

University of Wollongong

Research Online

Faculty of Engineering and Information
Sciences - Papers: Part A

Faculty of Engineering and Information
Sciences

January 2012

Modeling tie duration in ERGM-based dynamic network models

Pavel N. Krivitsky

University of Wollongong, pavel@uow.edu.au

Follow this and additional works at: <https://ro.uow.edu.au/eispapers>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Modeling tie duration in ERGM-based dynamic network models

Abstract

Krivitsky and Handcock (2010) proposed a Separable Temporal ERGM (STERGM) framework for modeling social networks, which facilitates separable modeling of the tie duration distributions and the structural dynamics of tie formation. In this note, we explore the hazard structures achievable in this framework, with rst- and higher-order Markov assumptions, and propose ways to model a variety of duration distributions in this framework.

Keywords

dynamic, network, models, modeling, duration, tie, ergm

Publication Details

Krivitsky, P. N.. (2012). Modeling tie duration in ERGM-based dynamic network models. United States: The Pennsylvania State University.

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

TECHNICAL REPORTS AND PREPRINTS

Number 12-02: April 2012

**Modeling Tie Duration in ERGM-Based
Dynamic Network Models**

Pavel N. Krivitsky

Abstract

Krivitsky and Handcock (2010) proposed a Separable Temporal ERGM (STERGM) framework for modeling social networks, which facilitates separable modeling of the tie duration distributions and the structural dynamics of tie formation. In this note, we explore the hazard structures achievable in this framework, with first- and higher-order Markov assumptions, and propose ways to model a variety of duration distributions in this framework.

1 Introduction

Modeling of dynamic networks — networks that evolve over time — has applications in many fields, particularly epidemiology and social sciences. Exponential-family random graph (p^*) 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, and Robins and Pattison (2001) first described this approach. Hanneke, Fu, and Xing (2010) also define and describe what they call a Temporal ERGM (TERGM), postulating an exponential family for the transition probability from a network at time t to a network at time $t + 1$.

Holland and Leinhardt (1977), Frank (1991), and others describe *continuous-time* Markov models for evolution of social networks (Doreian and Stokman, 1997), and the most popular parametrization is the *actor-oriented* model described by Snijders (2005), which can be viewed in terms of actors making decisions to make and withdraw ties to other actors.

Arguing that “social processes and factors that result in ties being formed are not the same as those that result in ties being dissolved”, Krivitsky and Handcock (2010) introduced a separable formulation of discrete-time models for network evolution parametrized in terms of a process that controls formation of new ties and a process that controls dissolution of extant ties, in which both processes are (possibly different ERGMs), calling them *Separable Temporal ERGMs* (STERGMs). Thus, the model separates the factors that affect *incidence* of ties — the rate at which new ties are formed — from their *duration* — how long they tend to last once they do. This latter aspect, combined with its discrete-time nature, in turn, allows straightforward modeling of complex tie hazard structures and duration distributions. In this work, we discuss how a variety of these can be modeled.

In Section 2, we review the STERGM framework. In Section 3 we discuss tie hazard structures that can be induced in the framework under the first-order Markov assumption — that the transition probability does not take into account duration explicitly, while in Section 4, we propose a variety of ways to model tie hazard explicitly.

2 Separable temporal ERGM

We now review the model proposed by Krivitsky and Handcock (2010) and define some additional notation. The following overview borrows heavily from Krivitsky (2012). Using their notation, let N be the set of $n = |N|$ actors of interest, labeled $1, \dots, n$, and let $\mathbb{Y} \subseteq N \times N$ be the set of dyads (potential ties) among the actors, with $(i, j) \in \mathbb{Y}$ directed if modeling directed relations and $\{i, j\} \in \mathbb{Y}$ for undirected networks. \mathbb{Y} may be a proper subset: for example, self-loops with $i = j$ are often excluded. Then, the set of possible networks \mathcal{Y} is the power set of dyads, $2^{\mathbb{Y}}$. For a network at time $t - 1$, \mathbf{y}^{t-1} , Krivitsky and Handcock (2010) define $\mathcal{Y}^+(\mathbf{y}^{t-1}) = \{\mathbf{y} \in 2^{\mathbb{Y}} : \mathbf{y} \supseteq \mathbf{y}^{t-1}\}$ be the set of

networks that can be constructed by forming zero or more ties in \mathbf{y}^{t-1} and $\mathcal{Y}^-(\mathbf{y}^{t-1}) = \{\mathbf{y} \in 2^{\mathbb{Y}} : \mathbf{y} \subseteq \mathbf{y}^{t-1}\}$ be the set of networks that can be constructed by dissolving zero or more ties in \mathbf{y}^{t-1} .

Given \mathbf{y}^{t-1} , the network \mathbf{Y}^t at time t is modeled as a consequence of some ties being formed according to a conditional ERGM

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

specified by model parameters $\boldsymbol{\theta}^+$, sufficient statistic \mathbf{g}^+ , and, optionally, a canonical mapping $\boldsymbol{\eta}^+$; and some dissolved according to a conditional ERGM

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

specified by (usually different) $\boldsymbol{\theta}^-$, \mathbf{g}^- , and $\boldsymbol{\eta}^-$. Their normalizing constants $c_{\boldsymbol{\eta}^+, \mathbf{g}^+}(\boldsymbol{\theta}^+, \mathbf{y}^{t-1})$ and $c_{\boldsymbol{\eta}^-, \mathbf{g}^-}(\boldsymbol{\theta}^-, \mathbf{y}^{t-1})$ sum their respective model kernels over $\mathcal{Y}^+(\mathbf{y}^{t-1})$ and $\mathcal{Y}^-(\mathbf{y}^{t-1})$, respectively. \mathbf{Y}^t is then evaluated by applying the changes in \mathbf{Y}^+ and \mathbf{Y}^- to \mathbf{y}^{t-1} : $\mathbf{Y}^t = \mathbf{y}^{t-1} \cup (\mathbf{y}^+ \setminus \mathbf{y}^{t-1}) \setminus (\mathbf{y}^{t-1} \setminus \mathbf{y}^-) = \mathbf{y}^+ \setminus (\mathbf{y}^{t-1} \setminus \mathbf{y}^-) = \mathbf{y}^- \cup (\mathbf{y}^+ \setminus \mathbf{y}^{t-1})$.

Although an ERGM is a model for a whole network, many ERGM sufficient statistics have a local interpretation in the form of *change statistics* $\Delta_{i,j} \mathbf{g}(\mathbf{y}) = \mathbf{g}(\mathbf{y} \cup \{(i,j)\}) - \mathbf{g}(\mathbf{y} \setminus \{(i,j)\})$, the effect that a single dyad (i,j) has on the model's sufficient statistic and thus on its conditional probability given the rest of the network. (Hunter, Handcock, Butts, Goodreau, and Morris, 2008) For models with dyadic independence, the conditional probability is the same as the marginal probability, so $\Pr_{\boldsymbol{\eta}, \mathbf{g}}(\mathbf{Y}_{i,j} = 1) = \text{logit}^{-1}(\boldsymbol{\eta}(\boldsymbol{\theta}) \cdot \Delta_{i,j} \mathbf{g}(\mathbf{y}))$. When applied to the dissolution phase, this is the probability of an extant tie being preserved during a given time step.

When discussing the tie hazard structure of a model, we define $a(\mathbf{y}_{i,j}^t)$, the *age* of a network tie (i,j) at time t that is present at time t , to be the number of time steps that had elapsed since the tie was formed, as of time t . This is in contrast to a tie's *duration*, which is a measure of how long a tie ultimately lasts, with the distinction being analogous to that between a person's age in a given year and their ultimate lifespan. In a STERGM, a tie cannot be formed and dissolved in the same time step, so $a(\mathbf{y}_{i,j}^t) \geq 1$. Notably $a(\mathbf{y}_{i,j}^t)$ is, implicitly, a function of $\mathbf{y}_{i,j}^{t-1}$, $\mathbf{y}_{i,j}^{t-2}$, etc., up to $\mathbf{y}_{i,j}^{t-(a(\mathbf{y}_{i,j}^t)+1)}$, at which point it becomes known how long ago the tie was formed.

3 Tie hazards for first-order Markov models

We begin by considering hazard properties of first-order Markov models: models where a network \mathbf{Y}^t only depends on networks \mathbf{Y}^{t-q} , $q > 1$, through \mathbf{Y}^{t-1} .

3.1 Constant hazard

When only dyad-independent, implicitly dynamic dissolution statistics — statistics that only depend on \mathbf{y}^{t-1} through \mathbf{y}^- — are used, such as edge counts, mixing counts, and actor and edge covariates, each dyad has a geometric (discrete memoryless) distribution, although depending on the statistics used and exogenous covariates, each dyad may have a different expected duration. (Krivitsky and Handcock, 2010) Being memoryless, the geometric distribution has a constant hazard function:

$$h_{\text{Geometric}(p)}(x) = \frac{f_{\text{Geometric}(p)}(x)}{1 - F_{\text{Geometric}(p)}(x-1)} = p.$$

This is the case described and applied by Krivitsky (2012).

3.2 Non-constant hazard through dyadic dependence

When dyadic dependence is introduced into the dissolution process, the marginal hazard function of each dyad may no longer be constant. For example, if the formation model “enforces” monogamy by “encouraging” formation of an actor’s first tie and “penalizing” the formation of the second tie, while the dissolution model has a statistic that reduces the dissolution hazard of ties when they are monogamous, say,

$$\mathbf{g}^+(\mathbf{y}^+) = \left(|\mathbf{y}^+|, \sum_{i \in N} 1_{|\mathbf{y}_i^+|=1} \right), \quad \mathbf{g}^-(\mathbf{y}^-) = \left(|\mathbf{y}^-|, \sum_{i \in N} 1_{|\mathbf{y}_i^-|=1} \right),$$

with $\boldsymbol{\theta}^+ = (-, +)$ and $\boldsymbol{\theta}^- = (+, +)$ — negative coefficient on formation phase edge counts and positive coefficient on dissolution edge count (to produce relatively slow network evolution), and positive coefficients on the counts of actors with degree 1.

The dissolution phase is a draw from a dyad-dependent ERGM, so deriving the exact hazard function for this model, even conditional on \mathbf{y}^{t-1} , is intractable, but heuristically, a tie may be found in one of two scenarios:

1. It is the only tie incident on the actors on which it is incident (i.e. actors i and j are both isolates without it).
2. It is not the only tie incident on the actors on which it is incident (i.e. either i or j has other ties).

The positive coefficient $\boldsymbol{\theta}_2^-$ increases the hazard of those ties which are not their actors’ only ties (ties in Scenario 2) so ties in Scenario 2 would have a relatively high hazard, with dissolution likely until only one tie is left. However, when only one tie is left (i.e. Scenario 1), $\boldsymbol{\theta}_2^-$ reduces the hazard of that tie *and* the positive coefficient $\boldsymbol{\theta}_2^+$ reduces the probability that another tie incident on either of the actors will be formed during a given time step.

This means that a new tie that does form between actors which already have ties will have a relatively high hazard, but so will other ties incident on those actors, and if the new tie is the “survivor”, its hazard will decrease and

Table 1: Simulated equilibrium statistics under the four parameter configurations given on page 5.

Parameter configuration	Network density	Prop. of actors with 1 tie
I	0.020	0.37
D	0.022	0.73
F	0.019	0.74
B	0.020	0.96

monogamy bias in formation phase (if present) will reduce any “competition” it might face. This results in a hazard function which is high at first and decreases over time. (A new tie that forms between actors that do not already have any ties will have a constant hazard.)

To illustrate this, we conduct a simulation of a dynamic network with 50 actors. We used R package **ergm** (Hunter et al., 2008; Handcock, Hunter, Butts, Goodreau, Krivitsky, and Morris, 2012) to simulate four runs of the network process, each 11,000 time steps, with the following parameter configurations:

$\theta_{\mathbf{I}} = (\theta^+, \theta^-) = (-6, 0, 2, 0)$ total dyadic independence;

$\theta_{\mathbf{D}} = (-6, 0, 2, 2)$ formation dyadic-independent, dissolution dyad-dependent;

$\theta_{\mathbf{F}} = (-6, 2, 2, 0)$ formation dyad-dependent, dissolution dyadic-independent;
and

$\theta_{\mathbf{B}} = (-6, 2, 2, 2)$ both formation and dissolution dyad-dependent.

Since the goal of this simulation is to contrast the differences in tie hazard functions due to monogamy bias parameters, these configurations all create short-duration dynamic network processes. The four configurations produce networks with different (equilibrium) degree distributions and densities, but the quantity of interest is the edgewise hazard function, which is effectively adjusted for the number of edges in the network.

We estimate the discrete hazard function

$$\widehat{\Pr}(X = x | X \geq x) = \frac{\# \text{ ties terminated with duration } x}{\# \text{ ties terminated with duration of at least } x}$$

for tie ages $x = 1, \dots, 15$, for each parameter configuration. Durations of ties formed in the first 1,000 time steps were excluded as burn-in. The results from the simulation are given in Figure 1 and Table 1. As expected, under temporal dyadic independence ($\theta_{\mathbf{I}}$), the hazard is constant — it is $1 - \text{logit}^{-1}(2) \approx 0.119$ — and dyadic dependence limited to tie formation ($\theta_{\mathbf{F}}$) does not change this. When dissolution is dyad-dependent (both $\theta_{\mathbf{B}}$ and $\theta_{\mathbf{D}}$), the hazard is initially high, but then declines, as expected. However, it declines to a slightly lower level when both formation and dissolution have a monogamy bias ($\theta_{\mathbf{B}}$): a monogamous tie not only has its hazard reduced in the dissolution but prevents any hazard-increasing “competitors” from arising in formation. This is not the case when formation is dyad-independent ($\theta_{\mathbf{D}}$), and a monogamous tie is always potentially

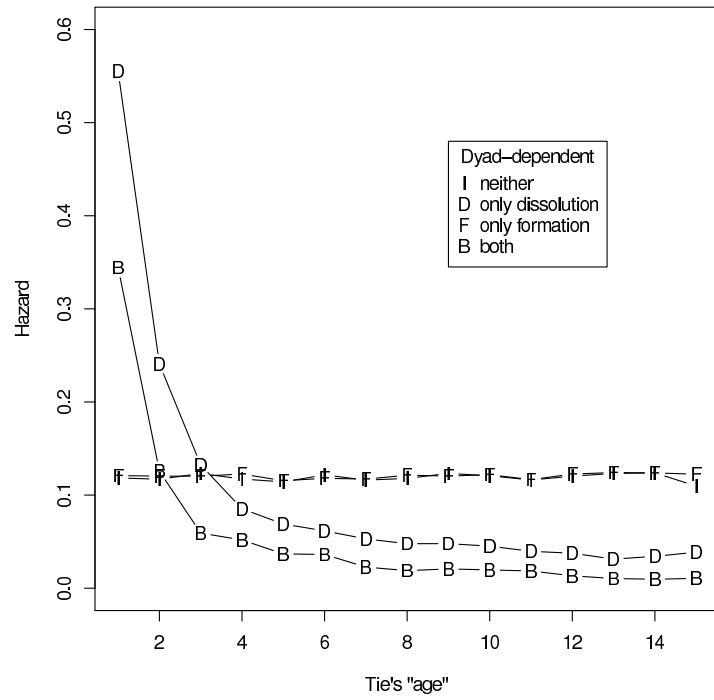


Figure 1: Estimated tie hazards under four parameter configurations given on page 5. Note that hazards at non-integral ages are meaningless, so the lines between data points are only drawn to make the series easier to follow. Note that constant hazard corresponds to geometric duration distribution.

subject to this “competition”.

Thus, non-constant dyad hazards can be induced by dyadic dependence in dissolution, and if thus induced, they may be affected by dependence in formation as well. The hazard of a given dyad during a given time step is a function only of the state of the network at the beginning of that time step, so even though the hazards are not constant, the Markov property of the process is preserved. In the following sections, we relax this, and describe explicit non-constant hazards.

4 Higher-order Markov specifications

One of the advantages discrete-time models have over continuous-time models is simpler control over the duration distribution. In the context of STERGMs and TERGMs in general, this is done by manipulating edgewise hazard functions. In this section, we describe several ways in which non-memoryless tie duration distributions can be induced.

4.1 Piecewise-constant hazard model

The simplest way to directly induce non-constant tie hazard is by modifying it by a fixed value for some set of age values. For example, let

$$\mathbf{g}^-(\mathbf{y}^-) = \left(|\mathbf{y}^-|, \sum_{(i,j) \in \mathbf{y}^-} 1_{a(\mathbf{y}_{i,j}^-) \in A} \right),$$

for some set $A \subset \mathbb{N}$. \mathbf{g}_2^- counts the number of ties in the network that whose age at the time point of interest is in set A . (For most practical purposes, A is a discrete interval.) With change statistic,

$$\Delta_{i,j} \mathbf{g}(\mathbf{y}^t, \mathbf{y}^{t-1}) = \left(\mathbf{y}_{i,j}^{t-1}, \mathbf{y}_{i,j}^{t-1} 1_{a(\mathbf{y}_{i,j}^-) \in A} \right),$$

leads to the probability of a tie being preserved in a time step of

$$\Pr_{\boldsymbol{\eta}, \mathbf{g}}(\mathbf{Y}_{i,j}^t = 1 | \mathbf{Y}_{i,j}^{t-1} = 1; \boldsymbol{\theta}^-) = \text{logit}^{-1} \left(\boldsymbol{\theta}_1^- + \boldsymbol{\theta}_2^- 1_{a(\mathbf{y}_{i,j}^-) \in A} \right),$$

and results in the probability of it being dissolved (the hazard)

$$h(a(\mathbf{y}_{i,j}^-)) = \Pr_{\boldsymbol{\eta}, \mathbf{g}}(\mathbf{Y}_{i,j}^t = 0 | \mathbf{Y}_{i,j}^{t-1} = 1; \boldsymbol{\theta}^-) = \text{logit}^{-1} \left(-\boldsymbol{\theta}_1^- - \boldsymbol{\theta}_2^- 1_{a(\mathbf{y}_{i,j}^-) \in A} \right).$$

In this case, the hazard can attain two values, and if $A = \{1, \dots, a_0\}$ for some a_0 , the duration distribution, which can be computed recursively from the hazard function h , as follows:

$$f(1) = h(1), \quad f(x) = h(x) \left(1 - \sum_{i=1}^{x-1} f(i) \right)$$

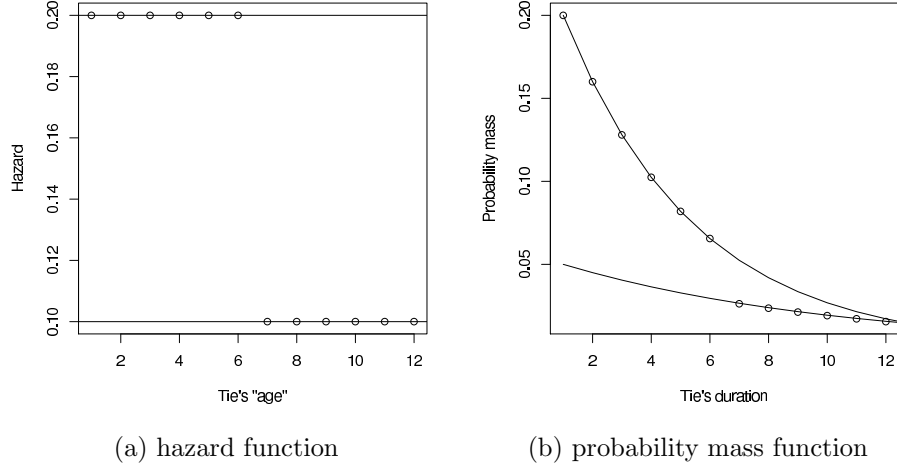


Figure 2: Piecewise-constant hazard function and resulting duration distribution.

giving duration distribution

$$f(x) = \begin{cases} (\text{logit}^{-1}(\theta_1^- + \theta_2^-))^{x-1} \text{logit}^{-1}(-\theta_1^- - \theta_2^-) & \text{for } x \leq a_0 \\ (\text{logit}^{-1}(\theta_1^- + \theta_2^-))^{a_0} (\text{logit}^{-1}(\theta_1^-))^{x-a_0-1} \text{logit}^{-1}(-\theta_1^-) & \text{for } x > a_0 \end{cases},$$

shown in Figure 2 for $a_0 = 6$, $\theta_1^- = \text{logit}(0.9)$ (hazard of 0.1) and $\theta_2^- = \text{logit}(0.8) - \theta_1^-$ (hazard of 0.2).

If A is finite, with $a_0 = \sup(A)$, this model for network evolution ($a_0 + 1$)th-order Markov: beyond a_0 , the hazard reverts to the baseline $\text{logit}^{-1}(-\theta_1^-)$, regardless of the state of networks prior to \mathbf{y}^{t-a_0-1} .

It is straightforward to extend this formulation to more hazard levels.

4.1.1 Application to formation

Rather than viewing age as an attribute of a tie, we can view it as an attribute of any dyad — the time since the last toggle in *either* direction — and in the formation phase, it can be used to, for example, penalize reformation of recently-dissolved ties.

4.2 Finite mixture model for duration

Another possible source of non-constant hazard in the duration distribution is unobserved (latent) classes of ties. An example of this is models for networks of sexual partnerships, which may be short-term or long-term. Dyad-dependent but (temporally) Markovian features of the model, such as a monogamy bias

for dissolution demonstrated in Section 3.2, can account for some of this. Alternatively, the duration distribution can be modeled as a finite mixture of simpler distributions: let there be m latent types of relationships, indexed $1, \dots, m$, $\mathbf{X}_1, \dots, \mathbf{X}_m$ be the duration distributions of different relationship types parametrized by $\boldsymbol{\omega}$, let $\boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_m$, $\pi_k > 0$, $\sum_{k=1}^m \pi_k = 1$, be their *incidence*. That is, at the time a tie forms, the probability that it is a tie of the type with duration distributed as \mathbf{X}_k is π_k . This is different from the tie class *prevalence* in the population, since that is also a function of duration of ties (that differs between types): the tie types with higher expected duration will be disproportionately more prevalent relative to their incidence. Let X be the marginal duration distribution of a tie. For notational convenience, let $\boldsymbol{\theta}^- = (\boldsymbol{\omega}, \boldsymbol{\pi})$.

Consider a simple scenario with long-term and short-term relationships, having $m = 2$, and given that a relationship is of type k , it evolves as first-order Markov, thus having a memoryless duration distribution $\text{Geometric}(\boldsymbol{\omega}_k)$, with $\boldsymbol{\omega}_1$ being the hazard of the short-term relationships and $\boldsymbol{\omega}_2$ being the hazard of the long-term relationships, so $\boldsymbol{\omega}_2 < \boldsymbol{\omega}_1$. Then, each type's pmf and cdf

$$f_{X_k}(x; \boldsymbol{\omega}_k) = (1 - \boldsymbol{\omega}_k)^{x-1} \boldsymbol{\omega}_k, \quad F_{X_k}(x; \boldsymbol{\omega}_k) = 1 - (1 - \boldsymbol{\omega}_k)^x,$$

leading to the marginal relationship duration distribution of

$$f_X(x; \boldsymbol{\theta}^-) = \sum_{k=1}^2 \pi_k (1 - \boldsymbol{\omega}_k)^{x-1} \boldsymbol{\omega}_k,$$

$$F_X(x; \boldsymbol{\theta}^-) = \sum_{k=1}^2 \pi_k (1 - (1 - \boldsymbol{\omega}_k)^x) = 1 - \sum_{k=1}^2 \pi_k (1 - \boldsymbol{\omega}_k)^x,$$

so

$$h_X(x; \boldsymbol{\theta}^-) = \frac{\sum_{k=1}^2 \pi_k (1 - \boldsymbol{\omega}_k)^{x-1} \boldsymbol{\omega}_k}{\sum_{k=1}^2 \pi_k (1 - \boldsymbol{\omega}_k)^{x-1}}.$$

Then, the probability of a tie aged x being preserved,

$$\begin{aligned} 1 - h_X(x; \boldsymbol{\theta}^-) &= 1 - \frac{\sum_{k=1}^2 \pi_k (1 - \boldsymbol{\omega}_k)^{x-1} \boldsymbol{\omega}_k}{\sum_{k=1}^2 \pi_k (1 - \boldsymbol{\omega}_k)^{x-1}} \\ &= \frac{\sum_{k=1}^2 \pi_k (1 - \boldsymbol{\omega}_k)^x}{\sum_{k=1}^2 \pi_k (1 - \boldsymbol{\omega}_k)^{x-1}}. \end{aligned}$$

Initially, the probability of a tie created in the previous time step of being preserved is

$$1 - h_X(x; \boldsymbol{\theta}^-) = \frac{\sum_{k=1}^2 \pi_k (1 - \boldsymbol{\omega}_k)^1}{\sum_{k=1}^2 \pi_k (1 - \boldsymbol{\omega}_k)^{1-1}} = 1 - \sum_{k=1}^2 \pi_k \boldsymbol{\omega}_k$$

with hazard $\sum_{k=1}^2 \pi_k \boldsymbol{\omega}_k$, the mean of the hazard of tie classes, weighted by their incidence in the population. If the tie survives, the conditional probability

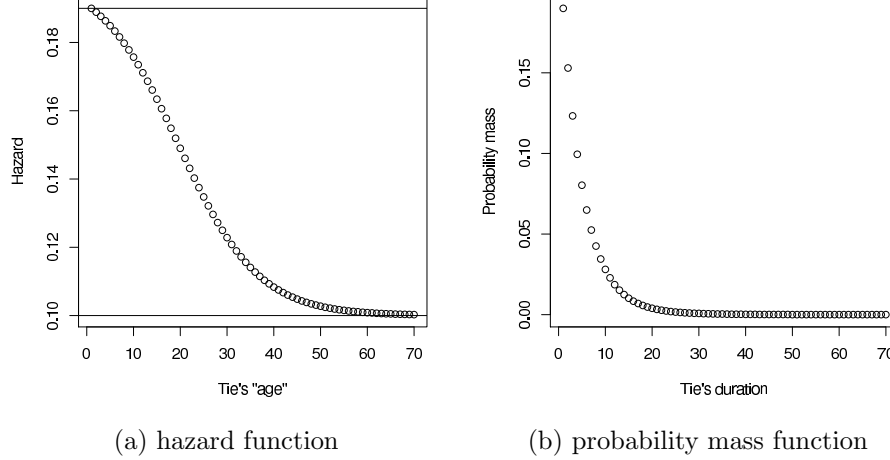


Figure 3: Probability mass function of a mixture of two geometric distributions ($0.9 \text{ Geometric}(0.2) + 0.1 \text{ Geometric}(0.1)$) and the resulting hazard function. In (a), the upper line is the hazard at a tie's first dissolution phase, $0.9 \cdot 0.2 + 0.1 \cdot 0.1$, and the lower line is the hazard of a tie which has persisted for a long time (0.1).

given its age that it was a long-term tie in the first place increases, per Bayes's Theorem, and, indeed,

$$\begin{aligned}
 \lim_{x \rightarrow \infty} (1 - h_X(x; \theta^-)) &= \lim_{x \rightarrow \infty} \frac{\sum_{k=1}^2 \pi_k (1 - \omega_k)^x}{\sum_{k=1}^2 \pi_k (1 - \omega_k)^{x-1}} \\
 &= \lim_{x \rightarrow \infty} \frac{\pi_2 (1 - \omega_2)^x}{\pi_2 (1 - \omega_2)^{x-1}} \\
 &= 1 - \omega_2,
 \end{aligned}$$

so the hazard converges to ω_2 , the hazard of long-term ties. Figure 3 for gives an example. Notably, while this process is no longer Markovian (of any order), it approaches a Markov process as the hazard converges.

More generally, for m types of ties, the marginal duration distribution of the mixture has

$$f_X(x; \theta^-) = \sum_{k=1}^m \pi_k f_{\mathbf{X}_k}(x; \omega), \quad F_X(x; \theta^-) = \sum_{k=1}^m \pi_k F_{\mathbf{X}_k}(x; \omega),$$

respectively, with the discrete hazard function

$$h_X(x; \theta^-) = \frac{f_X(x; \theta^-)}{1 - F_X(x-1; \theta^-)} = \frac{\sum_{k=1}^m \pi_k f_{\mathbf{X}_k}(x; \omega)}{1 - \sum_{k=1}^m \pi_k F_{\mathbf{X}_k}(x-1; \omega)}.$$

If, as in the example above, the combined hazard function converges to some positive value that depends only on θ^- , then this duration distribution can be

approximated in the STERGM framework as a curved exponential family. Let a_0 be the age after which the hazard is sufficiently close to constant. Then, setting

$$\boldsymbol{\eta}^-(\boldsymbol{\theta}^-) = (\text{logit}(1 - h_X(1; \boldsymbol{\theta}^-)), \dots, \text{logit}(1 - h_X(a_0; \boldsymbol{\theta}^-))),$$

and setting

$$\mathbf{g}^-(\mathbf{y}^-) = \left(\sum_{(i,j) \in \mathbf{y}^-} 1_{a(\mathbf{y}_{i,j}^-)=1}, \sum_{(i,j) \in \mathbf{y}^-} 1_{a(\mathbf{y}_{i,j}^-)=2}, \dots, \sum_{(i,j) \in \mathbf{y}^-} 1_{a(\mathbf{y}_{i,j}^-) \geq a_0} \right),$$

leading to a change statistic

$$\Delta_{i,j} \mathbf{g}(\mathbf{y}^t, \mathbf{y}^{t-1}) = \left(\mathbf{y}_{i,j}^{t-1} 1_{a(\mathbf{y}_{i,j}^-)=1}, \mathbf{y}_{i,j}^{t-1} 1_{a(\mathbf{y}_{i,j}^-)=2}, \dots, \mathbf{y}_{i,j}^{t-1} 1_{a(\mathbf{y}_{i,j}^-) \geq a_0} \right).$$

For any given dyad, if it has a tie at $\mathbf{y}_{i,j}^{t-1}$, all but one of these elements will be 0: if age $x < a_0$, then only x th element will be 1. Otherwise, only a_0 th element will be 1, giving the desired hazard structure.

4.3 Hazard induced by linear age effect

Finally, we describe a slight generalization of the piecewise-constant hazard, in which the log-odds of a dissolution (or, equivalently of preservation) of a tie are an affine function of the tie duration. Let

$$\mathbf{g}^-(\mathbf{y}^-) = \left(|\mathbf{y}^-|, \sum_{(i,j) \in \mathbf{y}^-} \left(a(\mathbf{y}_{i,j}^-) 1_{a(\mathbf{y}_{i,j}^-) < a_0} + a_0 1_{a(\mathbf{y}_{i,j}^-) \geq a_0} \right) \right),$$

with change statistic

$$\Delta_{i,j} \mathbf{g}(\mathbf{y}^t, \mathbf{y}^{t-1}) = \left(\mathbf{y}_{i,j}^{t-1}, \mathbf{y}_{i,j}^{t-1} \left(a(\mathbf{y}_{i,j}^-) 1_{a(\mathbf{y}_{i,j}^-) < a_0} + a_0 1_{a(\mathbf{y}_{i,j}^-) \geq a_0} \right) \right).$$

The restriction of the affine effect to the ages less than a_0 is to preserve the (potentially high-order) Markov property of the process, and to ensure that when $\boldsymbol{\theta}_2^- > 0$, as it would be in the short-term-long-term scenario above, no tie would have a nonzero probability of never dissolving.

This dissolution statistic could be used to approximate those in Section 4.2 more efficiently (in terms of computing power) than the approach described in that section.

5 Discussion

Given that a tie does exist, we showed via a simulation study that even in a first-order Markov model where all actors and dyads are *a priori* homogeneous,

a non-geometric duration distribution — non-constant hazard — can be induced by dyadic dependence in the dissolution process.

We have also outlined several ways in which one might explicitly model non-constant hazard durations, including piecewise-constant hazards for situations where the duration distribution is inferred from survival analysis, and for situations where there are substantive reasons to model duration distribution as a mixture.

6 Acknowledgements

This work was supported by NIH awards P30 AI27757 and 1R01 HD068395-01, NSF award HSD07-021607, ONR award N00014-08-1-1015, NICHD Grant 7R29HD034957, NIDA Grant 7R01DA012831, the University of Washington Networks Project, and Portuguese Foundation for Science and Technology Ciência 2009 Program. The author would also like to thank David Hunter and Mark Handcock, as well as the members of the University of Washington Network Modeling Group, especially Martina Morris and Steven Goodreau, for their helpful input and comments on the draft.

References

- Patrick Doreian and Franz N. Stokman, editors. *Evolution of social networks*. Routledge, 1997. ISBN 9056995383. 2
- Ove Frank. Statistical analysis of change in networks. *Statistica Neerlandica*, 45(3):283–293, 1991. 2
- Mark S. Handcock, David R. Hunter, Carter T. Butts, Steven M. Goodreau, Pavel N. Krivitsky, and Martina Morris. *ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks*. The Statnet Project, Seattle, WA, version 3.0-1 edition, March 2012. URL <http://CRAN.R-project.org/package=ergm>. Version 3.0-1. Project home page at <http://www.statnet.org>. 5
- Steve Hanneke, Wenjie Fu, and Eric P. Xing. Discrete temporal models of social networks. *Electronic Journal of Statistics*, 4:585–605, 2010. ISSN 1935-7524. doi: 10.1214/09-EJS548. 2
- Paul W. Holland and Samuel Leinhardt. A dynamic model for social networks. *Journal of Mathematical Sociology*, 5(1):5–20, 1977. 2
- David R. Hunter, Mark S. Handcock, Carter T. Butts, Steven M. Goodreau, and Martina Morris. **ergm**: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, 24(3):1–29, May 2008. ISSN 1548-7660. URL <http://www.jstatsoft.org/v24/i03>. 3, 5

- Pavel N. Krivitsky. Modeling of dynamic networks based on egocentric data with durational information. Technical Report 2012-01, Pennsylvania State University Department of Statistics, April 2012. 2, 4
- Pavel N. Krivitsky and Mark S. Handcock. A separable model for dynamic networks. *Under review*, November 2010. URL <http://arxiv.org/abs/1011.1937>. 1, 2, 4
- Garry Robins and Philippa Pattison. Random graph models for temporal processes in social networks. *Journal of Mathematical Sociology*, 25(1):5–41, 2001. 2
- Tom A. B. Snijders. Models for longitudinal network data. In Peter J. Carrington, John Scott, and Stanley S. Wasserman, editors, *Models and methods in social network analysis*, page 215. Cambridge University Press, 2005. 2