus y^{t-1}). \quad (5)$$

Since it is the networks y^{t-1} and y^t that are actually observed, y^+ and y^- may be regarded as latent variables, but it is possible to recover them given y^{t-1} and y^t, because a tie variable can only be in one of four states given in Table 1. Each possibility has a unique combination of tie variable states in y^{t-1} and y^t, so observing the network at the beginning and the end allows the two intermediate states to be determined as $y^+ = y^{t-1} \cup y^t$ and $y^- = y^{t-1} \cap y^t$.

If Y^+ is conditionally independent of Y^- given Y^{t-1} then

$$\Pr(Y^t = y^t \mid Y^{t-1} = y^{t-1}; \theta) = \Pr(Y^+ = y^+ \mid Y^{t-1} = y^{t-1}; \theta) \times \Pr(Y^- = y^- \mid Y^{t-1} = y^{t-1}; \theta) \quad (6)$$

We refer to the two factors on the RHS as the *formation model* and the *dissolution model*, respectively. Suppose that we can express $\theta = (\theta^+, \theta^-)$ where the formation model is parametrised by θ^+ and the dissolution model by θ^-.

Definition 1. We say that a dynamic model is *separable* if Y^+ is conditionally independent of Y^- given Y^{t-1} and the parameter space of θ is the product of the individual parameter spaces of θ^+ and θ^-.

We refer to such a model as separable because it represents an assumption that during a given discrete time step, the process by which the ties form does not interact with the process by which they dissolve: both are separated (in the conditional independence sense) from each other conditional on the state of the network at the beginning of the time step.

3.2.2. Generative Mechanism—Let some $\mathscr{Y}^+(y^{t-1}) \subseteq \{y \in 2^{\mathbb{Y}} : y \supseteq y^{t-1}\}$ be the sample space, under the model, of formation networks, starting from y^{t-1}; and let some $\mathscr{Y}^-(y^{t-1}) \subseteq \{y \in 2^{\mathbb{Y}} : y \subseteq y^{t-1}\}$ be the sample space of dissolution networks. The model postulates the following process for evolution of a random network at time $t-1$ to a random network at time t:

    a. Draw an intermediate network y^+ from the distribution

$$\Pr(Y^+ = y^+ \mid Y^{t-1} = y^{t-1}; \theta^+), y^+ \in \mathscr{Y}^+(y^{t-1}).$$

    b. Draw an intermediate network y^- from the distribution

$$\Pr(Y^- = y^- \mid Y^{t-1} = y^{t-1}; \theta^-), y^- \in \mathscr{Y}^-(y^{t-1}).$$

    c. Apply formations and dissolutions to y^{t-1} to produce y^t by evaluating (5).

Note that, as specified, this model is first order Markov, but Y^t can be further conditioned on Y^{t-2}, Y^{t-3}, etc, to produce higher order versions. We do not develop these models here.

3.2.3. Separable Temporal ERGM (STERGM)—A natural family of models for the components of the separable model is the ERGMs considered in Section 2.1. We focus on this rich class of models in the remainder of the paper. Specifically, we model:

$$\Pr(\boldsymbol{Y}^+ {=} \boldsymbol{y}^+ | \boldsymbol{Y}^{t-1} {=} \boldsymbol{y}^{t-1}; \theta^+) {=} \frac{\exp\ \left(\eta^+(\theta^+) \cdot \boldsymbol{g}^+(\boldsymbol{y}^+, \boldsymbol{y}^{t-1})\right)}{c_{\eta^+, \boldsymbol{g}^+}(\theta^+, \boldsymbol{y}^{t-1})}, \boldsymbol{y}^+ \in \mathscr{Y}^+(\boldsymbol{y}^{t-1})$$

and

$$\Pr(\boldsymbol{Y}^- {=} \boldsymbol{y}^- | \boldsymbol{Y}^{t-1} {=} \boldsymbol{y}^{t-1}; \theta^-) {=} \frac{\exp\ \left(\eta^-(\theta^-) \cdot \boldsymbol{g}^-(\boldsymbol{y}^-, \boldsymbol{y}^{t-1})\right)}{c_{\eta^-, \boldsymbol{g}^-}(\theta^-, \boldsymbol{y}^{t-1})}, \boldsymbol{y}^- \in \mathscr{Y}^-(\boldsymbol{y}^{t-1}),$$

with their normalizing constants $c_{\boldsymbol{\eta}^+, \boldsymbol{g}^+}(\theta^+, \boldsymbol{y}^{t-1})$ and $c_{\boldsymbol{\eta}^-, \boldsymbol{g}^-}(\theta^-, \boldsymbol{y}^{t-1})$ summing over $\mathscr{Y}^+(\boldsymbol{y}^{t-1})$ and $\mathscr{Y}^-(\boldsymbol{y}^{t-1})$, respectively.

We now derive the probability of transitioning from a given network at time $t - 1$, $\boldsymbol{y}^{t-1}$ to a given network at time t, $\boldsymbol{y}^t$. Based on (6), we have

$$\Pr(\boldsymbol{Y}^t {=} \boldsymbol{y}^t | \boldsymbol{Y}^{t-1} {=} \boldsymbol{y}^{t-1}; \theta) {=} \frac{\exp\ \left(\eta(\theta) \cdot \boldsymbol{g}(\boldsymbol{y}^t, \boldsymbol{y}^{t-1})\right)}{c_{\eta^+, \boldsymbol{g}^+}(\theta^+, \boldsymbol{y}^{t-1}) c_{\eta^-, \boldsymbol{g}^-}(\theta^-, \boldsymbol{y}^{t-1})}.$$

where $\boldsymbol{\eta} = (\boldsymbol{\eta}^+, \boldsymbol{\eta}^-)$ and $\boldsymbol{g}(\boldsymbol{y}^t, \boldsymbol{y}^{t-1}) = (\boldsymbol{g}^+(\boldsymbol{y}^{t-1} \cup \boldsymbol{y}^t, \boldsymbol{y}^{t-1}), \boldsymbol{g}^-(\boldsymbol{y}^{t-1} \cap \boldsymbol{y}^t, \boldsymbol{y}^{t-1}))$. As $\Pr(\boldsymbol{Y}^t = \boldsymbol{y}^t | \boldsymbol{Y}^{t-1} = \boldsymbol{y}^{t-1}; \theta)$ is, by construction, a valid probability mass function,

$$c_{\eta^+, \boldsymbol{g}^+}(\theta^+, \boldsymbol{y}^{t-1}) c_{\eta^-, \boldsymbol{g}^-}(\theta^-, \boldsymbol{y}^{t-1}) {=} c_{\eta, \boldsymbol{g}}(\theta, \boldsymbol{y}^{t-1}),$$

where

$$c_{\eta, \boldsymbol{g}}(\theta, \boldsymbol{y}^{t-1}) {=} \sum_{\boldsymbol{y}' \in \mathscr{Y}} \exp\ \left(\eta(\theta) \cdot \boldsymbol{g}(\boldsymbol{y}', \boldsymbol{y}^{t-1})\right).$$

This is the same form as (1). Thus, the STERGM class is a subclass of a first-order Markov TERGM of Hanneke and Xing (2007), described in Section 2.1: any transition process that can be expressed with $\boldsymbol{g}^+$, $\boldsymbol{g}^-$, $\boldsymbol{\eta}^+$ and $\boldsymbol{\eta}^-$ can be reproduced by a model in the TERGM class. However, the essential issue is the specification of models within these classes, and the value of the STERGM class is that it focuses specification on a viable and fecund region in the very broad class. In the parametrisation in terms of formation and dissolution, some flexibility is lost — the ability to have the formation and dissolution processes interact within a given time step. What is gained is ease of specification, tractability of the model, and substantial improvement in interpretability.

3.3. Interpretation

In contrast to statistics like *stability* in Section 2.2, the STERGM's sufficient statistics and parameters have an implicit direction: they affect directly either incidence or duration, but not both, and even statistics that do not explicitly incorporate the previous time step's network $\boldsymbol{y}^{t-1}$, incorporate it via the constraint of the phase in which they are used. This allows familiar cross-sectional ERGM sufficient statistics to be used, with their parameters acquiring intuitive interpretations in terms of the network evolution process. We call these inherited terms, for which $g_k^+(\boldsymbol{y}^+, \boldsymbol{y}^{t-1}) \equiv g_k^+(\boldsymbol{y}^+)$ and/or $g_k^-(\boldsymbol{y}^-, \boldsymbol{y}^{t-1}) \equiv g_k^-(\boldsymbol{y}^-)$, with no further dependence on $\boldsymbol{y}^{t-1}$, *implicitly dynamic*.

Such terms (and their corresponding coefficients) often have straightforward general interpretations for formation and dissolution phases. In particular, consider an implicitly dynamic statistic that counts the number of instances of a particular feature found in the network y^+ or y^-. Examples of features that might be counted include a tie, an actor with exactly d neighbors, or a tie between an actor in a set P_{k1} and an actor in the set P_{k2}.

3.3.1. Formation—A positive θ_k^+ corresponding to a particular g_k^+ increases the probability of those y^+ which have more instances of the feature counted by g_k^+ — greater values of $g_k^+(y^+)$. This affects the network process in two ways: the probability of forming those ties that create new instances of the feature counted by g_k^+ is increased and the probability of forming those ties that "disrupt" those instances would be reduced.

Conversely, negative θ_k^+ would result in higher probabilities for those networks with fewer instances of the feature counted by g_k^+, reducing the probability of forming ties to create more instances of the feature counted and increasing the probability of forming ties to "disrupt" the feature.

Notably, g_k^+ counts features in the network $y^+ = y^t \cup y^{t-1}$, rather than in the ultimately observed network y^t. This means that for some features, particularly those with dyadic dependence, the dissolution process may influence the feature so that it is present in y^+ but not in y^{t-1} or y^t. How frequently this occurs depends on the specific model and the rate of evolution of the network process: if a network process is such that the network changes little (in both formation and dissolution) during each time step, such interference is unlikely.

3.3.2. Dissolution—As in the formation phase, a positive θ_k^- corresponding to a particular g_k^- increases the probability of those y^- which have more instances of the feature counted by g_k^-, thus tending to preserve more instances of that feature (or dissolving ties to create more instances, as may be the case with dyadic-dependent terms), while a negative θ_k^- will increase the probability of networks with fewer instances of the feature in question, effectively causing the dissolution process to target those features, and also refrain from dissolving ties whose dissolution would create those features. It is important to note that the dissolution phase ERGM determines which ties are *preserved* during the time step, and the parameters should be interpreted accordingly.

Again, it is $y^- = y^t \cap y^{t-1}$ on which statistics are evaluated, so the formation process can interact with the dissolution process as well.

These principles mean that many of the vast array of network statistics developed for ERGMs (Morris et al., 2008, for example) can be readily adapted to STERGM modeling, retaining much of their interpretation. In the Appendix, we develop and give interpretations to the fundamental edge count, selective mixing by actor attribute, and degree distribution terms.

3.3.3. Explicitly Dynamic Terms—At the same time, some effects on formation and dissolution may depend on specific features of y^{t-1}. For instance, consider a social process in which an actor having multiple partners (e.g., "two-timing") is actively punished, so having more than one partner in y^{t-1} increases the hazard of losing all of one's partners in y^t. (Such an effect may be salient in a sexual partnership network.) This dissolution effect cannot be modeled by implicitly dynamic terms, because it cannot be reduced to merely increasing or reducing the tendency of Y^- to have particular features. For example, a positive coefficient on a statistic counting the number of actors with no partners (isolates)

would increase the weight of those $\boldsymbol{y}^-$ that have more isolates, affecting the dissolution of the sole tie of an actor with only one partner just as much as it would affect the dissolution of ties of an actor with more than one partner.

On the other hand, an *explicitly dynamic* model term that counts the number of actors with no partners in $\boldsymbol{y}^-$ *only among those actors who had two or more partners in* $\boldsymbol{y}^{t-1}$ would, with a positive coefficient, increase the probability of a transition directly from having two partners to having none. Beyond that, its interpretation would be no different than that of an implicitly dynamic dissolution term.

3.4. Continuous-Time Markov Models

Although the focus of this paper is on discrete-time models for network evolution, the separability paradigm can be applied to continuous-time network evolution models such as those of Holland and Leinhardt (1977). There, network evolution is modeled as a continuous-time Markov process such that the intensity of transition between two networks that differ by more than one dyad is 0, while the evolution of the network is controlled by $\lambda(\boldsymbol{y}^t;\theta):\mathscr{Y} \to \mathbb{R}_+^{\mathscr{Y}}$, with each $\lambda_{i,j}(\boldsymbol{y}^t;\theta)$ being the intensity associated with toggling each dyad (i, j).

In that scenario, separation of formation and dissolution is realized by formulating $\theta = (\theta^+, \theta^-)$ and

$$\lambda_{i,j}(\boldsymbol{y}^t;\theta^+, \theta^-) = \begin{cases} \lambda_{i,j}^+(\boldsymbol{y}^t;\theta^+) & \text{if } \boldsymbol{y}_{i,j}^t = 0 \\ \lambda_{i,j}^-(\boldsymbol{y}^t;\theta^-) & \text{if } \boldsymbol{y}_{i,j}^t = 1 \end{cases},$$

with $\lambda_{i,j}^+(\boldsymbol{y}^t;\theta^+)$ and $\lambda_{i,j}^-(\boldsymbol{y}^t;\theta^-)$ being formation- and dissolution-specific intensities. Indeed, Holland and Leinhardt (1977) use a formulation of this general sort. Notably, unlike the discrete-time process, this separation requires only separation of parameters and no additional independence assumptions. This is because under the Markov assumption and with no chance of more than one dyad toggling coincidentally at a specific time, dyads effectively evolve independently in a sufficiently small interval (i.e., $[t, t + h], h \to 0$), and dyadic independence in network evolution *a fortiori* implies separability between which ties form and which ties dissolve.

An exponential-family form for $\lambda_{i,j}$,

$$\lambda_{\eta,\boldsymbol{g},(i,j)}(\boldsymbol{y}^t;\theta) = \begin{cases} \exp\left(\eta^+(\theta^+) \cdot \boldsymbol{g}^+(\boldsymbol{y}^t \cup \{(i,j)\}, \boldsymbol{y}^t)\right) & \text{if } \boldsymbol{y}_{i,j}^t = 0 \\ \exp\left(\eta^-(\theta^-) \cdot \boldsymbol{g}^-(\boldsymbol{y}^t \setminus \{(i,j)\}, \boldsymbol{y}^t)\right) & \text{if } \boldsymbol{y}_{i,j}^t = 1 \end{cases},$$

may be viewed as the limiting case of the discrete-time STERGM, in which the amount of time represented by each time step shrinks to zero.

4. Likelihood-Based Inference for TERGMs

In this section, we consider inference based on observing a series of $T + 1$ networks, $\boldsymbol{y}^0, \ldots, \boldsymbol{y}^T$. Hanneke and Xing (2007) proposed to fit TERGMs by finding the conditional MLE under an order k Markov assumption,

$$\hat{\theta} = \arg \max_{\theta} \prod_{t=k}^{T} \Pr(\boldsymbol{Y}^t = \boldsymbol{y}^t \mid \boldsymbol{Y}^{t-k} = \boldsymbol{y}^{t-k}, \dots, \boldsymbol{Y}^{t-1} = \boldsymbol{y}^{t-1}; \theta), \quad (7)$$

computing a Method-of-Moments estimator (equivalent in their case to the MLE) with a simulated Newton-Raphson zero-finding algorithm. We extend the work of Hunter and Handcock (2006) and Geyer and Thompson (1992) to compute the conditional MLE for curved exponential-family transition models (that is, cases where $\eta(\theta)\ \ \theta$).

For simplicity, we consider models with first-order Markov dependence. There is no loss of generality, since as long as the order of Markov dependence k is finite, we can define the depended-upon network $\boldsymbol{y}^{t-1}$ to implicitly "store" whatever information about $\boldsymbol{y}^{t-1}, \dots,$ $\boldsymbol{y}^{t-k+1}$ is needed to compute the transition probability.

The conditional MLE (7) can then be obtained by maximizing the log-likelihood

$$l(\theta) = \eta(\theta) \cdot \left(\sum_{t=1}^{T} \boldsymbol{g}(\boldsymbol{y}^t, \boldsymbol{y}^{t-1}) \right) - \log \left(\prod_{t=1}^{T} c_{\eta, \boldsymbol{g}}(\theta, \boldsymbol{y}^{t-1}) \right).$$

For any two values of the model parameter θ^0 and θ, the log-likelihood-ratio is

$$l(\theta) - l(\theta^0) = (\eta(\theta) - \eta(\theta^0)) \cdot \left(\sum_{t=1}^{T} \boldsymbol{g}(\boldsymbol{y}^t, \boldsymbol{y}^{t-1}) \right) - \log \left(\prod_{t=1}^{T} \frac{c_{\eta, \boldsymbol{g}}(\theta, \boldsymbol{y}^{t-1})}{c_{\eta, \boldsymbol{g}}(\theta^0, \boldsymbol{y}^{t-1})} \right).$$

The main difficulty is in evaluating the ratio of the normalizing constants. These conditional normalizing constants depend on networks at times $0, \dots, T - 1$. However, these ratios can still be expressed as

$$\prod_{t=1}^{T} \frac{c_{\eta, \boldsymbol{g}}(\theta, \boldsymbol{y}^{t-1})}{c_{\eta, \boldsymbol{g}}(\theta^0, \boldsymbol{y}^{t-1})}$$

$$= \prod_{t=1}^{T} \sum_{\boldsymbol{y} \in \mathcal{Y}} \exp \left(((\eta(\theta) \right.$$

$$\left. - \eta(\theta^0)) \right.$$

$$\left. \cdot \boldsymbol{g}(\boldsymbol{y}, \boldsymbol{y}^{t-1}) \right)$$

$$\times \frac{\exp \left(\eta(\theta^0) \cdot \boldsymbol{g}(\boldsymbol{y}, \boldsymbol{y}^{t-1}) \right)}{c_{\eta, \boldsymbol{g}}(\theta^0, \boldsymbol{y}^{t-1})} \quad (8)$$

$$= \prod_{t=1}^{T} \sum_{\boldsymbol{y} \in \mathcal{Y}} \exp \left(((\eta(\theta) \right.$$

$$\left. - \eta(\theta^0)) \right.$$

$$\left. \cdot \boldsymbol{g}(\boldsymbol{y}, \boldsymbol{y}^{t-1}) \right)$$

$$\times \Pr(\boldsymbol{Y}^t$$

$$= \boldsymbol{y} \mid \boldsymbol{Y}^{t-1}$$

$$= \boldsymbol{y}^{t-1}; \theta^0).$$

The expression (8) is a product of expectations over the conditional distribution under the model of Y^t given Y^{t-1} at θ^0, each of which can be estimated by simulation, allowing the algorithm of Hunter and Handcock (2006) to be applied to fit a TERGM to network series data.

These results also make it possible to assess the goodness-of-fit of a model via an analyses of deviance. Specifically, we can compute the change in log-likelihood from the null model ($\eta(\theta) = 0$) to the conditional MLE. To do this, we extended the bridge sampler of Hunter and Handcock (2006) to this setting.

5. Application to the Dynamics of Friendship

As an application of this model, we consider the friendship relations among 26 students during their first year at a Dutch secondary school (Knecht, 2008). The friendship nominations were assessed at four time points at intervals of three months starting at the beginning of their secondary schooling. The friendship data are directed and were assessed by asking students to indicate classmates whom they considered good friends. There were 17 girls and 9 boys in the class. The data included covariates collected on each student. Here, we consider the sex of the student, as it is a primary determinant of the friendship ties. We also consider a dyadic covariate indicating if each pair of students had gone to the same primary school. These data were used to illustrate the actor-oriented approach to modeling by Snijders et al. (2010) (whom we follow). That paper should be consulted for details of the data set and an alternative analysis.

Some of the data at time points two through four were missing due to student absence when the survey was taken. These were accommodated using the approach of Handcock and Gile (2010) under the assumption that the unobserved data pattern was amenable to the model; this assumption is reasonable as long as student absence is conditionally independent of the unobserved ties given the observed ties. One student left the class after time point 1. This could have been accommodated in a number of ways, depending on the assumptions one is willing to make. Here we considered the networks with this student omitted both as a nominator and nominee of friendships. As Snijders et al. (2010) note, each student was allowed to nominate at most 12 classmates at each time point. In general, inference needs to incorporate features of sampling design such as this one. We discuss how in Section 6. However, its effect here is negligible: in the ($4 \times 25 =$) 100 student reports, only 3 nominated the maximum number.

Our objective is to explain the observed structural patterns of change in the network over the course of the year. We build a model including both exogenous and endogenous structural effects, following the same approach and motivations as Snijders et al. (2010). For the formation component we include terms for the propensity of students to choose friends of the same or opposite sex (i.e., overall propensities to nominate friends that are homophilous on sex or not). We include a term to measure the propensity of friendships to be reciprocal. We include information on the primary school co-attendance via a count of the number of times students nominate other students with whom they went to primary school. To capture any overall propensity of students to nominate other students who are popular we include an overall outdegree popularity effect (Snijders et al., 2010, equation (12)). To model transitivity effects we include two terms. The first is aggregate transitive ties that aims to capture a tendency toward transitive closure consistent with local hierarchy. The second is an aggregate cyclical ties term to capture anti-hierarchical closure. The terms in the model are structurally largely consistent with the terms chosen in Snijders et al. (2010). A similar model was considered for the dissolution process. Specifics of these terms are given in the Appendix.

We fit the model using the conditional MLE procedure of Section 4. Computationally this is implemented using a variant of the MCMC approach of Hunter and Handcock (2006). To monitor the statistical properties of the MCMC algorithm we use the procedures by Hunter et al. (2008a).

Table 2 reports the estimates for the model assuming homogeneity of parameters over time. The outdegree popularity effect had a correlation of 0.995 with the edges effect and was omitted from the model.

As for the standard ERGM, the individual θ coefficients can be interpreted as conditional log-odds ratios. There is also a relative risk interpretation that is often simpler. For example, the exponential of the primary school coefficient is the relative risk of formation or preservation (depending on the phase) of friendship between two students from the same primary school compared to two students from different primary schools with the same values of the other covariates and structural effects. The probabilities involved are conditional on these other covariates and structural effects. The interpretation for non-binary and multiple covariates is similar: $\exp(\theta\Delta)$ is the relative risk of friendship between two students compared to two students with vector of covariates differing by Δ (and with the same values of the other structural effects).

The standard errors of Table 2 are obtained from the information matrix in the likelihood evaluated at the MLE to which we have added the (small) MCMC standard error obtained using the procedure given by Hunter et al. (2008b).

The networks at the earlier time points are strongly sexually segregated, and we see strong homophily by sex in the formation of ties. This effect is mildly stronger for boys than for girls. We do not see an overall disinclination for girls to nominate boys (relative to other combinations). In other words, the boys are about as likely to form friendships as the girls. As expected, we see a high degree of reciprocity in the formation of ties. There is a strong transitive closure effect, with a positive coefficient on transitive tie formation and a negative coefficient on cyclical tie formation. This suggests a strong hierarchical tendency in the formation of ties. We see that students who attended the same primary school are much more likely to form ties.

These structural terms have less influence on the dissolution of ties. There is some modest evidence that boy to boy ties are less likely to dissolve than other mixtures of sexes. (Recall that parameters represent a measure of persistence, so that negative parameters are associated with shorter durations). As expected, we see the dissolution of ties is strongly retarded by the presence of a reciprocal tie. As in the formation process, there is a strong transitive closure effect suggesting a strong hierarchical tendency in the dissolution of ties. Once a hierarchical triad is formed it will tend to endure longer. Students who attended the same primary school are not significantly more likely to have persistent ties.

As the data measure a social process that is developing in time, we do not need to assume that the process is in temporal equilibrium; thus we could estimate separate parameters for the change between each pair of successive time points. One such model specifies different overall rates of tie formation or dissolution at each time point but retains homogeneous parameters for the other terms. Another allows all the parameters to vary at each time point.

Table 3 gives the analysis of deviance for formation and dissolution models nested above and below those in Table 2. For the formation process we see the full time-homogeneous model in Table 2 significantly improves on the null and Erdős-Rényi model (Edges (hom.)). Specifying different overall rates of tie formation at each time point does not significantly improve the fit, nor does a full time-heterogeneous model with different structural

parameters at each time point. For the dissolution process, we again see the full time-homogeneous model significantly improves on the null and Erd s-Rényi model. However, there is some evidence that specifying time-heterogeneous versions improve the fit. An inspection of the time-heterogeneous models indicates that most of the improvement is due to the increase in hierarchical tendency over time. Initially this transitive closure does not retard tie dissolution, but it does over time.

6. Discussion

This paper introduces a statistical model for networks that evolve over time. It builds on the foundations of exponential-family random graph models for cross-sectional networks and inherits the flexibility and interpretability of these models. In addition, it leverages the inferential and computational tools that have been developed for ERGMs over the last two decades.

As we showed in Section 2, parameters used in models currently in use directly affect both the incidence of ties (at a given time point) and the duration of ties (over time). STERGMs have one set of parameters control formation of new ties and another control dissolution (or non-dissolution) of extant ties. Such a separable parametrisation controls the attribution of incidence and duration and greatly improves the interpretability of the model parameters, all without sacrificing the ability to explicitly incorporate effects of specific features of past networks, if needed.

It is important to emphasize that STERGMs jointly model the formation and dissolution of ties. While the two processes are modeled as conditionally independent within a time step, they are modeled as dependent over time. More importantly, they allow the structure of the incidence to be identified in the presence of the durational structure.

In addition, the model has computational advantages. The likelihood function can be decomposed and the components computed relatively easily. All computations in this paper were completed using the **ergm** (Hunter et al., 2008b; Handcock et al., 2012) package from the **statnet** (Handcock et al., 2008) suite of libraries for social network analysis in R (R Development Core Team, 2009).

The model is directly applicable to both directed and undirected networks. It can be easily tuned to applications by appropriate choices of terms for both the formation and dissolution processes, as we show in Section 5. Because it is based on ERGMs, it will share in advances made on those models as well. The model is very useful for simulating realistic dynamic networks. This is because of the sequential specification, the tractable parameters and the relatively light computation burden.

As illustrated in Section 5, missing data on the relational information can be dealt with in likelihood-based inference using the approach of Handcock and Gile (2010). If the longitudinal data are partially observed due to either a sample design or a missing data process and is amenable to the model then their method is directly applicable.

The assumption of within-step independence of formation and dissolution is an important one, and its appropriateness depends on the substantive setting and the basic nature of the process. Some settings do not allow a separable formulation at all. For example an affiliation network of players to teams in some sports, with a realization observed during every game, imposes a hard constraint that a player must belong to exactly one team at a time, and no team can have more or fewer than a particular number of players, so the basic unit of network change is teams trading players, rather than a player joining or quitting a team. In settings that do allow simpler atomic changes, separability may be a plausible approximation

if the amount of change between the discrete time steps is relatively small — that each time step represents a fairly small amount of time. As the length of the time step increases, the separability approximation may become less and less plausible. For example, a marriage network, even though it has a hard constraint of each actor having at most one spouse at a time, could be plausibly approximated in a separable framework (using, e.g., $\mathcal{Y}^+(y^{t-1}) \equiv \{y \in 2^{\mathbb{Y}} : y \supseteq y^{t-1} \wedge \forall_{i \in N} |y_i| \leq 1\}$) if each discrete time step represented one month (since relatively few people divorce and remarry in the same month) but not if it represented ten years. More generally, the simpler the formation and dissolution processes are within a time-step and the weaker the dependence between them, the more plausible the assumption. (Of course, continuous-time Markov models, to which these models asymptote, do not require an independence assumption at all.)

As with the data used in Section 5, restriction on the number of alters reportable is a common feature of network surveys. Other examples of this censoring include the Add Health friendship networks (Harris et al., 2003) and Sampson's monastery data (Sampson, 1968). To the extent that these are features of the sampling design, they should be reflected in the likelihood. Per Section 3.2.3, a STERGM can be represented as a TERGM (1), which allows the sample space $\mathcal{Y}$ of Y^t to be constrained to reflect this design. Changing $\mathcal{Y}$ only affects $c_{\eta,g}(\theta)$ in $l(\theta)$ — the kernel of the model remains separable. This situation is similar to that with censored data in survival analysis where the likelihood is altered to reflect the censoring while the model, and its interpretation, is unchanged.

Since assuming separability between formation and dissolution grants significant benefits to interpretability, it would be useful to be able to test if separability may be assumed in a given network process. Some avenues for such tests include comparing goodness-of-fit of a given model in modeling a transition $y^0 \to y^2$ to its modeling a transition $y^0 \to y^1 \to y^2$ (with homogeneous parameters). Or, if only one transition is available, a transition $y^0 \to y^1$ to a transition $y^0 \to Y^{\frac{1}{2}} \to y^1$, with a latent intermediate network $Y^{\frac{1}{2}}$. Development of such tests is beyond the scope of this work and is subject for future research.

The STERGM framework allows a number of extensions to the model. Over time, networks do not merely change ties: actors enter and leave the network, and actors' own attributes change. It is possible to incorporate the network size adjustment developed by Krivitsky et al. (2011) into these dynamic models. We have focused on longitudinal data. It is possible to fit the model based on egocentrically sampled data when the data includes durational information on the relational ties (Krivitsky and Handcock, 2008; Krivitsky, 2009).

Acknowledgments

Appendix

Separable TERGM Terms

In this appendix we derive and discuss some fundamental model terms that can be used in a STERGM.

A.1. Edge Counts

A.1.1. Formation—Let $g^+(y^+, y^{t-1}) = |y^+|$. This is equivalent to $g(y^t, y^{t-1}) = |y^t \cup y^{t-1}|$. If $y_{i,j}^{t-1} = 1$, $y_{i,j}^t \vee y_{i,j}^{t-1} = 1$, so the state of $y_{i,j}^t$ has no effect on $g(y^t, y^{t-1})$, but if $y_{i,j}^{t-1} = 0$, $y_{i,j}^t \vee y_{i,j}^{t-1} = y_{i,j}^t$, and the change in $g(y^t, y^{t-1})$ is 1. This means that, in the absence of other formation terms, θ^+ represents the log-odds of a given tie variable, that does not already have a tie, gaining a tie. Then $\text{logit}^{-1}(\theta^+)$ is the expected fraction of tie variables empty at time $t-1$ gaining a tie at time t. In the presence of other terms, these log-odds become conditional log-odds-ratios.

A.1.2. Dissolution—Let $g^-(y^-, y^{t-1}) = |y^-|$, or, equivalently, $g(y^t, y^{t-1}) = |y^t \cap y^{t-1}|$. If $y_{i,j}^{t-1} = 0$, $y_{i,j}^t \wedge y_{i,j}^{t-1} = 0$, so the state of $y_{i,j}^t$ has no effect on $g(y^t, y^{t-1})$, but if $y_{i,j}^{t-1} = 1$, $y_{i,j}^t \wedge y_{i,j}^{t-1} = y_{i,j}^t$, and the change in $g(y^t, y^{t-1})$ is 1. Then, in the absence of other dissolution terms, θ^- represents the log-odds of a given tie that exists at $t-1$ *surviving* to t, and $\text{logit}^{-1}(\theta^-)$ is the expected fraction of ties extant at time $t-1$ surviving to time t. Depending on the problem, the interpretation of $-\theta^-$ might be more useful: $\text{logit}^{-1}(-\theta^-)$ is the expected fraction of extant ties being dissolved — the hazard.

The formation phase can only affect non-tied pairs of actors, so if the dissolution phase statistics have dyadic independence, the formation process has no effect on duration distribution: in the absence of other dissolution terms, the duration distribution of a tie is geometric (with support $\mathbb{N}$) with expected value (Casella and Berger, 2002, pp. 621–622)

$$\left(\text{logit}^{-1}(-\theta^-) \right)^{-1} = 1 + \exp(\theta^-).$$

A.2. Selective Mixing

Selective mixing in the formation model can be represented by a vector of statistics $g^+(y^+, y^{t-1}) = |y_{P_{k_1}, P_{k_2}}^+|$, with notation described for (4). However, in the context of a STERGM, they have a direction.

A.2.1. Formation—Let $g_{k_1,k_2}^+(y^+, y^{t-1}) = |y_{P_{k_1}, P_{k_2}}^+|$ (equivalently, $g_{k_1,k_2}(y^t, y^{t-1}) = |y_{P_{k_1}, P_{k_2}}^t \cup y_{P_{k_1}, P_{k_2}}^{t-1}|$). The change in its value due to adding a tie (i, j) (absent in y^{t-1}) is $1_{i \in P_{k_1} \wedge j \in P_{k_2}}$, so θ_{k_1,k_2}^+ is the conditional log-odds-ratio due to the effect of i belonging to group k_1 and j belonging to group k_2 of a dyad (i, j), that does not already have a tie, gaining a tie. If the formation phase has no other terms, then the odds that $Y_{i,j}^t = 1$ given that $Y_{i,j}^{t-1} = 0$ are

$$\text{Odds}_{\eta,g}(Y_{i,j}^t = 1 | Y_{i,j}^{t-1} = 0, i \in P_{k_1} \wedge j \in P_{k_2}; \theta_{k_1,k_2}^+) = \exp\left(\theta_{k_1,k_2}^+ \right).$$

A.2.2. Dissolution—Similar to the formation case, selective mixing can be represented by a vector of statistics $g^-(y^-, y^{t-1}) = |y_{P_{k_1}, P_{k_2}}^-|$. Then, $g_{k_1,k_2}(y^t, y^{t-1}) = |y_{P_{k_1}, P_{k_2}}^t \cap y_{P_{k_1}, P_{k_2}}^{t-1}|$, and θ_{k_1,k_2}^- is the conditional log-odds-ratio due to the effect of i belonging to group k_1 and j belonging to group k_2 of an extant tie (i, j) being preserved until the next time step.

A.3. Degree Distribution

Unlike the first two examples, degree distribution statistics — counts of actors with a particular degree or range of degrees — introduce dyadic dependence into the model. As with many other such terms, closed forms for many quantities of interest are not available, and conditional log-odds are not as instructive, but the general results for implicitly dynamic terms from Section 3.3 provide a useful heuristic, with the caveats discussed in that section.

In practice, these terms are often used in conjunction with other terms, so we only discuss their effect on the formation and dissolution probabilities conditional on other terms — their effect over and above other terms, with those terms' coefficients held fixed.

A.3.1. Formation—Let $\boldsymbol{y}_i$ be the set of neighbors to whom i has ties in $\boldsymbol{y}$. A formation degree count term has the form $g_k^+(\boldsymbol{y}^+, \boldsymbol{y}^{t-1}) = \sum_{i \in N} 1_{|\boldsymbol{y}_i^+|=d}$: the number of actors i in $\boldsymbol{y}^+$ whose degree is d. The corresponding TERGM statistic $g_k(\boldsymbol{y}^t, \boldsymbol{y}^{t-1}) = \sum_{i \in N} 1_{|\boldsymbol{y}_i^t \cup \boldsymbol{y}_i^{t-1}|=d}$. We discuss the cases of $d = 0$ and $d = 1$, with the cases for $d > 1$ being similar to the $d = 1$ case.

$d = 0$ By increasing the weight of those formation networks that have fewer isolates, a negative coefficient on this term increases the chances of a given actor gaining its first tie within a given time step. Conversely, a positive coefficient reduces the chances of an actor gaining its first tie. Because the term does not distinguish between different nonzero degrees, it mainly affects transitions from isolation to degree 1, not affecting further tie formation on that actor positively or negatively.

$d = 1$ Unlike the statistic for $d = 0$, which can only be decreased by adding ties, the statistic for $d = 1$ can be both increased and decreased (by making isolates into actors with degree 1 and by making actors with degree 1 into actors with degree 2 and higher, respectively). Thus, the effect of this term is two-sided: with a positive coefficient, it both increases the chances of an actor gaining its first tie and reduces the chances of an actor gaining its second tie, while having relatively little effect on an actor with two ties gaining a third tie. A negative coefficient reduces the chances of an actor gaining its first tie, but if an actor already has one tie, it increases the chances that the actor gains a second tie.

A.3.2. Dissolution—The analogous term in the dissolution model is

$g_k^-(\boldsymbol{y}^-, \boldsymbol{y}^{t-1}) = \sum_{i \in N} 1_{|\boldsymbol{y}_i^-|=d}$: same as formation, but applied to $\boldsymbol{y}^-$, and
$g_k(\boldsymbol{y}^t, \boldsymbol{y}^{t-1}) = \sum_{i \in N} 1_{|\boldsymbol{y}_i^t \cap \boldsymbol{y}_i^{t-1}|=d}$.

$d = 0$ A negative coefficient on this term in the dissolution phase increases the weight of dissolution networks that have fewer isolates, and thus reduces the chances of a given actor losing its only tie, while a positive coefficient increases the chances of an actor losing its only tie. It may also have a modest effect on actors with more than one tie, since there is a positive probability of an actor losing more than one tie in the same time step.

$d = 1$ As in the case of formation, the effect of this term is two-sided: with a positive coefficient — to preserve or create networks with more "monogamous" ties — the chances of an actor losing its only tie decrease while the chances of an actor losing its second tie increase. (If an actor has 3 or more ties, the effect is weaker.)

A negative coefficient on this term both increases the chances that an actor's last tie will be dissolved and reduces the chances that an actor with more than one tie has any ties dissolve.

A.4. Other Standard Statistics

Most statistics used in standard ERGM can be used in STERGM as implicitly dynamic statistics. For example, standard formation statistics are

Reciprocity: $\sum_{(i,j)\in\mathbb{Y},i<j} \boldsymbol{y}_{i,j}^+ \boldsymbol{y}_{j,i}^+$

Transitive ties: $\sum_{(i,j)\in\mathbb{Y}} \boldsymbol{y}_{i,j}^+ \max_{k\in N}(\boldsymbol{y}_{i,k}^+ \boldsymbol{y}_{k,j}^+)$

Cyclical ties: $\sum_{(i,j)\in\mathbb{Y}} \boldsymbol{y}_{i,j}^+ \max_{k\in N}(\boldsymbol{y}_{k,i}^+ \boldsymbol{y}_{j,k}^+)$

Outdegree popularity (sqrt): $\sum_{(i,j)\in\mathbb{Y}} \boldsymbol{y}_{i,j}^+ \sqrt{|\boldsymbol{y}_{\cdot,j}^+|}$

Edge covariate: For a covariate $x \in \mathbb{R}^{n\times n}$, $\sum_{(i,j)\in\mathbb{Y}} \boldsymbol{y}_{i,j}^+ \boldsymbol{x}_{i,j}$

The corresponding dissolution statistics have the same form, with $\boldsymbol{y}^+$ replaced by $\boldsymbol{y}^-$.

References

Butts CT. A relational event framework for social action. Sociological Methodology. 2008; 38(1):155–200.

Casella, G.; Berger, RL. Statistical Inference. second ed.. Pacific Grove, CA: Duxbury; 2002. Duxbury Advanced Series.

Doreian, P.; Stokman, FN., editors. Evolution of Social Networks. Mahwah, NJ: Routledge; 1997.

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

Geyer CJ, Thompson EA. Constrained Monte Carlo maximum likelihood for dependent data (with discussion). J. R. Statist. Soc. B. 1992; 54(3):657–699.

Handcock, MS. Statistical models for social networks: Inference and degeneracy. In: Breiger, R.; Carley, K.; Pattison, P., editors. Dynamic Social Network Modeling and Analysis. Washington, DC: Committee on Human Factors, Board on Behavioral, Cognitive, and Sensory Sciences. National Academy Press; 2003. p. 229-240.

Handcock MS, Gile KJ. Modeling social networks from sampled data. Annals of Applied Statistics. 2010; 4(1):5–25.

Handcock MS, Hunter DR, Butts CT, Goodreau SM, Krivitsky PN, Morris M. ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks. 2012 Statnet Project URL http://www.statnet.org. Version 3.0-3. URL http://cran.r-project.org/package=ergm.

Handcock MS, Hunter DR, Butts CT, Goodreau SM, Morris M. statnet: Software tools for the representation, visualization, analysis and simulation of network data. Journal of Statistical Software. 2008; 24(1):1–11. URL http://www.jstatsoft.org/v24/i01. [PubMed: 18612375]

Hanneke S, Fu W, Xing EP. Discrete temporal models of social networks. Electronic Journal of Statistics. 2010; 4:585–605.

Hanneke, S.; Xing, EP. Discrete temporal models of social networks. In: Airoldi, E.; Blei, DM.; Fienberg, SE.; Goldenberg, A.; Xing, EP.; Zheng, AX., editors. Statistical Network Analysis: Models, Issues, and New Directions: ICML 2006 Workshop on Statistical Network Analysis, Pittsburgh, PA, USA, June 29, 2006, Revised Selected Papers, vol. 4503 of Lecture Notes in Computer Science. Springer; 2007. p. 115-125.

Harris, KM.; Florey, F.; Tabor, J.; Bearman, PS.; Jones, J.; Udry, JR. Tech. rep. University of North Carolina; 2003. The national longitudinal study of adolescent health: Research design. URL http://www.cpc.unc.edu/projects/addhealth/design/

Holland PW, Leinhardt S. A dynamic model for social networks. Journal of Mathematical Sociology. 1977; 5(1):5–20.

Hunter DR, Goodreau SM, Handcock MS. Goodness of fit for social network models. J. Am. Statist. Ass. 2008a; 103(481):248–258.

Hunter DR, Handcock MS. Inference in curved exponential family models for networks. Journal of Computational & Graphical Statistics. 2006; 15(3):565–583.

Hunter DR, Handcock MS, Butts CT, Goodreau SM, Morris M. ergm: A package to fit, simulate and diagnose exponential-family models for networks. Journal of Statistical Software. 2008b; 24(3):1–29. URL http://www.jstatsoft.org/v24/i03. [PubMed: 18612375]

Knecht, A. PhD Dissertation. University of Utrecht; 2008. Friendship Selection and Friends: Influence. Dynamics of Networks and Actor Attributes in Early Adolescence.

Koehly LM, Goodreau SM, Morris M. Exponential family models for sampled and census network data. Sociological Methodology. 2004; 34(1):241–270.

Krivitsky, PN. Ph.D. thesis. Seattle, WA: University of Washington; 2009. Statistical Models for Social Network Data and Processes.

Krivitsky, PN.; Handcock, MS. Joint Statistical Meetings 2008. Denver, CO, USA: American Statistical Association; 2008. Separable models for dynamic social networks.

Krivitsky PN, Handcock MS, Morris M. Adjusting for network size and composition effects in exponential-family random graph models. Statistical Methodology. 2011; 8(4):319–339. [PubMed: 21691424]

Morris M, Handcock MS, Hunter DR. Specification of exponential-family random graph models: Terms and computational aspects. Journal of Statistical Software. 2008; 24(4):1–24. URL http://www.jstatsoft.org/v24/i04. [PubMed: 18612375]

Morris M, Kretzschmar M. Concurrent partnerships and the spread of HIV. AIDS. 1997; 11(5):641–648. [PubMed: 9108946]

R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2009. Version 2.6.1. URL http://www.R-project.org

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

Sampson, SF. Ph.D. thesis (university micofilm, no 69-5775). Ithaca, New York: Department of Sociology, Cornell University; 1968. A Novitiate in a Period of Change: An Experimental and Case Study of Social Relationships.

Snijders TAB. The statistical evaluation of social network dynamics. Sociological Methodology. 2001; 31(1):361–395.

Snijders, TAB. Models for longitudinal network data. In: Carrington, PJ.; Scott, J.; Wasserman, SS., editors. Models and Methods in Social Network Analysis. Cambridge, UK: Cambridge University Press; 2005. p. 215

Snijders, TAB.; Steglich, CEG.; Schweinberger, M. Modeling the co-evolution of networks and behavior. In: van Montfort, K.; Oud, J.; Satorra, A., editors. Longitudinal Models in the Behavioral and Related Sciences. Vol. chap. 3. Mahwah, NJ: Routledge Academic; 2007. p. 41-72.European Association for Methodology Series

Snijders TAB, van de Bunt CG, Steglich CEG. Introduction to stochastic actor-based models for network dynamics. Social Networks. 2010; 31(1):44–60.

Table 1

Possible transitions of a single tie variable

$y_{i,j}^{t-1}$	$\xrightarrow{}$	$(y_{i,j}^+, y_{i,j}^-)$	$\xrightarrow{}$	$y_{i,j}^t$
0	$\rightarrow$	$(0, 0)$	$\rightarrow$	0
0	$\rightarrow$	$(1, 0)$	$\rightarrow$	1
1	$\rightarrow$	$(1, 0)$	$\rightarrow$	0
1	$\rightarrow$	$(1, 1)$	$\rightarrow$	1

Table 2

MLE parameter estimates for the longitudinal friendship network

Parameter	Formation est. (s.e.)	Dissolution est. (s.e.)
Edges	−3.336 (0.320)[***]	−1.132 (0.448)[*]
Homophily (girls)	0.480 (0.269)	0.122 (0.394)
Homophily (boys)	0.973 (0.355)[**]	1.168 (0.523)[*]
F→M heterophily	−0.358 (0.330)	−0.577 (0.609)
Primary school	0.650 (0.248)[**]	0.451 (0.291)
Reciprocity	1.384 (0.280)[***]	2.682 (0.523)[***]
Transitive ties	0.886 (0.247)[***]	1.121 (0.264)[***]
Cyclical ties	−0.389 (0.133)[**]	−1.016 (0.231)[***]

Significance levels:

[*]
0.05

[**]
> 0.01

[***]
> 0.001

Table 3

Analysis of deviance for the longitudinal friendship network, comparing time-homogeneous (hom.) and time-heterogeneous (het.) parametrisations

Model	Formation			Dissolution		
	residual dev. (d.f.)	explained dev. (d.f.)	AIC	residual dev. (d.f.)	explained dev. (d.f.)	AIC
Null	1838 (1326)		1838	459 (331)		459
Edges (hom.)	924 (1325)	915 (1)***	926	431 (330)	28 (1)***	433
Full (hom.)	819 (1318)	104 (7)***	835	350 (323)	82 (7)***	366
Full (hom. except edges)	818 (1316)	2 (2)	838	344 (321)	6 (2)	364
Full (het.)	795 (1302)	23 (14)	843	314 (307)	30 (14)**	362

Significance levels:

*
  0.05

**
  > 0.01

***
  > 0.001