Network Science : page 1 of 28. © Cambridge University Press 2018
doi:10.1017/nws.2017.36

Simultaneous and temporal autoregressive network models

DANIEL K. SEWELL

Department of Biostatistics, University of Iowa, Iowa City, IA 52242, USA
(*e-mail:* daniel-sewell@uiowa.edu)

Abstract

While logistic regression models are easily accessible to researchers, when applied to network data there are unrealistic assumptions made about the dependence structure of the data. For temporal networks measured in discrete time, recent work has made good advances (Almquist & Butts, 2014), but there is still the assumption that the dyads are conditionally independent given the edge histories. This assumption can be quite strong and is sometimes difficult to justify. If time steps are rather large, one would typically expect not only the existence of temporal dependencies among the dyads across observed time points but also the existence of simultaneous dependencies affecting how the dyads of the network co-evolve. We propose a general observation-driven model for dynamic networks that overcomes this problem by modeling both the mean and the covariance structures as functions of the edge histories using a flexible autoregressive approach. This approach can be shown to fit into a generalized linear mixed model framework. We propose a visualization method that provides evidence concerning the existence of simultaneous dependence. We describe a simulation study to determine the method's performance in the presence and absence of simultaneous dependence, and we analyze both a proximity network from conference attendees and a world trade network. We also use this last data set to illustrate how simultaneous dependencies become more prominent as the time intervals become coarser.

Keywords: *dependence structures, dynamic networks, generalized linear mixed models, multivariate probit, observation-driven model*

1 Introduction

Co-occurrence data involves observing a set of interactions, or edges, between a set of actors. The observed edge set and actor set together form a network object. Such networks arise in multitudinous contexts, and the analysis of network objects has been of extreme importance to scientists in a wide range of fields. In particular, the analysis of network dynamics is an extremely interesting and often difficult area to work in, as temporal dependencies are added to an already complex network dependence structure.

Several classes of models for temporally measured, or dynamic, networks have been proposed, mostly over the last two decades. Each of these classes comes with pros and cons, as one would expect. The network literature is vast even for dynamic networks, and so we only touch on a few of the key classes of models before presenting our proposed approach.

Modeling dynamic networks using continuous-time Markov processes has a long history beginning with Holland & Leinhardt (1977) and continuing with several other works (e.g., Wasserman, 1980; Leenders, 1995). A very impactful work continuing the adoption of continuous-time Markov processes is the stochastic actor-oriented model (Snijders, 1996), which has since seen much methodological and software development (Ripley *et al*., 2013). In this framework, each actor forms a new edge or breaks an existing edge in order to maximize that actor's so-called objective function. This function can represent homophily on attributes or structures of the network itself, such as transitivity and reciprocity. This class of models has been very popular and useful, and allows for wide flexibility in constructing the objective function.

Another popular class of models used for static networks is the exponential random graph (ERG) models, proposed by Frank & Strauss (1986) and developed further in countless works. The ERG family of models was extended to dynamic networks by Robins & Pattison (2001), and later extended by Hanneke *et al.* (2010) and others. The temporal ERG model, or TERG model, in contrast to the stochastic actor-oriented model, assumes the network data to be generated according to a discrete time Markov process. The general idea in these ERG models is to put the probabilistic structure of the observed networks in terms of functions of sufficient statistics. These statistics often correspond to a count of some topological feature, such as triangles or k-stars. The TERGM is quite flexible in the sufficient statistics that can be included in the model, is parsimonious, and can handle complex dependencies in the network. Similar in spirit is the Separable TERGM (Krivitsky & Handcock, 2014), where both the formation and dissolution process are modeled. Unfortunately, there are a variety of problems that arise with these types of ERG models. There is the intractable normalizing constant that must be approximated, as well as degeneracy issues, or non-existence of the maximum likelihood estimators. See, e.g., Okabayashi (2011) and Jin & Liang (2013) for more on this, as well as Hummel *et al.* (2012) for remedies to some of these problems.

Stochastic blockmodels (Holland *et al*., 1983; Wang & Wong, 1987; Snijders & Nowicki, 1997) have been one of the most widely used and studied class of models for networks. The mixed membership blockmodel (Airoldi *et al*., 2008) was extended for dynamic networks by Xing *et al.* (2010). While quite useful, blockmodels suffer from an inability to capture network dependencies induced by complex features such as transitivity or reciprocity.

A large number of models fall into the class of latent space models. These models originated with Hoff *et al.* (2002) for static networks, and expanded in a variety of ways (see, e.g., Handcock *et al*., 2007; Krivitsky *et al*., 2009). These models were then extended to the dynamic context by Sarkar & Moore (2005), Durante & Dunson (2014), and Sewell & Chen (2015). Scalability remains an issue with latent space models, though some attempts have been made to alleviate this (Raftery *et al*., 2012; Salter-Townshend & Murphy, 2013), and determining the dimensionality of the latent space has attracted relatively little serious work, the main exception being work done by Durante & Dunson (2014).

Our proposed work builds off of the logistic network regression models proposed by Almquist & Butts (2013, 2014). This model provides a simple yet flexible framework for capturing the temporal dependency by modeling the mean as a function of sufficient statistics constructed from previous observations of the network. Their model has distinct advantages such as scalability, flexibility, and easy accessibility to anyone familiar with generalized linear models. The authors derive this model from the TERGM based on a clear set of assumptions. The most controversial of these is that the network dyads are conditionally independent given the network history. The problem is that the simultaneous dependence is ignored, i.e., the dependence between the co-evolving dyads. These simultaneous dependencies play an important role in the evolution of the network, especially as the intervals at which the network is observed increase (Lerner *et al.*, 2013). It is well known that ignoring extra variation in the data can, in contexts similar to our own, lead to inconsistent estimation and attenuated estimates of the parameters (Demidenko, 2013). Thus ignoring simultaneous dependence in the data will in many cases lead to poor estimation; we shall demonstrate this analytically in Section 2.3 and empirically in Section 6.

Cox (1981) used the terms "parameter driven" and "observation driven" models to describe two approaches for modeling binary time series data. In the context of dynamic network analysis, we can think of the latent space approach as the analog to parameter-driven models, where the temporal dependencies of the network are driven through some latent variables evolving through, say, a Markov process. Our proposed model follows what may be considered an observation-driven approach, where both the simultaneous and temporal dependencies are driven by some functions of the lagged observed networks. More specifically, our proposed approach captures temporal dependence through modeling the mean as a function of lagged networks and similarly captures the simultaneous dependence through modeling the covariance as a function of lagged networks.

An important motivation for this work was accessibility to appropriate network methodology for those without extensive statistical background. We believe that those familiar with generalized linear mixed models (GLMMs) (see Section 4) should be able to easily understand and utilize our proposed approach, and software will be made available on the author's website to further facilitate accessibility. While using a familiar framework, we account for both temporal and simultaneous dependence, thus avoiding the adverse inferential impacts that we otherwise would expect to occur by ignoring these two sources of variation.

In Section 2, we present our proposed methodology, as well as some suggestions for appropriately choosing the mean and covariance functions. In Section 3, we describe our approach to estimation, with the details and selected proofs given in the appendix. Section 4 generalizes our approach by fitting our method into the familiar GLMM framework. In Section 5, we describe a visualization approach to evaluating the evidence regarding the existence and impact of simultaneous dependence in the data. In Section 6, we present a simulation study that examines the performance of our model in the presence and absence of simultaneous dependencies. In Section 7, we analyze two real data sets, illustrating the utility of our method and the importance of accounting for simultaneous dependence in real data, as well as illustrating how simultaneous dependence becomes more prominent as time intervals become coarser.

2 Methodology

2.1 Context and notation

We assume that we have n objects, or *actors*, each of which may have some interactions or relationships with the other actors. If such an interaction/relationship exists between actors i and j, we say there is an *edge* between them. We assume that the set of actors are constant over time, though the edges themselves may exist during any subset of all possible time points. Here, we assume the data are collected at discrete time points. Collectively, the set of actors and the time-varying set of edges define the dynamic network. The data obtained can then be represented by a three-dimensional tensor, or equivalently a sequence of adjacency matrices, where each adjacency matrix, denoted as A_t, $t = 0, 1, \ldots, T$, is an $n \times n$ matrix corresponding to the edges that exist at time t. That is, the (i, j)-th entry of A_t, A_{ijt}, equals one if there is an edge from i to j at time t and zero otherwise. The diagonal entries of each adjacency matrix hold no meaning unless so-called self-loops are allowed, that is, an actor may send an edge to itself. For the purposes of clarity in our exposition, we will assume in Section 2 that such self-loops are allowed as this helps facilitate the mathematical description of the model and its properties; it is trivial to translate the presented model to the context of no self-loops. However, because (1) self-loops are relatively rare in practice, and (2) the derivations of our estimation algorithm requires additional non-trivial steps when self-loops are not allowed, the derivations provided in our appendices assume the diagonal elements of the A_t's are meaningless. Additionally, the data in Sections 6 and 7 do not have self-loops.

We also assume there exist some exogenous covariate information with which we would like to explain or predict the edge probabilities. These covariates may by static (e.g., race or gender) or time-varying (e.g., income or marital status). In the remainder of the paper, we will treat the covariates as though they are time-varying with the understanding that static covariates may be treated as such simply by replicating them from one time point to the next. We denote the dyadic covariate information by the $n \times n$ matrices $X_{\ell t}$, $\ell = 1, \ldots, p_1$, $t = 1, \ldots, T$. For notational convenience, we will denote a linear combination of equal sized matrices as $\langle \boldsymbol{\beta}, \mathscr{X}_t \rangle := \sum_{\ell=1}^{p_1} \beta_\ell X_{\ell t}$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{p_1})$ and $\mathscr{X}_t$ is a 3-dimensional array whose ℓth slice is $X_{\ell t}$.

As will be seen shortly, we shall be focusing on covariance structures, and hence it is natural to implement a probit type model for our binary dyadic data (although we will generalize the work in Section 4). We thus assume that there are some underlying matrices of normal random variables A_t^* that directly correspond to A_t via the surjective function $A_{ijt} = \mathbf{1}_{\{A_{ijt}^* > 0\}}$.

2.2 Observation-driven model

The proposed model is an observation-driven approach, rather than parameter-driven. That is, we may write the conditional mean of A_t^* as a function of $A_0, \ldots, A_{t-1}$ rather than as a function of some unobservable noise process. Observation-driven approaches for temporal binary data have been well studied in simpler contexts. While some complicated mean functions have been proposed (e.g., Shephard, 1995), often it is the simple and intuitive

$$\mathbb{E}(A^*_{ijt}|A_{ij(t-1}, A_{it(t-2)}, \ldots) = \sum_{\ell=1}^{p_1} \beta_\ell X_{\ell t}[i,j] + \sum_{\ell=1}^{p_2} \theta_\ell A_{ij(t-\ell)},$$

(e.g., Cox, 1981; Zeger & Qaqish, 1988), where $X[i,j]$ is the (i,j)-th entry of the matrix X. However, this simplistic mean function is insufficient for complex network objects. With this in mind, we will allow the second term of the mean of A^*_t to be $\langle \boldsymbol{\theta}, \mathscr{G}_t \rangle := \langle \boldsymbol{\theta}, \mathscr{G}(A_{t-1}, A_{t-2}, \ldots) \rangle$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{p_2})$, and $\mathscr{G}_t$ maps the previous adjacency matrices onto the space of $n \times n \times p_2$ tensors, i.e., $\mathscr{G}_t$ uses the previous adjacency matrices to construct p_2 new $n \times n$ matrices.

Note that p_2 does not refer to the number of lagged time points as in the simple binary time series model, but rather can encompass a number of salient features of the previous adjacency matrices, such as stability, reciprocity, or transitivity. As a simple example, if we include stability and reciprocity for up to a lag of two time points, then $p_2 = 4$ and the slices of $\mathscr{G}_t$ are A_{t-1}, A'_{t-1}, A_{t-2}, and A'_{t-2}. These p_2 covariates involving functions of the lagged network can thus be used in sophisticated ways to explain the temporal dependencies, i.e., the dependence between A_{ijt} and $A_{k\ell s}$, $t \neq s$. For examples of other ways to construct $\mathscr{G}_t$, see Table 1 or the appendices of Almquist & Butts (2014).

Networks are complex objects, however, and attempting to capture all dependencies through the mean structure alone is insufficient, particularly as the intervals between time points grow larger. One would typically expect not only the existence of *temporal* dependencies through which the network at varying time points are dependent, but also *simultaneous* dependencies that dictate how the dyads of the network co-evolve. Thus, we should be quite concerned with appropriately modeling the second moments of the A^*_{ijt}'s.

With this motivation in mind, we begin with the following multivariate probit model. Let $\mathscr{A}_t$ be equal to $\mathrm{vec}(A^*_t)$. Then set

$$\mathbb{E}(A^*_t|A_{t-1}, A_{t-2}, \ldots) = \langle \boldsymbol{\beta}, \mathscr{X}_t \rangle + \langle \boldsymbol{\theta}, \mathscr{G}_t \rangle \tag{1}$$

$$\mathrm{Cov}(\mathscr{A}_t) = \Sigma_{A^*,t}. \tag{2}$$

Note that $\Sigma_{A^*,t}$ determines the covariance structure among the n^2 dyads, and hence has $\mathscr{O}(n^4)$ parameters. Clearly, it would not be possible to estimate such an unconstrained $\Sigma_{A^*,t}$ outside of the context of small n large T, nor is this unconstrained covariance structure what one would expect to see in reality. Going to the extreme of constraining $\Sigma_{A^*,t}$ to be the identity matrix (and thus ignoring simultaneous dependence entirely) leads to the model presented in Almquist & Butts (2014), and hence what is presented here can be thought of as an alternative generalization of their methods (the TERGM is the original motivation for and generalization of their approach).

2.3 Ignoring simultaneous dependencies

Here, we make a short note on estimation errors associated with ignoring existing variablity in the data. Demidenko (2013) gives a short discussion on these types of issues with regard to GLMMs (see chapter 7). For our context, suppose we may write the normal random variables A^*_{ijt}'s as

$$A^*_{ijt} = \langle \boldsymbol{\beta}, \mathscr{X}_t \rangle [i,j] + \langle \boldsymbol{\theta}, \mathscr{G}_t \rangle [i,j] + s_{it} + r_{jt} + E_{ijt},$$

where s_{it}, r_{it}, and E_{ijt} are zero mean normal random variables (possibly correlated in complex ways, though letting $s_{it}, r_{it} \perp E_{ijt} \forall i, j, t$). Then we have the following proposition, the proof of which is given in Appendix 8.

Proposition.

$$\mathbb{P}(A_{ijt} = 1 | \boldsymbol{\beta}, \boldsymbol{\theta}) = \boldsymbol{\Phi}\left(\frac{\mathbb{E}(A^*_{ijt})}{\sqrt{Var(E_{ijt}) + Var(s_{it} + r_{jt})}} \right), \tag{3}$$

where $\boldsymbol{\Phi}(\cdot)$ *is the CDF of a standard normal distribution, and* $\mathbb{E}(A^*_{ijt})$ *is given in Equation (1).*

Now consider the very simple example where we have

$$\begin{pmatrix} s_{it} \\ r_{it} \end{pmatrix} \overset{iid}{\sim} N\left(\mathbf{0}, \begin{pmatrix} \tau_s & 0 \\ 0 & \tau_r \end{pmatrix} \right)$$

and constant variance for the E_{ijt}'s. We can quickly see that should we ignore simultaneous dependence, any attempts to estimate $(\boldsymbol{\beta}, \boldsymbol{\theta})$ would in fact unintentionally lead to the attenuated estimation of $(\boldsymbol{\beta}, \boldsymbol{\theta})$ scaled by $Var(E_{ijt}) + \tau_s + \tau_r$. For more general cases, when $Var(s_{it} + r_{jt})$ is time dependent or dependent on the actors i and j, it is unclear what, if anything, any naive estimates of $(\boldsymbol{\beta}, \boldsymbol{\theta})$ are actually estimating.

2.4 Simultaneous and temporal autoregressive model

A middle ground between fully ignoring simultaneous dependence and using a saturated covariance matrix $\Sigma_{A^*,t}$ would be to assume that there ought to be some connection with the covariance between two dyads and the actors that are incident on those two dyads. This simple and intuitive idea will eventually lead us to a model resembling the social relations model (Warner *et al.*, 1979), having the form

$$A^*_{ijt} = \text{ mean structure } + \text{ sender effects } + \text{ receiver effects } + \text{ residuals}$$

(the final form is given in Equation (10)). To get there, we begin by introducing the following definition.

Definition. An $n \times n$ matrix A^* has a *role-based additive covariance structure* if

$$\begin{aligned} &\mathrm{Cov}(A^*_{ij}, A^*_{k\ell}) \\ &\quad = \Sigma_s[i,k] + \Sigma_r[j,\ell] + \Sigma_{sr}[i,\ell] + \Sigma_{sr}[k,j] + \sigma^2_R 1_{[\{(i,j)=(k,\ell)\} \cup \{(i,j)=(\ell,k)\}]} \\ &\qquad + \sigma^2_\epsilon 1_{[(i,j)=(k,\ell)]}, \end{aligned} \tag{4}$$

where Σ_s, Σ_r, and Σ_{sr} are $n \times n$ covariance matrices that represents, respectively, the covariance among the senders of the dyads, the receivers of the dyads, and between the senders and the receivers, and where σ^2_R and σ^2_ϵ correspond to pair and dyad variance, respectively.

A role-based additive covariance structure can be interpreted to mean that the covariance between any two dyads (i,j) and (k,ℓ) can be explained by how similar

i and k are as senders, how similar j and ℓ are as receivers, how i and ℓ relate to each other as sender and receiver, respectively, and similarly for k and j, the variability due to reciprocated dyads, and the inherent variability between the dyads.

The role-based additive covariance structure has a nice representation that lends itself well to estimation. To demonstrate this, we provide the following theorem.

Theorem. *The following are equivalent.*

1. *The A^*_{ijt}'s are jointly normal with a role-based additive covariance structure and mean given by Equation (1).*
2. $$\mathscr{A}_t \sim N\Big(\mathrm{vec}(\langle\boldsymbol{\beta}, \mathscr{X}_t\rangle + \langle\boldsymbol{\theta}, \mathscr{G}_t\rangle), \\ J_n \otimes \Sigma_{st} + \Sigma_{rt} \otimes J_n + \mathbb{1}_n \otimes \Sigma_{srt} \otimes \mathbb{1}'_n + \mathbb{1}'_n \otimes \Sigma'_{srt} \otimes \mathbb{1}_n \\ + \sigma^2_R M_R + (\sigma^2_\epsilon + \sigma^2_R) I_{n^2}\Big), \tag{5}$$
   where $\mathbb{1}_k$ is the $k \times 1$ vector of 1's, J_k equals $\mathbb{1}_k\mathbb{1}'_k$, and I_k is the $k \times k$ identity matrix, and where M_R is a matrix such that for $1 \leqslant i \neq j \leqslant n$, $M_R[(j-1)n + i, (i-1)n + j] = 1$ and $M_r[\ell, m] = 0$ everywhere else.
3. *$A^*_t = \langle\boldsymbol{\beta}, \mathscr{X}_t\rangle + \langle\boldsymbol{\theta}, \mathscr{G}_t\rangle + \boldsymbol{s}_t\mathbb{1}' + \mathbb{1}\boldsymbol{r}' + E_t$, where*
   $$\begin{pmatrix} \boldsymbol{s}_t \\ \boldsymbol{r}_t \end{pmatrix} \overset{iid}{\sim} N\left(\mathbf{0}, \begin{pmatrix} \Sigma_{st} & \Sigma_{srt} \\ \Sigma'_{srt} & \Sigma_{rt} \end{pmatrix}\right), \\ (E_t[i,j], E_t[j,i])' \overset{iid}{\sim} N\big(\mathbf{0}, \sigma^2_\epsilon I_2 + \sigma^2_R J_2\big). \tag{6}$$

The proof is given in Appendix 8.

Unconstrained, the covariance structure of Equation (6) still has $O(n^2)$ parameters to be estimated. The question then is how to appropriately, yet parsimoniously, represent the covariance structure of $(\boldsymbol{s}_t, \boldsymbol{r}_t)$. In response, we pose the following question: if the features found in $(A_{t-1}, A_{t-2}, \ldots)$ can appropriately capture the temporal dependence through the mean structure, may we not also capitalize on the information stored in $(A_{t-1}, A_{t-2}, \ldots)$ to estimate the simultaneous dependence through the covariance structure? (This is similar in principle to ARCH models. See Engle, 1982). We propose using an autoregressive model on the covariance structure of $(\boldsymbol{s}_t, \boldsymbol{r}_t)$ as well as on the mean structure of A^*_t, so that $\mathrm{Cov}(\mathscr{A}_t | \mathscr{A}_{t-1}, \mathscr{A}_{t-2}, \ldots)$ is some function of $(\mathscr{A}_{t-1}, \mathscr{A}_{t-2}, \ldots)$.

Specifically, we consider $\mathrm{Cov}(\boldsymbol{s}_t, \boldsymbol{r}_t)$ with the following structure:

$$\Sigma_{st} = \sum_{k=1}^{K_s} \tau_{sk} H_{skt} \qquad \Sigma_{rt} = \sum_{k=1}^{K_r} \tau_{rk} H_{rkt} \qquad \Sigma_{srt} = \sum_{k=1}^{K_{sr}} \tau_{srk} H_{srkt} \tag{7}$$

where τ_{sk}, τ_{rk}, and τ_{srk} are positive valued parameters, H_{skt}, H_{rkt}, and H_{srkt} are functions of $(A_{t-1}, A_{t-2}, \ldots)$, and $H_{skt}, H_{rkt} \in \mathbb{S}^n_+$ for all k. Here, $\mathbb{S}^n_+$ denotes the positive semi-definite (PSD) cone. Writing $\mathrm{Cov}(\boldsymbol{s}_t, \boldsymbol{r}_t)$ in this manner, i.e., as a linear combination of PSD matrices, is similar in principle to covariance structures studied for many decades (e.g., Anderson, 1973). Constructing the covariance matrices in this manner allows us to use the data to represent complex simultaneous dependence, while reducing the number of parameters from $O(n^2)$ to $K_s + K_r + K_{sr}$.

Note that this does not automatically ensure that $\Sigma_{A^*,t} \in \mathbb{S}_+^{n^2}$, and so some care is still needed. To ensure that we have a valid covariance matrix, we constrain $K_{sr} \leqslant \min\{K_s, K_r\}$, and for $1 \leqslant k \leqslant K_{sr}$ impose the constraint that

$$\begin{pmatrix} \tau_{sk} H_{skt} & \tau_{srk} H_{srkt} \\ \tau_{srk} H'_{srkt} & \tau_{rk} H_{rkt} \end{pmatrix} \in \mathbb{S}_+^{(2n)}. \tag{8}$$

The structure found in Equation (7) allows us to further decompose $\boldsymbol{s}_t$ and $\boldsymbol{r}_t$ as

$$\begin{aligned} \boldsymbol{s}_t &= \sum_{k=1}^{K_s} \boldsymbol{s}_{kt}, \quad \boldsymbol{s}_{kt} \overset{\text{ind}}{\sim} N(\boldsymbol{0}, \tau_{sk} H_{skt}) \\ \boldsymbol{r}_t &= \sum_{k=1}^{K_r} \boldsymbol{r}_{kt}, \quad \boldsymbol{r}_{kt} \overset{\text{ind}}{\sim} N(\boldsymbol{0}, \tau_{rk} H_{rkt}) \end{aligned} \qquad \mathrm{Cov}(\boldsymbol{s}_{kt}, \boldsymbol{r}_{k't}) = \begin{cases} \tau_{srk} H_{srkt} & \text{if } 1 \leqslant k = k' \leqslant K_{sr} \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

This then results in having our multivariate probit model with role-based additive covariance structure represented as

$$A_t^* = \langle \boldsymbol{\beta}, \mathscr{X}_t \rangle + \langle \boldsymbol{\theta}, \mathscr{G}_t \rangle + \left(\sum_{k=1}^{K_s} \boldsymbol{s}_{kt} \right) \mathbb{1}' + \mathbb{1} \left(\sum_{k=1}^{K_r} \boldsymbol{r}_{kt} \right)' + E_t. \tag{10}$$

2.5 *Broader context of sender/receiver effects*

By first assuming an intuitive form for the covariance of the dyads, we are able to arrive at a multivariate mixed effects probit model for the dynamic network, using individual sender and receiver effects. The use of individual sender and receiver effects has a long history in network analysis, starting with Warner *et al.* (1979). In nearly all cases, the additive sender and receiver effects can be put within the framework described above by setting $K_s = K_r = K_{sr} = 1$ and $H_{s1t} = H_{r1t} = H_{sr1} = I_n$. An important work using this is the p_2 model of Duijn *et al.* (2004). This work was built off of the p_1 model of Holland & Leinhardt (1981), which was not motivated by modeling an appropriate covariance structure. Latent space models have incorporated additive sender/receiver effects as well, such as Hoff (2005) (which also incorporated multiplicative effects), and Krivitsky *et al.* (2009).

The above-referenced works are all concerned with static networks. Westveld & Hoff (2011) used the ideas of sender and receiver effects to model the covariance of the data for dynamic networks. As with the others, they constrain $K_s = K_r = K_{sr} = 1$ and $H_{s1t} = H_{r1t} = H_{sr1} = I_n$, while also assuming AR processes on the sender and receiver effects (and on the residuals). While there is merit in this approach, we still prefer capturing the temporal dependency through the observation-driven model. This is primarily because one may utilize specific network features, such as stability, reciprocity, transitivity, etc. to help explain the temporal dependencies. In this way, one may argue that there is more flexibility, and researchers can investigate the specific effects of various network features.

The way in which we use sender and receiver effects here differs in two important ways from previous uses. First, the constraints on the covariance matrix of the dyads are relaxed to allow $\Sigma_{A^*,t}$ to be dense, thus generalizing the way that researchers have in the past used sender and receiver effects in their models. Second, we incorporate past data to make the parameter space parsimonious. That is, a dense covariance matrix with $O(n^4)$ unknowns can, by leveraging past information, be estimated using $K_s + K_r + K_{sr}$ parameters. For an example of how we may do this in practice, see Section 2.6.

2.6 An example of operationalization

One of the strengths of Equations (1) and (7) is the flexibility in choosing the features of the previous adjacency matrices to be used in constructing the mean and covariance functions. In this subsection, we provide an example, based on sociological principles as well as previous research in statistical models for networks, with the intention that researchers using the STAR model may use whatever network features are most appropriate for their particular context.

Fortunately for the analyst looking at dynamic network data, there has been much focus in the social science literature on the salient structures of networks. To quote Wasserman & Faust (1994),

> Many researchers have shown, using empirical studies, that social network data possess strong deviations from randomness. … data often fail to agree with predictions from [models with assumptions, such as equal popularity, lack of transitivity, or no reciprocity].

Krackhardt & Handcock (2007) made note that it has long been argued that "the triad, not the dyad, is the fundamental social unit that needs to be studied" (see also Simmel & Wolff, 1950), which further emphasizes that transitivity is, to quote Wasserman & Faust (1994) again, "indeed a compelling force in the organization of social groups."

These notions then motivate the construction of $\mathscr{G}_t$, the three-dimensional tensor whose ℓth slice is denoted by $\mathscr{G}_{\ell t}$, as given in Table 1. We can categorize these eight structures of the network in the following terms. $\mathscr{G}_{1t}$ and $\mathscr{G}_{2t}$ correspond to first-order structures, that is, features of the network that relate to individual actors only. $\mathscr{G}_{3t}$ and $\mathscr{G}_{4t}$ correspond to second-order structures, that is, features of the network that relate to dyads. $\mathscr{G}_{5t}$ to $\mathscr{G}_{8t}$ correspond to third-order structures, that is, features of the network that relate to triads. In particular, $\mathscr{G}_{5t}$ to $\mathscr{G}_{7t}$ correspond to transitivity in the network, i.e., the probability that a transitive relation exists, while $\mathscr{G}_{8t}$ corresponds to a cycle, i.e., the probability that a three-cycle will be completed. These last four structures are depicted visually in Figure 1, where we are considering the probability of an edge from i to j and visualizing the transitive and cyclic triadic relations involving the third actor k. One note regarding $\mathscr{G}_{1t}$ to $\mathscr{G}_{8t}$ is that these same features could of course be trivially extended to more than just a lag of 1 whenever appropriate.

Table 1. Example of how to construct $\mathscr{G}_t$, incorporating first-, second-, and third-order structures.

(out degree)	$\mathscr{G}_{1t} = A_{t-1}J_n$	$\mathscr{G}_{1t}[i,j] = \sum_{k=1}^n A_{ik(t-1)}$
(in degree)	$\mathscr{G}_{2t} = J_nA_{t-1}$	$\mathscr{G}_{2t}[i,j] = \sum_{k=1}^n A_{kj(t-1)}$
(stability)	$\mathscr{G}_{3t} = A_{t-1}$	$\mathscr{G}_{3t}[i,j] = A_{ij(t-1)}$
(reciprocity)	$\mathscr{G}_{4t} = A'_{t-1}$	$\mathscr{G}_{4t}[i,j] = A_{ji(t-1)}$
(transitivity 1)	$\mathscr{G}_{5t} = A_{t-1}A_{t-1}$	$\mathscr{G}_{5t}[i,j] = \sum_{k=1}^n A_{ik(t-1)}A_{kj(t-1)}$
(transitivity 2)	$\mathscr{G}_{6t} = A_{t-1}A'_{t-1}$	$\mathscr{G}_{6t}[i,j] = \sum_{k=1}^n A_{ik(t-1)}A_{jk(t-1)}$
(transitivity 3)	$\mathscr{G}_{7t} = A'_{t-1}A_{t-1}$	$\mathscr{G}_{7t}[i,j] = \sum_{k=1}^n A_{ki(t-1)}A_{kj(t-1)}$
(cycle)	$\mathscr{G}_{8t} = A'_{t-1}A'_{t-1}$	$\mathscr{G}_{8t}[i,j] = \sum_{k=1}^n A_{ki(t-1)}A_{jk(t-1)}$

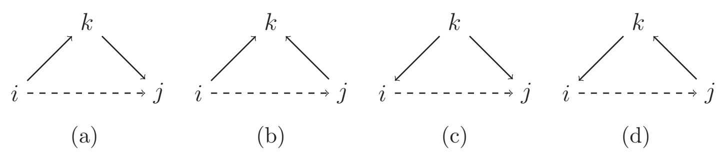


Fig. 1. Network structures which are being summed over k to determine the mean of A^*_{ijt}. (a) $\mathscr{G}_{5t}[i,j]$. (b) $\mathscr{G}_{6t}[i,j]$. (c) $\mathscr{G}_{7t}[i,j]$. (d) $\mathscr{G}_{8t}[i,j]$.

Intuitively, Σ_{st} and Σ_{rt} ought to reflect how similar actors behave as senders and receivers, respectively. We therefore suggest setting $K_s = K_r = 2$, $K_{sr} = 1$, and

$$
\begin{aligned}
H_{s1t} &= H_{r1t} = H_{sr1t} = I_n \\
H_{s2t} &= D^{-1/2}_{out,(t-1)} A_{t-1} A'_{t-1} D^{-1/2}_{out,(t-1)}, \\
H_{r2t} &= D^{-1/2}_{in,(t-1)} A'_{t-1} A_{t-1} D^{-1/2}_{in,(t-1)},
\end{aligned}
\tag{11}
$$

where $D_{out,(t-1)}$ and $D_{in,(t-1)}$ are diagonal matrices whose diagonal entries are the out-degrees and in-degrees of A_{t-1}, respectively. The (i,j)-th entry of H_{s2t} then is the number of actors to whom both i and j sent edges scaled by the geometric mean of the total number of actors to whom i and j each sent edges. In this manner, we are capturing the intended notion of similarity between senders while enforcing H_{s2t} to be PSD. In fact, H_{s2t} is a valid correlation matrix. Similarly for H_{r2t}, a note on the practical implementation of this is that to avoid the possibility of dividing by zero anywhere, in our analyses we set the diagonal of A_{t-1} to be $\mathbb{1}$ when computing H_{s2t} and H_{r2t}. To ensure that the covariance of $(\boldsymbol{s}_t, \boldsymbol{r}_t)$ is PSD, and hence the covariance of $\mathscr{A}_t$ is PSD, we constrain

$$
\Omega := \begin{pmatrix} \tau_{s1} & \tau_{sr1} \\ \tau_{sr1} & \tau_{r1} \end{pmatrix} \in \mathbb{S}^2_+. \tag{12}
$$

2.7 *Undirected networks*

The above proposed methodology has focused on directed dynamic networks. Simplifying to an undirected dynamic network implies that Equations (4) and (5) can be written

$$
\begin{aligned}
&\mathrm{Cov}(A^*_{ijt}, A^*_{k\ell t}) = \Sigma_{st}[i,k] + \Sigma_{st}[j,\ell] + \Sigma_{st}[i,\ell] + \Sigma_{st}[k,j] + \sigma^2 1_{[(i,j)=(k,\ell)]} \\
&\quad \Leftrightarrow \mathrm{Cov}(\mathscr{A}_t) = J_n \otimes \Sigma_{st} + \Sigma_{st} \otimes J_n + \mathbb{1} \otimes \Sigma_{st} \otimes \mathbb{1}' + \mathbb{1}' \otimes \Sigma_{st} \otimes \mathbb{1} + \sigma^2 I.
\end{aligned}
\tag{13}
$$

The estimation algorithm given in Section 3 can be adapted to the undirected case; some of the details which are not obvious are given in Appendix 8. In the analysis of Section 7.1, we set

$$\Sigma_{st} = \tau_s H_{st}, \text{ where } H_{st} = D_{(t-1)}^{-1/2} A_{t-1} A_{t-1} D_{(t-1)}^{-1/2} \tag{14}$$

and D_t is the diagonal matrix whose diagonal entries are the degrees of the actors corresponding to A_t, i.e., $A_t\mathbb{1}$. For autoregressive mean terms, we used

$$\begin{array}{lll} \textbf{(degree)} & \mathcal{G}_{1t} = A_{t-1}J_n + J_nA_{t-1} & \mathcal{G}_{1t}[i,j] = \sum_{k=1}^n \left(A_{ik(t-1)} + A_{jk(t-1)}\right) \\ \textbf{(stability)} & \mathcal{G}_{2t} = A_{t-1} & \mathcal{G}_{2t}[i,j] = A_{ij(t-1)} \\ \textbf{(triangle)} & \mathcal{G}_{3t} = A_{t-1}A_{t-1} & \mathcal{G}_{3t}[i,j] = \sum_{k=1}^n A_{ik(t-1)}A_{jk(t-1)}. \end{array}$$

3 Variational Bayes estimation

From a Bayesian perspective, we would like to make posterior inference regarding the mean parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ as well as the variance components τ_{sk}'s, τ_{rk}'s, and τ_{srk}'s. In what follows, we will assume the particular formulation given in Section 2.6. Thus of interest is deriving $\pi(\boldsymbol{\beta}, \boldsymbol{\theta}, \Omega, \tau_{s2}, \tau_{r2}, \sigma_R^2 | \{A_t\}_{t=0}^T)$. Note that just as with any probit model, σ_ϵ^2 is constrained to equal 1 for identifiability. We assign the following priors on the model parameters.

$$\begin{aligned} (\boldsymbol{\beta}', \boldsymbol{\theta}')' &\sim N(\mathbf{0}, \operatorname{diag}(\sigma_\beta^2, \ldots, \sigma_\beta^2, \sigma_\theta^2, \ldots, \sigma_\theta^2)), \\ \tau_{s2} &\sim IG(a_{s0}, b_{s0}), \\ \tau_{r2} &\sim IG(a_{r0}, b_{r0}), \\ \Omega &\sim IW(a_{\Omega 0}, B_{\Omega 0}), \\ \sigma_R^2 &\sim IG(a_{R0}, b_{R0}), \end{aligned}$$

where $\operatorname{diag}(\sigma_\beta^2, \ldots, \sigma_\beta^2, \sigma_\theta^2, \ldots, \sigma_\theta^2)$ is the $(p_1 + p_2) \times (p_1 + p_2)$ diagonal matrix whose first p_1 diagonal entries are σ_β^2 and whose last p_2 diagonal entries are σ_θ^2, $IG(a, b)$ is the inverse gamma distribution with shape parameter a and scale parameter b, and $IW(a, B)$ denotes the inverse Wishart distribution with degrees of freedom a and scale matrix B.

Rather than implementing a computationally expensive MCMC algorithm, we implement a mean field variational Bayes (VB) algorithm. This estimation technique finds an approximation of the posterior distribution such that the Kullback–Leibler divergence between this approximation and the true posterior distribution is minimized. This minimization is done under the constraint that the approximated posterior density is a product of densities corresponding to a partition of the unknown model parameters. See, e.g., Gelman *et al.* (2004) (Chapter 13) for a brief overview of variational methods.

While much faster than MCMC, one issue with the VB algorithm is a negative bias of the variance components. In our analyses, we found that the bias was so strong in σ_R^2 as to render the reciprocity effects negligible, which led to poorer performance overall. To address this, first consider further data augmentation via the $n \times n$ symmetric matrices of dyad-pair specific random effects R_t, such that $R_t[i, j] = R_t[j, i] \overset{iid}{\sim} N(0, \sigma_R^2)$. That is, we now have the equivalent form of

Equation (10)

$$A_t^* = \langle \boldsymbol{\beta}, \mathscr{X}_t \rangle + \langle \boldsymbol{\theta}, \mathscr{G}_t \rangle + \left(\sum_{k=1}^{K_s} \boldsymbol{s}_{kt} \right) \mathbb{1}' + \mathbb{1} \left(\sum_{k=1}^{K_r} \boldsymbol{r}_{kt} \right)' + R_t + \widetilde{E}_t, \tag{15}$$

where $\widetilde{E}_t$ is a matrix of *iid* normal random variables with zero mean and variance σ_ϵ^2. To prohibit σ_R^2 from shrinking to zero, we treat it as a hyperparameter for the R_t's. While not ideal, this seemed to improve overall performance.

The specific form of the approximated posterior is

$$\begin{aligned} &\pi(\boldsymbol{\beta}, \boldsymbol{\theta}, \tau_{s2}, \tau_{r2}, \Omega, \sigma_R^2, \{A_t^*\}_{t=1}^T, \{\boldsymbol{s}_{1t}, \boldsymbol{r}_{1t}, \boldsymbol{s}_{2t}, \boldsymbol{r}_{2t}\}_{t=1}^T, \{R_t\}_{t=1}^T | \{A_t\}_{t=0}^T) \\ &\quad \approx q_1(\boldsymbol{\beta}, \boldsymbol{\theta}) q_2(\tau_{s2}, \tau_{r2}, \Omega) q_3(\{\mathscr{A}_t\}_{t=1}^T) q_4(\{\boldsymbol{s}_{1t}, \boldsymbol{r}_{1t}, \boldsymbol{s}_{2t}, \boldsymbol{r}_{2t}\}_{t=1}^T) q_5(\{R_t\}_{t=1}^T) q_6(\sigma_R^2). \end{aligned} \tag{16}$$

This is an iterative scheme, in which we use the parameters from, say, q_ℓ to estimate q_m and vice versa. The closed-form solutions to the VB updates are given in Appendix 8. The derivations for the sender and receiver effects are also provided, as these are not straightforward due to the fact that the derivations must be taken with respect to the distribution of $A_t^* \circ (J_n - I_n)$ rather than A_t^*, as given in Equation (10).

The VB approach is quite fast and yields good point estimates. This comes at a cost, however. VB algorithms may get stuck in local modes, and which local mode one ends up in may be highly dependent on the starting values (see, e.g., Bickel *et al.*, 2013; Salter-Townshend & Murphy, 2013, for more detailed studies using variational approaches). Additionally, by partitioning the parameters and forcing them to be independent in the approximate posterior, the posterior probability regions are typically much too concentrated. In our context, we found that a Gibbs sampler obtained similar posterior means, though wider credible intervals. The MCMC algorithm was simply too slow in practice for networks of medium to large size, however.

4 Generalizing to weighted networks

In this section, we demonstrate how to generalize our approach to weighted networks in which the dyads are not constrained to $\{0, 1\}$. We accomplish this by placing our work within the framework of a GLMM. Most researchers, statisticians or not, are familiar with GLMMs that are often the tool of choice for modeling dependent non-Gaussian data. The general framework assumes that a function of the means of the random variables are themselves correlated (typically Gaussian) random variables, thus allowing researchers to control for the correlation among the data. Specifically, for some response vector $\boldsymbol{y}$, covariate matrix X, random variables γ, and design matrix Z, we write

$$g\Big(\mathbb{E}(\boldsymbol{y})\Big) = X\boldsymbol{\beta} + Z\gamma. \tag{17}$$

(Note that the notation in Equation (17) is not linked to anything previously given, but is rather a general form for a GLMM).

Up to this point, we have assumed a probit model, as this was a natural approach to dealing with complex dependencies in binary data. This is equivalent to a GLMM using the normal inverse cumulative distribution function as the link function g.

Placing our proposed methods within the GLMM framework allows us to use other link functions such as a logit() for logistic regression, as well as allowing us to model other types of non-Gaussian data; e.g., should our network data be count, as is often the case, we may use a log link corresponding to a Poisson or Negative Binomial family of distributions. Countless texts describe these models, and in fact GLMMs are so prevalent that many fields have books or articles demonstrating how to apply GLMMs to their specific subject area (e.g., Bolker *et al.*, 2009; Gbur, 2012; Krueger & Montgomery, 2014; Bharadwaj, 2016).

We wish to maintain the covariance structures detailed in Section 2.4, and in particular that implied by Equation (15) but generalize it to other link functions and other data types. This can be done by setting

$$
\begin{aligned}
& g\Big(\mathbb{E}(\mathscr{A}_t|\mathscr{A}_{t-1},\mathscr{A}_{t-2},\ldots)\Big) \\
& \quad = (\mathrm{vec}^{-}(X_{1t}),\mathrm{vec}^{-}(X_{2t}),\ldots,\mathrm{vec}^{-}(\mathscr{G}_{1t}),\mathrm{vec}^{-}(\mathscr{G}_{2t}),\ldots)\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\theta}\end{pmatrix} + Z\gamma_t, \\
& Z = \begin{pmatrix}\mathbb{1}'_{K_s}\otimes Z_s & \mathbb{1}'_{K_r}\otimes Z_r & Z_{rec}\end{pmatrix}, \\
& \gamma_t = \begin{pmatrix}\boldsymbol{s}'_{1t} & \cdots & \boldsymbol{s}'_{K_st} & \boldsymbol{r}'_{1t} & \cdots & \boldsymbol{r}'_{K_rt} & \mathscr{R}'_t\end{pmatrix}',
\end{aligned}
\tag{18}
$$

where $\mathscr{R}_t$ contains the lower triangular elements of R_t (i.e., $\mathscr{R}_t = (R_{21t}, R_{31t},\ldots, R_{n(n-1)t})$), and where $\mathrm{vec}^{-}(M)$ for some $n\times n$ square matrix M is the standard $\mathrm{vec}(M)$, while omitting the diagonals; hence $\mathrm{vec}^{-}(M)$ will be an $n(n-1)\times 1$ vector. To construct Z_s, we may stack $I_{n,(-1,\cdot)}$, $I_{n,(-2,\cdot)}$, …, and $I_{n,(-n,\cdot)}$ to form a $n(n-1)\times n$ matrix, where $I_{n,(-i,\cdot)}$ is the $n\times n$ identity matrix with the *ith* row removed. Z_r is simply $I_n\otimes\mathbb{1}_{n-1}$. Constructing the $n(n-1)\times n(n-1)/2$ matrix Z_{rec} is perhaps the most involved, but can be accomplished by the following pseudocode:

Set all elements of Z_{rec} to 0.

for $i\in\{1,2,\ldots,n\}$ **do**
  for $j\in\{1,2,\ldots,n\}\setminus i$ **do**
    $r\leftarrow(n-1)(j-1)+i-1_{[i>j]}$
    if $i>j$ **then** $c=n(j-1)-\frac{j(j+1)}{2}+i$ **else** $c=n(i-1)-\frac{i(i+1)}{2}+j$
    $Z_{rec}[r,c]\leftarrow 1$
  end
end

By placing our methods within the GLMM framework, we provide an easy way to handle a wide range of data types as well as overdispersion.

5 Evidence of simultaneous dependence

We now begin to address determining whether or not simultaneous dependence exists. Just as with mixed models, we could check the intraclass correlation between the pairs of residuals $E_t[i,j]$ and $E_t[j,i]$ to evaluate the importance of simultaneous reciprocity. That is, estimate

$$
\frac{\sigma_R^2}{\sigma_R^2+1}. \tag{19}
$$

The issue is not so straighforward for the other types of simultaneous dependence. Consider the case where the variance of A^*_{ijt} does not depend on the actors i and

j nor the time t, the off diagonals of H_{srk} are 0 for all k, and the H_{sk}'s and H_{rk}'s have been scaled such that the diagonal entries are 1 (as is true in our example of Section 2.6). Then analogously to Equation (19), one may consider the vector

$$\mathbf{v}/(\mathbf{v}'\mathbb{1}) \quad \text{where} \quad \mathbf{v} = (\tau_{s1}, \tau_{s2}, \ldots, \tau_{sK_s}, \tau_{r1}, \ldots, \tau_{rK_r}, \sigma_R^2, 1). \tag{20}$$

Though Equation (20) appears similar to a vector of intraclass correlations, these two things are in fact not comparable. Equation (20) is only a ratio of variance components, while Equation (19) is a veritable correlation. In the context of a directed network, there are seven correlations we could consider: $Cor(A^*_{ijt}, A^*_{k\ell t})$, $Cor(A^*_{ijt}, A^*_{kit})$, $Cor(A^*_{ijt}, A^*_{kjt})$, $Cor(A^*_{ijt}, A^*_{i\ell t})$, $Cor(A^*_{ijt}, A^*_{ijt})$,$Cor(A^*_{ijt}, A^*_{j\ell t})$, and $Cor(A^*_{ijt}, A^*_{jit})$. Moreover, these seven correlations very well may differ based on which actors we are considering! Instead, we present a visualization method that may be used to assess the evidence regarding the existence and impact of simultaneous dependence.

The main idea is that we would like to evaluate how much of our posterior distributions of $\left(\{\boldsymbol{s}_{kt}\}_{k=1}^{K_s}, \{\boldsymbol{r}_{kt}\}_{k=1}^{K_r}\right)$, $t = 1, \ldots, T$, are located within some small ball around zero. If there is no simultaneous dependence, then we would expect the posterior distributions to reflect this in having most of their mass near zero. Hence, we are concerned with

$$\begin{aligned}\mathscr{P}_{\epsilon,t} &:= \int_{\mathscr{B}_\epsilon} dF\left(\{\boldsymbol{s}_{kt}\}_{k=1}^{K_s}, \{\boldsymbol{r}_{kt}\}_{k=1}^{K_r} \mid \{A_t\}_{t=1}^T\right) \\ &= \mathbb{P}\Big(\|(\boldsymbol{s}'_{1t}, \ldots, \boldsymbol{s}'_{K_s t}, \boldsymbol{r}'_{1t}, \ldots, \boldsymbol{r}'_{K_r t})\| < \epsilon \mid \{A_t\}_{t=1}^T\Big),\end{aligned} \tag{21}$$

where $\mathscr{B}_\epsilon$ represents the ball around zero of radius ϵ. This probability is very easily and accurately estimated using a Monte Carlo approximation using draws from q_4. We can then plot $\mathscr{P}_{\epsilon,t}$ vs. ϵ to obtain a visualization of the magnitude of our individual effects at each time point.

Our estimate of this high-dimensional posterior distribution, q_4, has the surprising characteristic that most of the probability mass lies within a thin shell far from the posterior mean (intuitively, this is because the volume of $\mathscr{B}_\epsilon$ grows exponentially with n). Therefore, we need some comparison for the $\mathscr{P}_{\epsilon,t}$'s. It may be helpful to compare the posterior for $\|(\boldsymbol{s}'_{1t}, \ldots, \boldsymbol{s}'_{K_s t}, \boldsymbol{r}'_{1t}, \ldots, \boldsymbol{r}'_{K_r t})\|$ with the distribution of the magnitude of a $N(\mathbf{0}, \frac{p(\sigma_R^2+1)}{(1-p)(K_s+K_r)} I_{n(K_s+K_r)})$ random variable for some $p \in (0,1)$. The distribution of this comparative random variable arises from letting the ratio of variances in Equation (20) sum to a proportion p for these simultaneous dependence terms (and letting each of the $K_s + K_r$ terms contribute equally); that is, what does the distribution of $\|(\boldsymbol{s}'_{1t}, \ldots, \boldsymbol{s}'_{K_s t}, \boldsymbol{r}'_{1t}, \ldots, \boldsymbol{r}'_{K_r t})\|$ look like if simultaneous dependence accounts for $p(100)\%$ of the variance of the A^*_{ijt}'s compared with the inherent noise? Though there well may be better comparative distributions, what we have described provides a reasonable frame of reference by which we may evaluate the strength of the evidence of simultaneous dependence as given by the posterior distribution for the sender and receiver effects. By looking at the visualization rather than just the ratio of variance components, we do not throw away the effects of the off-diagonal elements of the covariance matrices Σ_{st} and Σ_{rt} nor the entirety of Σ_{srt} when evaluating the evidence of the existence of simultaneous dependence.

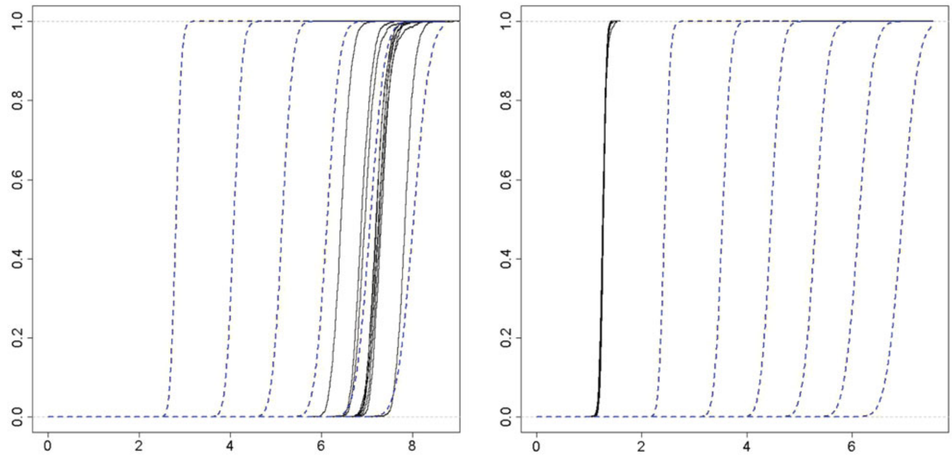


Fig. 2. Empirical example of the visualization of the existence of simultaneous dependence. The horizontal axis corresponds to the ϵ radius of a ball $\mathscr{B}_\epsilon$ about zero, and the vertical axis is $\mathscr{P}_{\epsilon,\cdot}$. Each solid line corresponds to a time point ($T = 10$), and the dotted lines correspond to the comparative random variable having proportion of variance attributable to simultaneous dependence of, from left to right, $p = 0.05, 0.1, 0.15, 0.2, 0.25, 0.3$. The left panel corresponds to data generated with simultaneous dependence and the right panel without.

The distribution of the magnitude of the comparative random variable can be evaluated in the following way. Let $\boldsymbol{x} \sim N_n(\mathbf{0}, \sigma^2 I_n)$ (e.g., $\sigma^2 = p(\sigma_R^2 + 1)/((1-p)(K_s + K_r))$). Then, let $Y^2 := \boldsymbol{x}'\boldsymbol{x}/\sigma^2 \sim \chi^2(n)$. Then $Y \sim \chi(n)$ and thus

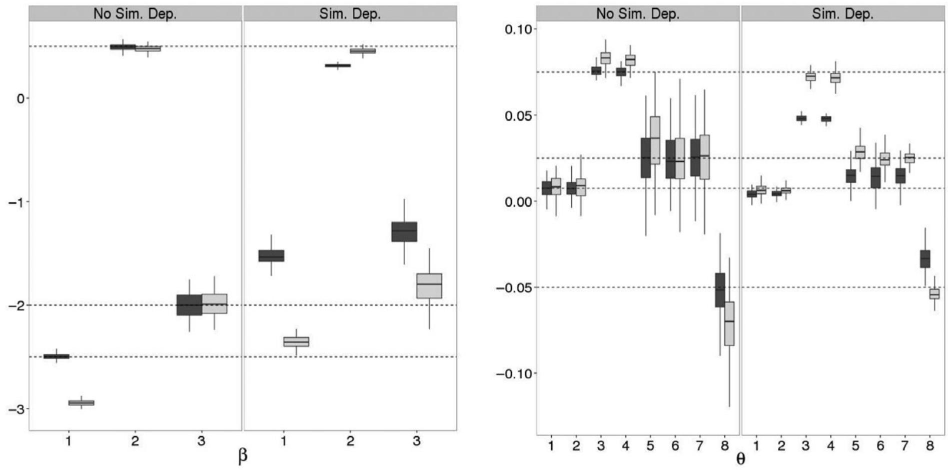
$$\mathbb{P}(\|\boldsymbol{x}\| \leqslant \epsilon) = \mathbb{P}(Y \leqslant \frac{\epsilon}{\sigma}) = \frac{\gamma(n/2, (\epsilon/\sigma)^2/2)}{\Gamma(n/2)}, \tag{22}$$

where $\gamma(\cdot,\cdot)$ is the lower incomplete gamma function. Using this we can directly compute $\mathscr{P}_\epsilon$ corresponding to this comparative random variable.

Figure 2 provides an empirical demonstration of the proposed visualization technique using the results from an arbitrarily chosen simulated data set as described in Section 6; note that we used the variance of the estimated R_t's as a proxy for σ_R^2. The left panel corresponds to data generated with simultaneous dependence and the right panel without. The solid lines correspond to the individual effects at a particular time point, and the dotted lines correspond to the comparative noise for $p \in \{0.05, 0.1, 0.15, 0.2, 0.25, 0.3\}$.

6 Simulation study

We performed a simulation study in order to investigate two things. First, what is the effect of ignoring simultaneous dependence when it exists? Second, what is the effect of modeling simultaneous dependence when it does not exist? Specifically, we wish to investigate the effects on the mean parameters, as these will typically be the parameters of interest to the researcher. To this end, we simulated 100 network data sets where there was simultaneous dependence and 100 without such dependencies.



(a) (b)

Fig. 3. Posterior means of (a) $\boldsymbol{\beta}$ and (b) $\boldsymbol{\theta}$ from analyzing the simulated data sets described in Section 6. Note that $\boldsymbol{\theta}_3$ and $\boldsymbol{\theta}_4$ have been scaled by 1/10 for visualization purposes. Horizontal dotted lines indicate true values of the parameters; the true $\boldsymbol{\beta}$ equals $(-2.5, 0.5, -2)$, and the true $\boldsymbol{\theta}$ equals $(0.0075, 0.0075, 0.75, 0.75, 0.025, 0.025, 0.025, -0.05)$. Lightly shaded boxplots correspond to accounting for simultaneous dependence in the model; dark shaded boxplots correspond to ignoring the simultaneous dependence.

For each of these 200 data sets, we fit two models, one accounting for and the other ignoring these dependencies.

Each simulated data set had $n = 100$ and $T = 10$. We incorporated two covariates as well as an intercept (i.e., $p_1 = 3$). The first dyadic covariate was a binary variable taking values 0 or 1 with equal probability; this covariate was treated as constant over time. The second covariate was constructed by first simulating n AR(1) processes with autoregressive coefficient equal to 0.9 and transition variance equal to 0.05, and then at each time point taking the distance between the corresponding cross-sectional views of the AR(1) time series. The coefficients were then set to be $\boldsymbol{\beta} = (-2.5, 0.5, -2)$ for the intercept, first covariate, and second covariate, respectively. We set $\boldsymbol{\theta} = (0.0075, 0.0075, 0.75, 0.75, 0.025, 0.025, 0.025, -0.05)$, corresponding to $\mathcal{G}_{1t}, \ldots, \mathcal{G}_{8t}$, respectively, where the $\mathcal{G}_{\ell t}$'s are as given in Section 2.6. Note that $\boldsymbol{\theta}_3$ and $\boldsymbol{\theta}_4$ needed to be on different scales, as these were the only coefficients corresponding to network structures taking values in $\{0, 1\}$ rather than $\{0, 1, \ldots, n-1\}$. For the simulations with simultaneous dependence, we set $\tau_{s2} = 0.2$, $\tau_{r2} = 0.1$, the diagonal of Ω to be $(0.25, 0.5)$, the off-diagonals of Ω equal to 0.1, and $\sigma_R^2 = 0.5$.

The results are given graphically in Figure 3. Figure 3(a) shows the boxplots of the estimates of the 3×1 vector $\boldsymbol{\beta}$. The columns correspond to the true model, and the shade of the boxplots corresponds to whether or not simultaneous dependence was accounted for. From this, we see that in the presence of simultaneous dependence, our proposed approach does a much better job at estimating the true values of $\boldsymbol{\beta}$ than when the simultaneous dependence is ignored. In the absence of simultaneous dependence, with the exception of the intercept (arguably of little importance in most research settings) our proposed approach performs very comparably to the models that ignore simultaneous dependence. We can reach the same conclusions looking at Figure 3(b), which gives the boxplots of the estimates of the 8×1 vector $\boldsymbol{\theta}$.

In summary, accounting for simultaneous dependence in the model is extremely important in obtaining more accurate estimates of the coefficients in the mean function, and doing so even in the absence of simultaneous dependence does not seem to do much harm in the estimation. If concerns persist, one may perform the visualization described previously, as seen in Figure 2, to determine whether or not to include simultaneous dependence in the final model.

7 Data analyses

We now look at two real data sets with the intent of illustrating how our approach can be implemented in practice both for directed and undirected data. In the last example, we illustrate the change in impact from simultaneous dependence as the time intervals vary from fine to coarse.

7.1 Conference proximity network

We first look at a proximity network taken from conference goers at The Last Hope Conference, collected and made available by the OpenAMD Project (OpenAMD, 2008). The 2008 conference goers had the option to wear an RFID badge, which tracked their movements throughout the conference. Thus, we are able to construct a proximity network, connecting two actors if they spent time close to one another. This type of network is quite important in, e.g., infectious disease (Vanhems *et al.*, 2013) and the study of human behavior and organization (Eagle & Pentland, 2006). Our undirected network data consisted of 1,190 actors over 29 hours (i.e., $T = 29$). We set $A_{ijt} (= A_{jit})$ to be 1 if actors i and j visited the same location during the tth hour.

Figure 4(a) shows the evidence of simultaneous dependence. From this plot, we see that there is very strong evidence of such dependencies even though the time intervals are rather fine (1 hour). Figure 4(b) shows the posterior means for the autoregressive terms when ignoring simultaneous dependence (dark gray) and when accounting for it (light gray). Notice that the estimates are, with the exception of stability, quite different; indeed, ignoring simultaneous dependence leads to a negative estimate for the effect of triangles, which seems very unlikely given previous work done on structural balance theory.

7.2 World trade data

The second data set that we consider here is that of a world trade network. We let A_{ijt} be 1 if country i exports to country j at time t. This data were collected from the Correlates of War Project (Barbieri & Keshk, 2012; Barbieri *et al.*, 2009). Along with the export/import data, we used as covariates religious makeup of a country (Maoz & Henderson, 2013), defense pacts, neutrality pacts, non-aggression pacts, and ententes (Gibler, 2009). We analyze this data in two ways. First, we focus on a larger number of countries that exist over recent years. We then look at a smaller subset of countries that all exist over a longer period of time and look at how the evidence for simultaneous dependence changes as the time intervals get coarser.

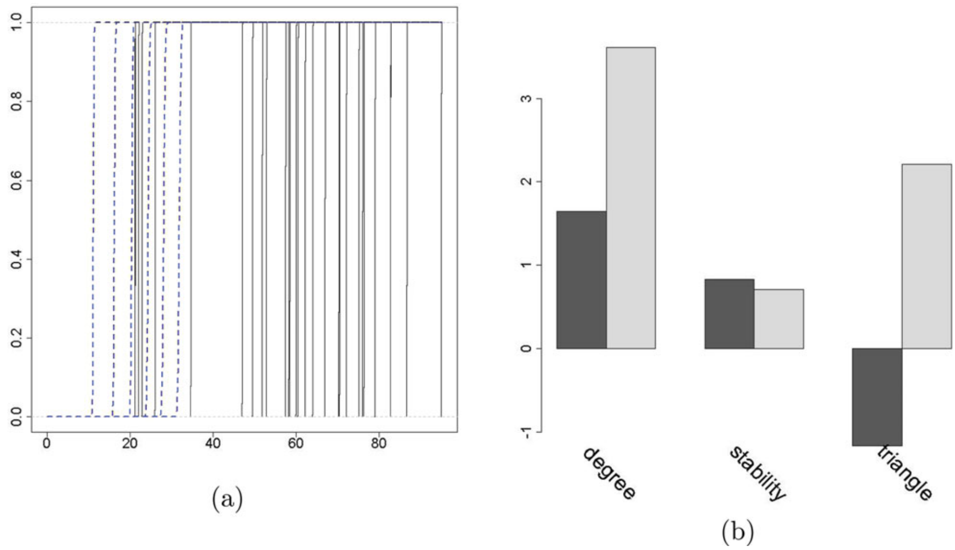


Fig. 4. Results from the AMD proximity network data. (a) Plot of $\mathscr{P}_{\epsilon,t}$ vs. ϵ. Each solid curve corresponds to the individual effects from a particular time point. The dotted lines correspond to the comparative random variable setting $p = 0.05, 0.1, 0.15, 0.2, 0.25, 0.3$. See Section 5 for details. (b) Posterior means for the coefficients of $\boldsymbol{\theta}$. Dark gray indicates ignoring simultaneous dependence, while light gray indicates accounting for this dependence in the model.

7.2.1 179 nations from 1993 to 2009

We consider all countries that exist and are involved in trade on an annual basis over the period from 1993 to 2009. For each of these countries, we have the measurements of the proportion of their population that belongs to each of the main world religions and the sub-branches of these religions (a total of 30 categories). These measurements only occur once every 5 years, which we interpolated to construct annual religious data. We then constructed the dyadic covariates by taking the Hellinger distance of two multinomial distributions whose probability vectors equal those nations' vector of proportions of religious adherents. Letting p_{it} be the 30×1 vector of the ith nation's proportion of religious adherents, this is equivalent to setting the dyadic covariate between i and j equal to $\sqrt{1 - \sum_{r=1}^{30} \sqrt{p_{itr} p_{jtr}}}$. The four types of pacts each were simply binary variables indicating whether or not countries i and j were engaged in such a pact during year t.

Figure 5(a) depicts the evidence of simultaneous dependence. From this, we see that we there is evidence of non-negligible simultaneous dependence, though much less so than in the AMD network data. Figure 5(b) shows the posterior means for the covariates and Figure 5(c) shows the same for the autoregressive terms, where again dark gray indicates ignoring simultaneous dependence and light gray indicates accounting for it in the model. As is consistent with the simulation results, when there is weaker simultaneous dependence in the data, these estimates are more in agreement. There are still some differences, mostly manifested in the attenuation of the estimates as well as more dramatic differences in the triadic effects.

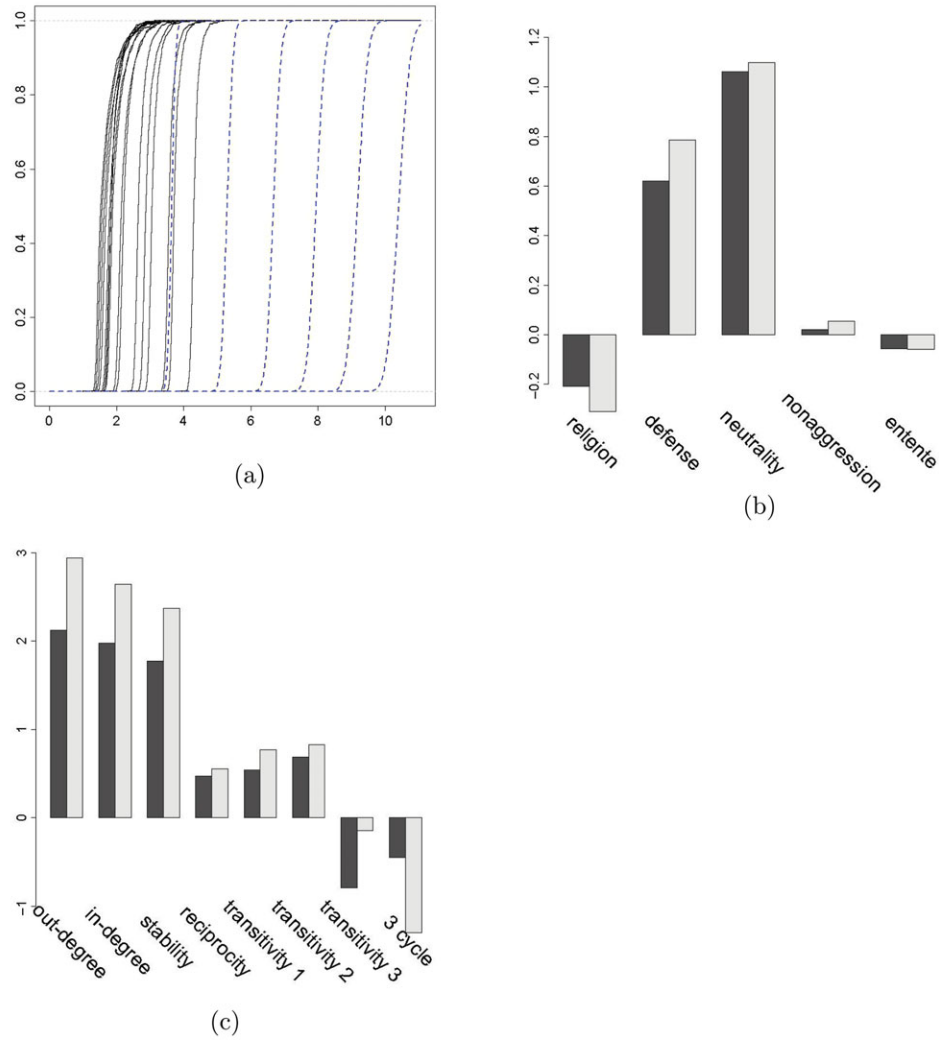


Fig. 5. Results from the world trade network data. (a) Plot of $\mathscr{P}_{\epsilon,t}$ vs. ϵ. Each curve corresponds to the random effects from a particular time point. The dotted lines correspond to the comparative random variable setting $p = 0.05, 0.1, 0.15, 0.2, 0.25, 0.3$. See Section 5 for details. (b) Posterior means of the covariates ($\boldsymbol{\beta}$). Dark gray indicates ignoring simultaneous dependence, while light gray indicates accounting for this dependence in the model. (c) Posterior means of the autoregressive terms ($\boldsymbol{\theta}$). Dark gray indicates ignoring simultaneous dependence, while light gray indicates accounting for this dependence in the model.

7.2.2 Evaluating the effect of the time interval on simultaneous dependence

As we have just seen, even at annual increments we see the presence of simultaneous dependence. We now show how this presence increases as the time intervals become coarser. We now consider the time interval from 1900 to 2000. This naturally diminishes the number of nations that exist during the entirety of the specified time

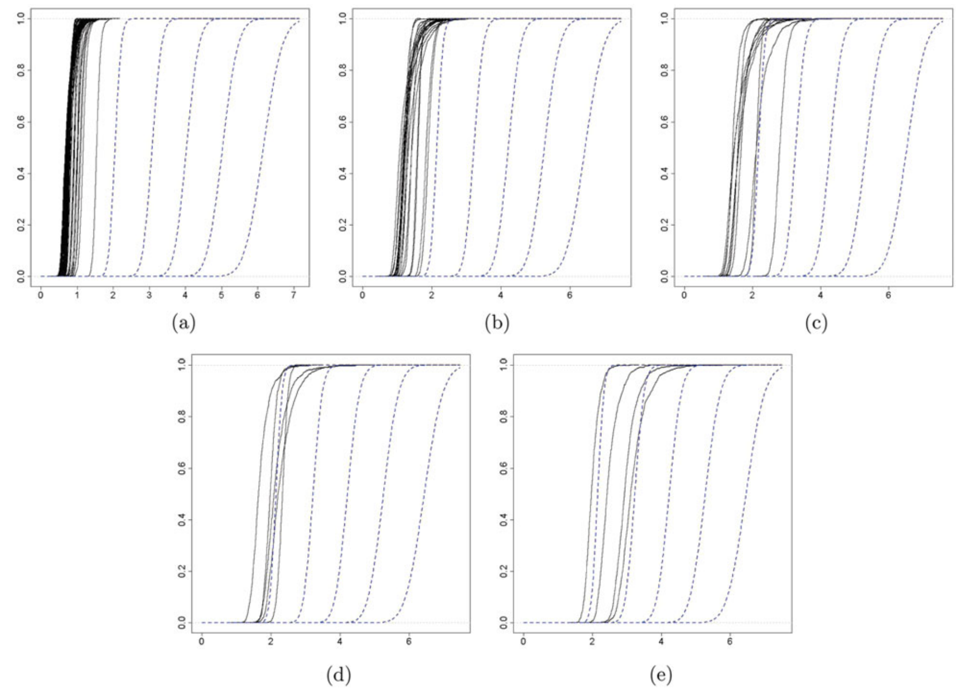


Fig. 6. World trade data: Plots of $\mathscr{P}_{\epsilon,t}$ vs. ϵ. Each curve corresponds to the random effects from a particular time point. The dotted lines in each figure correspond to the comparative random variable setting $p = 0.05, 0.1, 0.15, 0.2, 0.25, 0.3$. See Section 5 for details. Coarser time intervals lead to stronger evidence of simultaneous dependence. (a) Annual. (b) Every 5 years. (c) Every 10 years. (d) Every 20 years. (e) Every 25 years.

interval, and we are left with 28 nations. We apply our model to these 28 nations looking at every year, every 5 years, every 10 years, every 20 years, and every 25 years. Intuition (as well as previous work by Lerner *et al.*, 2013) tells us that the simultaneous dependence should grow as the time interval becomes larger, and in fact this is what we see.

Figure 6 gives the evidence of the simultaneous dependence for the five data sets. We can see that simultaneous dependence increases with the coarseness of the time interval, as shown by the increasing trend for the location of the thin shell of posterior probability mass for the individual effects. To corroborate this, we also implemented the TERGM model on the five different data sets (collected every 1, 5, 10, 20, and 25 years). To capture the simultaneous dependencies, we included as ERGM terms the counts of reciprocated ties, transitive triangles, and three-cycles. Figure 7 shows the trends of these parameter estimates for the five data sets, where the values for each parameter have been normalized by the corresponding parameter value from the 25 year interval data. We see that the strength of the effect sizes increase as the time between observations increases (we actually show the negative of the three-cycle coefficients for visual clarity), thus corroborating our finding that the simultaneous dependence does in fact increase.

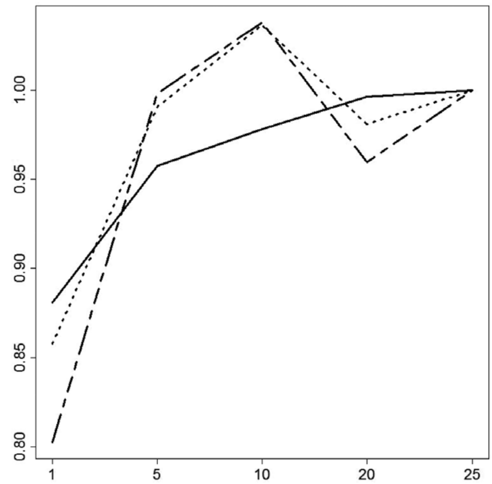


Fig. 7. TERGM coefficient estimates for reciprocity (solid), transitive triples (dotted), and cyclic triples (dash-dot) (negative coefficients given for the cyclic triples). Horizontal axis corresponds to the spacing of observations for the data set used. The increasing trend in the strength of the effect sizes corroborates our finding of increasing simultaneous dependence.

8 Discussion

In this paper, we have adapted the dynamic logistic network regression model of Almquist & Butts (2013) by introducing a framework for capturing not only temporal dependencies through an autoregressive mean structure but also simultaneous dependence through an autoregressive covariance structure. We demonstrated that ignoring simultaneous dependence leads to negative inferential consequences. The methods outlined here account for both complex temporal and simultaneous dependencies in a parsimonious way, while keeping within a familiar framework.

Like many other statistical models for network data, scalability is an issue for all but very simple simultaneous dependence structures. While the VB estimation method proposed for the STAR model is quick for small to medium data sets, the requirement to invert large covariance matrices prohibits this methodology in its current state from being scaled up to extremely large networks.

We have also described how our work may be placed within the familiar GLMM framework. While it is beyond the scope of this paper to thoroughly discuss model selection problems involving, e.g., covariance structures or link functions, it is the author's hope that previous and ongoing GLM and GLMM research (e.g., Chen & Tsurumi, 2010) can be used to build upon the proposed work in this area. Further, while we have shown practical operationalizations of the proposed method for binary data in Section 2.6, we leave it for future work to describe the specifics of sophisticated covariance structures (i.e., $H_{\cdot,t}$'s that are more complicated than I_n) for other data types.

Other future work that would be valuable to the network analysis community would be to provide a thorough comparison of the available methods for discrete temporal network data, such as the proposed approach, TERGM (Hanneke *et al.*, 2010) and STERGM (Krivitsky & Handcock, 2014), latent space models for dynamic networks (Durante & Dunson, 2014; Sewell & Chen, 2015), and dynamic stochastic blockmodels (Xing *et al.*, 2010). It would be important to know which method ought to be used in various contexts, and under what circumstances the conclusions from these models might differ.

References

Airoldi, E. M., Fienberg, S. E., & Xing, E. P. (2008). Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, **9**, 1981–2014.

Almquist, Z. W., & Butts, C. T. (2013). Dynamic network logistic regression: A logistic choice analysis of inter-and intra-group blog citation dynamics in the 2004 us presidential election. *Political Analysis*, **21**(4), 430–448.

Almquist, Z. W., & Butts, C. T. (2014). Logistic network regression for scalable analysis of networks with joint edge/vertex dynamics. *Sociological Methodology*, **44**(1), 273–321.

Anderson, T. W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *The Annals of Statistics*, **1**(1), 135–141.

Barbieri, K., & Keshk, O. (2012). Correlates of war project trade data set codebook, version 3.0. Retrieved April 29, 2015, from `http://correlatesofwar.org`.

Barbieri, K., Keshk, O. M. G., & Pollins, B. M. (2009). Trading data: Evaluating our assumptions and coding rules. *Conflict Management and Peace Science*, **26**(5), 471–491.

Bharadwaj, H. M. (2016). Generalized linear mixed models in hearing science. *The Journal of the Acoustical Society of America*, **139**(4), 2101–2101.

Bickel, P., Choi, D., Chang, X., & Zhang, H. (2013). Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *The Annals of Statistics*, **41**(4), 1922–1943.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, **24**(3), 127–135.

Chen, G., & Tsurumi, H. (2010). Probit and logit model selection. *Communications in Statistics*, **40**(1), 159–175.

Cox, D. R. (1981). Statistical analysis of time series: Some recent developments. *Scandinavian Journal of Statistics*, **8**(2), 93–115.

Demidenko, E. (2013). *Mixed models: theory and applications with R*. Hoboken, NJ: John Wiley & Sons.

Duijn, M. A. J., Snijders, T. A. B., & Zijlstra, B. J. H. (2004). p2: A random effects model with covariates for directed graphs. *Statistica Neerlandica*, **58**(2), 234–254.

Durante, D., & Dunson, D. B. (2014). Nonparametric bayes dynamic modelling of relational data. *Biometrika*, **101**(4), 125–138.

Eagle, N., & Pentland, A. (2006). Reality mining: Sensing complex social systems. *Personal and Ubiquitous Computing*, **10**(4), 255–268.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**(4), 987–1007.

Frank, O., & Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, **81**(395), 832–842.

Gbur, E. (2012). *Analysis of generalized linear mixed models in the agricultural and natural resources sciences*. Madison, WI: Soil Science Society of America.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2004). *Bayesian data analysis* (3rd ed.). Boca Raton, USA: Chapman & Hall/CRC.

Gibler, D. M. (2009). *International military alliances, 1648–2008*. Washington, DC: CQ Press.

Handcock, M. S., Raftery, A. E., & Tantrum, J. M. (2007). Model-based clustering for social networks. *Journal of the Royal Statistical Society, Series A*, **170**(2), 301–354.

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

Hoff, P. D. (2005). Bilinear mixed-effects models for dyadic data. *Journal of the American Statistical Association*, **100**(469), 286–295.

Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, **97**(460), 1090–1098.

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

Holland, P. W., & Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, **76**(373), 33–50.

Holland, P. W., Laskey, K. B., & Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks*, **5**(2), 109–137.

Hummel, R. M., Hunter, D. R., & Handcock, M. S. (2012). Improving simulation-based algorithms for fitting ERGMS. *Journal of Computational and Graphical Statistics*, **21**(4), 920–939.

Jin, I. H., & Liang, F. (2013). Fitting social network models using varying truncation stochastic approximation mcmc algorithm. *Journal of Computational and Graphical Statistics*, **22**(4), 927–952.

Krackhardt, D., & Handcock, M. S. (2007). Heider vs simmel: Emergent features in dynamic structures. In E. M. Airoldi, D. M. Blei, S. E. Fienberg, A. Goldenberg, E. P. Xing, & A. X. Zheng (Eds.), *Statistical network analysis: Models, issues, and new directions*, ICML 2006 Workshop on Statistical Network Analysis (pp. 14–27). Pittsburgh, PA: Springer.

Krivitsky, P. N., & Handcock, M. S. (2014). A separable model for dynamic networks. *Journal of the Royal Statistical Society, Series B*, **76**(1), 29–46.

Krivitsky, P. N., Handcock, M. S., Raftery, A. E., & Hoff, P. D. (2009). Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social Networks*, **31**(3), 204–213.

Krueger, D. C., & Montgomery, D. C. (2014). Modeling and analyzing semiconductor yield with generalized linear mixed models. *Applied Stochastic Models in Business & Industry*, **30**(6), 691–707.

Leenders, R. Th. A. J. (1995). Models for network dynamics: A markovian framework. *The Journal of Mathematical Sociology*, **20**(1), 1–21.

Lerner, J., Indlekofer, N., Nick, B., & Brandes, U. (2013). Conditional independence in dynamic networks. *Journal of Mathematical Psychology*, **57**(6), 275–283.

Maoz, Z., & Henderson, E. A. (2013). The world religion dataset, 1945–2010: Logic, estimates, and trends. *International Interactions*, **39**(3), 265–291.

Okabayashi, S. (2011). *Parameter estimation in social network models*. Ph.D. thesis. University of Minnesota, Minneapolis, MN.

OpenAMD. (2008). AMD hope RFID data. Retrieved September 21, 2015, from `http://networkdata.ics.uci.edu/data.php?d=amdhope`.

Raftery, A. E., Niu, X., Hoff, P. D., & Yeung, K. Y. (2012). Fast inference for the latent space network model using a case-control approximate likelihood. *Journal of Computational and Graphical Statistics*, **21**(4), 901–919.

Ripley, R., Boitmanis, K., & Snijders, T. A. B. (2013). *Rsiena: Siena – Simulation investigation for empirical network analysis*. R package version 1.1-232.

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

Salter-Townshend, M., & Murphy, T. B. (2013). Variational bayesian inference for the latent position cluster model for network data. *Computational Statistics & Data Analysis*, **57**(1), 661–671.

Sarkar, P., & Moore, A. W. (2005). Dynamic social network analysis using latent space models. *ACM SIGKDD Explorations Newsletter*, **7**(2), 31–40.

Sewell, D. K., & Chen, Y. (2015). Latent space models for dynamic networks. *Journal of the American Statistical Association*, **110**(512), 1646–1657.

Shephard, N. (1995). *Generalized linear autoregressions*. Economics Papers 8. Economics Group, Nuffield College, University of Oxford.

Simmel, G., & Wolff, K. H. (1950). *The sociology of Georg Simmel*. vol. 92892. New York, NY: Simon and Schuster.

Snijders, T. A. B. (1996). Stochastic actor-oriented models for network change. *Journal of Mathematical Sociology*, **21**(1–2), 149–172.

Snijders, T. A. B., & Nowicki, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification*, **14**(1), 75–100.

Vanhems, P, Barrat, A., Cattuto, C., Pinton, J.-F., Khanafer, N., Régis, C., . . . Voirin, N. (2013). Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PLoS One*, **8**(9), e73970.

Wang, Y. J., & Wong, G. Y. (1987). Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association*, **82**, 8–19.

Warner, R. M., Kenny, D. A., & Stoto, M. (1979). A new round robin analysis of variance for social interaction data. *Journal of Personality and Social Psychology*, **37**(10), 1742–1757.

Wasserman, S. (1980). Analyzing social networks as stochastic processes. *Journal of the American Statistical Association*, **75**(370), 280–294.

Wasserman, S., & Faust, K. (1994). *Social network analysis: methods and applications*. New York, NY: Cambridge University Press.

Westveld, A. H., & Hoff, P. D. (2011). A mixed effects model for longitudinal relational and network data, with applications to international trade and conflict. *The Annals of Applied Statistics*, **5**(2A), 843–872.

Xing, E. P., Fu, W., & Song, L. (2010). A state-space mixed membership blockmodel for dynamic network tomography. *The Annals of Applied Statistics*, **4**(2), 535–566.

Zeger, S. L., & Qaqish, B. (1988). Markov regression models for time series: A quasi-likelihood approach. *Biometrics*, **44**(4), 1019–1031.

Appendix A: Closed-form updates for VB

Before giving the closed form of the q's, let us first provide a little notation that will be used. Let $I^- = J_n - I_n$, i.e., the matrix of ones with zeros on the diagonal. Let $tr(A)$ be the trace of some square matrix A. For a matrix Σ, let $\Sigma_{(i,j)}$ denote the 2×2 submatrix obtained from the ith and jth rows and columns. Let $\mathscr{A}_t^-$ denote $\text{vec}^-(A_t^*)$. Let $trN(\boldsymbol{\mu}, \Sigma)$ be the truncated normal; we will not add any notation specifying the varying domain as this should be obvious in our context from the data, which A_{ijt}^* are restricted to the positive reals and which to the negative reals. Finally, let $\vec{X}_t$ denote the $n(n-1) \times (p_1 + p_2)$ matrix such that

$$\vec{X}_t = (\text{vec}^-(X_{1t}), \ldots, \text{vec}^-(X_{p_1 t}), \text{vec}^-(\mathscr{G}_{1t}), \ldots, \text{vec}^-(\mathscr{G}_{p_2 t})).$$

Result 1. *$q_1(\boldsymbol{\beta}, \boldsymbol{\theta}) \stackrel{\mathscr{D}}{=} N(\boldsymbol{\mu}_m, \Sigma_m)$, where*

$$\Sigma_m^{-1} = \text{diag}(1/\sigma_\beta^2, \ldots, 1/\sigma_\beta^2, 1/\sigma_\theta^2, \ldots, 1/\sigma_\theta^2) + \sum_{t=1}^{T} \vec{X}_t' \vec{X},$$

$$\boldsymbol{\mu}_m = \Sigma_m \left(\sum_{t=1}^{T} \vec{X}_t' (M_{A_t} - \text{vec}^-((\boldsymbol{\mu}_{s_1 t} + \boldsymbol{\mu}_{s_2 t})\mathbb{1}') - \text{vec}^-(\mathbb{1}(\boldsymbol{\mu}_{r_1 t} + \boldsymbol{\mu}_{r_2 t})') - \text{vec}^-(M_{R_t})) \right).$$

Result 2. $q_2(\tau_{s2}, \tau_{r2}, \Omega) \overset{\mathscr{D}}{=} IG(a_s, b_s)IG(a_r, b_r)IW(a_\Omega, B_\Omega)$ *where*

$$
\begin{aligned}
a_s &= a_{s0} + nT/2 & b_s &= b_{s0} + \tfrac{1}{2}\textstyle\sum_{t=1}^{T}\left[tr(\widetilde{\Sigma}_{srt(s)}H_{st}^{-1}) + \boldsymbol{\mu}_{s_2t}'H_{st}^{-1}\boldsymbol{\mu}_{s_2t}\right]\\
a_r &= a_{r0} + nT/2 & b_r &= b_{r0} + \tfrac{1}{2}\textstyle\sum_{t=1}^{T}\left[tr(\widetilde{\Sigma}_{srt(r)}H_{rt}^{-1}) + \boldsymbol{\mu}_{r_2t}'H_{rt}^{-1}\boldsymbol{\mu}_{r_2t}\right]\\
a_\Omega &= a_{\Omega 0} + nT & B_\Omega &= B_{\Omega 0} + \textstyle\sum_{t=1}\sum_{i=1}^{n}\left[\widetilde{\Sigma}_{srt(sr)(i,n+i)} + (\boldsymbol{\mu}_{s_1ti}, \boldsymbol{\mu}_{r_1ti})'(\boldsymbol{\mu}_{s_1ti}, \boldsymbol{\mu}_{r_1ti})\right],
\end{aligned}
$$

$\widetilde{\Sigma}_{srt(s)}$ *is the first n rows and first n columns of* $\widetilde{\Sigma}_{srt}$, $\widetilde{\Sigma}_{srt(r)}$ *is the second n rows and second n columns of* $\widetilde{\Sigma}_{srt}$, *and* $\widetilde{\Sigma}_{srt(sr)}$ *is the last* $(2n)$ *rows and* $(2n)$ *columns of* $\widetilde{\Sigma}_{srt}$.

Result 3. $q_3(\{\mathscr{A}_t^-\}_{t=1}^T) \overset{\mathscr{D}}{=} \prod_{t=1}^T trN(M_{A_t}, I)$ *where*

$$
M_{A_t} = \vec{X}_t\boldsymbol{\mu}_m + \mathrm{vec}^-\big((\boldsymbol{\mu}_{s_1t} + \boldsymbol{\mu}_{s_2t})\mathbb{1}'\big) + \mathrm{vec}^-\big(\mathbb{1}(\boldsymbol{\mu}_{r_1t} + \boldsymbol{\mu}_{r_2t})'\big) + \mathrm{vec}^-\big(M_{R_t}\big).
$$

Result 4. $q_4(\{\boldsymbol{s}_{1t}, \boldsymbol{r}_{1t}, \boldsymbol{s}_{2t}, \boldsymbol{r}_{2t}\}_{t=1}^T) \overset{\mathscr{D}}{=} \prod_{t=1}^T N\left((\boldsymbol{\mu}_{s_1t}', \boldsymbol{\mu}_{r_1t}', \boldsymbol{\mu}_{s_2t}'\boldsymbol{\mu}_{r_2t}')', \widetilde{\Sigma}_{srt}\right)$, *where*

$$
\widetilde{\Sigma}_{srt}^{-1} = \begin{pmatrix} 1 & 0 & 1 & 0\\ 0 & 1 & 0 & 1\\ 1 & 0 & 1 & 0\\ 0 & 1 & 0 & 1 \end{pmatrix} \otimes (n-1)I_n + \begin{pmatrix} 0 & 1 & 0 & 1\\ 1 & 0 & 1 & 0\\ 0 & 1 & 0 & 1\\ 1 & 0 & 1 & 0 \end{pmatrix} \otimes I^-
$$
$$
+ \left(\begin{array}{c|cc} a_\Omega B_\Omega^{-1} \otimes I_n & \multicolumn{2}{c}{0} \\ \hline 0 & \frac{a_s}{b_s}H_{s1t}^{-1} & 0 \\ & 0 & \frac{a_r}{b_r}H_{r1t}^{-1} \end{array}\right)
$$

$$
\begin{pmatrix} \boldsymbol{\mu}_{s_1t}\\ \boldsymbol{\mu}_{r_1t}\\ \boldsymbol{\mu}_{s_2t}\\ \boldsymbol{\mu}_{r_2t} \end{pmatrix} = \widetilde{\Sigma}_{srt} \begin{pmatrix} \left(\text{rev-vec}^-\left(M_{A_t} - \vec{X}_t\boldsymbol{\mu}_m\right) - M_{R_t}\right)\mathbb{1}\\ \left(\text{rev-vec}^-\left(M_{A_t} - \vec{X}_t\boldsymbol{\mu}_m\right)' - M_{R_t}\right)\mathbb{1}\\ \left(\text{rev-vec}^-\left(M_{A_t} - \vec{X}_t\boldsymbol{\mu}_m\right) - M_{R_t}\right)\mathbb{1}\\ \left(\text{rev-vec}^-\left(M_{A_t} - \vec{X}_t\boldsymbol{\mu}_m\right)' - M_{R_t}\right)\mathbb{1} \end{pmatrix}
$$

and rev-vec$^-(\cdot)$ *is the matrix (with zero diagonal elements) constructed by reversing the* vec$^-(\cdot)$ *operator.*

Derivation:

We first provide some preliminary results:

1. For some $n \times 1$ vectors $\mathbf{a}_1$ and $\mathbf{a}_2$, $tr(D_{a_1}I^-I^-D_{a_2}) = (n-1)\mathbf{a}_1'\mathbf{a}_2$, where D_a denotes a diagonal matrix whose entries are $\mathbf{a}$.
2. For some $n \times n$ matrix A, $tr(I^-D_a(A \circ I^-)) = \mathbf{a}'(A \circ I^-)\mathbb{1}$.
3. $tr(I^-D_{a_1}I^-D_{a_2}) = \mathbf{a}_1'I^-\mathbf{a}_2$.

Also note that since $\mathrm{vec}^-(A)'\,\mathrm{vec}^-(A) = \mathrm{vec}(A \circ I^-)'\mathrm{vec}(A \circ I^-) = tr((A \circ I^-)'(A \circ I^-))$, we may consider the conditional probability of $\mathscr{A}_t|\boldsymbol{s}_{1t}, \boldsymbol{r}_{1t}, \boldsymbol{s}_{2t}\boldsymbol{r}_{2t}, \cdot$ as proportional (with respect to the sender and receiver effects) to the matrix normal distribution kernel of $A_t^* \circ I^-$.

Letting $\tilde{A}_t = (A_t^* - \langle \boldsymbol{\beta}, \mathscr{X}_t \rangle + \langle \boldsymbol{\theta}, \mathscr{G}_t \rangle) \circ I^-$, we have, dropping the subscript t,

$$\begin{aligned}
&\log(\pi(A^*|\boldsymbol{s}_1, \boldsymbol{r}_1, \boldsymbol{s}_2, \boldsymbol{r}_2, \cdot)) \\
&\quad = \text{const} - \frac{1}{2} tr \left[(\tilde{A} - D_{s1}I^- - D_{s2}I^- - I^-D_{r1} - I^-D_{r2})' \right. \\
&\qquad \left. \times (\tilde{A} - D_{s1}I^- - D_{s2}I^- - I^-D_{r1} - I^-D_{r2}) \right] \\
&\quad = \text{const} - \frac{1}{2} tr \left[I^-D_{s1}D_{s1}I^- - 2I^-D_{s1}\tilde{A} + 2I^-D_{s1}D_{s2}I^- + 2I^-D_{s1}I^-D_{r1} \right. \\
&\qquad + 2I^-D_{s1}I^-D_{r2} - 2I^-D_{s2}\tilde{A} + I^-D_{s2}D_{s2}I^- + 2I^-D_{s2}I^-D_{r1} + 2I^-D_{s2}I^-D_{r2} \\
&\qquad \left. + D_{r1}I^-I^-D_{r1} + 2D_{r1}I^-I^-D_{r2} + D_{r2}I^-I^-D_{r2} - 2D_{r1}I^-\tilde{A} - 2D_{r2}I^-\tilde{A} \right] \\
&\quad = \text{const} - \frac{1}{2} \left[(n-1)\boldsymbol{s}_1'\boldsymbol{s}_1 - 2\boldsymbol{s}_1\tilde{A}\mathbb{1} + 2(n-1)\boldsymbol{s}_1'\boldsymbol{s}_2 + 2\boldsymbol{s}_1'I^-\boldsymbol{r}_1 + 2\boldsymbol{s}_1'I^-\boldsymbol{r}_2 - 2\boldsymbol{s}_2'\tilde{A}\mathbb{1} \right. \\
&\qquad + (n-1)\boldsymbol{s}_2'\boldsymbol{s}_2 + 2\boldsymbol{s}_2'I^-\boldsymbol{r}_1 + 2\boldsymbol{s}_2'I^-\boldsymbol{r}_2 + (n-1)\boldsymbol{r}_1'\boldsymbol{r}_1 + 2(n-1)\boldsymbol{r}_1'\boldsymbol{r}_2 + (n-1)\boldsymbol{r}_2'\boldsymbol{r}_2 \\
&\qquad \left. - 2\boldsymbol{r}_1'\tilde{A}'\mathbb{1} - 2\boldsymbol{r}_2'\tilde{A}'\mathbb{1} \right].
\end{aligned}$$

Combining the expected value of this under q with $\mathbb{E}_q(\log(\pi(\boldsymbol{s}_{1t}, \boldsymbol{r}_{1t}, \boldsymbol{s}_{2t}, \boldsymbol{r}_{2t}|\tau_{s2}, \tau_{r2}, \Omega, A_{t-1})))$ yields Result 4. $\square$

Result 5. *$q_5(\{R_t\}_{t=1}^T) \stackrel{\mathscr{D}}{=} \prod_t \prod_{i<j} N(M_{R_t}[i,j], \tilde{\sigma}_R^2)$ where*

$$\begin{aligned}
M_{R_t}[i,j] &= \tilde{\sigma}_R^2(\tilde{A}_{ijt} + \tilde{A}_{jit}), \\
\tilde{\sigma}_R^2 &= \frac{b_R/a_R}{1 + 2b_R/a_R}, \\
\tilde{A}_{ijt} &= \text{rev-vec}^-\left(M_{A_t} - \vec{X}_t\boldsymbol{\mu}_m\right)[i,j] - \boldsymbol{\mu}_{s_1t}[i] - \boldsymbol{\mu}_{s_2t}[i] - \boldsymbol{\mu}_{r_1t}[j] - \boldsymbol{\mu}_{r_2t}[j].
\end{aligned}$$

For the purposes of computing the parameters for the other q's, assume for $i < j$ that $M_{R_t}[j,i] = M_{R_t}[i,j]$.

Result 6. *$q_6(\sigma_R^2) \stackrel{\mathscr{D}}{=} IG(a_R, b_R)$ where*

$$\begin{aligned}
a_R &= a_{R0} + \frac{Tn(n-1)}{4} \\
b_R &= b_{R0} + \frac{1}{2}\sum_t \sum_{i<j} \left(\tilde{\sigma}_R^2 + M_{R_t}[i,j]^2\right)
\end{aligned}$$

Result 7. *For the undirected case, $q_4(\{\boldsymbol{s}_t\}_{t=1}^T) = \prod_{t=1}^T N(\boldsymbol{\mu}_{st}', \tilde{\Sigma}_{st})$, where*

$$\begin{aligned}
\boldsymbol{\mu}_{st} &= \tilde{\Sigma}_{st}\mathbb{E}(A_t^* \circ I^-)\mathbb{1} \\
\tilde{\Sigma}_{st}^{-1} &= (n-1)I_n + I^- + \frac{a_s}{b_s}H_{st}^{-1}
\end{aligned}$$

Derivation: Define $I^\triangle$ as the square matrix with ones on the upper triangle and zero everywhere else (the diagonal is also zero). As before, it is helpful to provide some preliminary results:

1. For some $n \times 1$ vector $\mathbf{a}$, $tr(D_\mathbf{a}(I^\triangle I^{\triangle\prime} + I^{\triangle\prime}I^\triangle)\mathbf{a}) = (n-1)\mathbf{a}'\mathbf{a}$.
2. For some $n \times n$ matrix A,

$$tr(D_\mathbf{a}(\tilde{A}'I^\triangle + \tilde{A}I^{\triangle\prime})) = tr(D_\mathbf{a}I^-A) = \mathbf{a}'(A \circ I^-)\mathbb{1}.$$

3. $2 \cdot tr(D_\mathbf{a}I^{\triangle\prime}D_\mathbf{a}I^\triangle) = \mathbf{a}'I^-\mathbf{a}$.

To show this last, note that the ith diagonal of $D_{\mathbf{a}} I^{\triangle\prime} D_{\mathbf{a}} I^{\triangle} = \sum_{j=1}^{i-1} \mathbf{a}_i \mathbf{a}_j$, and hence the trace equals $\sum_{i=1}^{n} \sum_{j=1}^{i-1} \mathbf{a}_i \mathbf{a}_j = \mathbf{a}' I^{\triangle\prime} \mathbf{a} = \mathbf{a}' I^{\triangle} \mathbf{a}$. This then implies that $2 \cdot tr(D_{\mathbf{a}} I^{\triangle\prime} D_{\mathbf{a}} I^{\triangle}) = \mathbf{a}' I^{\triangle\prime} \mathbf{a} + \mathbf{a}' I^{\triangle} \mathbf{a} = \mathbf{a}' I^{-} \mathbf{a}$.

Let $\tilde{A}_t = (A_t^* - \langle \boldsymbol{\beta}, \mathscr{X}_t \rangle - \langle \boldsymbol{\theta}, \mathscr{G}_t \rangle) \circ I^{\triangle}$. Then we have, dropping the subscript t,

$$
\begin{aligned}
&\log(\pi(A_t^* | \boldsymbol{s})) \\
&\quad = \text{const} - \frac{1}{2} tr\left[(\tilde{A} - D_s I^{\triangle} - I^{\triangle} D_s)' (\tilde{A} - D_s I^{\triangle} - I^{\triangle} D_s) \right] \\
&\quad = \text{const} - \frac{1}{2} tr\left[D_s (I^{\triangle} I^{\triangle\prime} + I^{\triangle\prime} I^{\triangle}) D_s + 2 D_s I^{\triangle\prime} D_s I^{\triangle} - 2 D_s (\tilde{A}' I^{\triangle} + \tilde{A} I^{\triangle\prime}) \right] \\
&\text{const} - \frac{1}{2} \left[\boldsymbol{s}' \left((n-1) I + I^{-} + \frac{1}{\tau_s} H_s^{-1} \right) \boldsymbol{s} - 2 \boldsymbol{s}' (A^* \circ I^{-}) \mathbb{1} \right].
\end{aligned}
$$

Combining the expected value of this under q with $\mathbb{E}_q(\log(\pi(\boldsymbol{s}_t | \tau_s, A_{t-1})))$ yields Result 7. □

Appendix B: Proofs

B.1 Proposition of Section 2.3

Proof. Letting $m_{ijt} = \langle \boldsymbol{\beta}, \mathscr{X}_t \rangle [i,j] + \langle \boldsymbol{\theta}, \mathscr{G}_t \rangle [i,j]$ and $V := Var(s_{it} + r_{jt})$, we have

$$
\begin{aligned}
\mathbb{P}(A_{ijt} = 1 | \boldsymbol{\beta}, \boldsymbol{\theta}) &= \mathbb{E}\Big(\mathbb{E}\big(A_{ijt} | s_{it} + r_{jt}, \boldsymbol{\beta}, \boldsymbol{\theta}\big) | \boldsymbol{\beta}, \boldsymbol{\theta} \Big) \\
&= \mathbb{E}\left(\boldsymbol{\Phi}\left(\frac{s_{it} + r_{jt} + m_{ijt}}{\sqrt{Var(E_{ijt})}} \right) \Big| \boldsymbol{\beta}, \boldsymbol{\theta} \right) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\frac{s_{it} + r_{jt} + m_{ijt}}{\sqrt{Var(E_{ijt})}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}} \frac{1}{\sqrt{2\pi V}} e^{-\frac{(s_{it} + r_{jt})^2}{2V}} dZ\, d(s_{it} + r_{jt}) \\
&= \mathbb{P}(Z \sqrt{Var(E_{ijt})} - (s_{it} + r_{jt}) < m_{ijt}).
\end{aligned}
$$

Since $Z \sqrt{Var(E_{ijt})} - (s_{it} + r_{jt}) \sim N(0, Var(E_{ijt}) + V)$, our result holds. □

B.2 Theorem of Section 2.4

Proof. It is obvious that the mean of each A_{ijt}^* are equivalent for (I), (II), and (III), and that the covariance between any A_{ijt}^* and $A_{k\ell t}^*$ as given by (III) satisfies Equation (4).

It is straightforward to check that $\sigma_R^2 M_R + (\sigma_\epsilon^2 + \sigma_R^2) I_{n^2}$ satisfies the final two terms in Equation (4), and that this is the covariance matrix of $\mathrm{vec}(E_t)$. Note that for any two n-dimensional vectors $\mathbf{a}$ and $\mathbf{b}$, we have that

1. $\mathrm{vec}(\mathbf{a}\mathbf{b}') = \mathbf{b} \otimes \mathbf{a}$,
2. $\mathrm{Cov}(\mathbb{1} \otimes \mathbf{a}) = J_n \otimes \mathrm{Cov}(\mathbf{a})$,
3. $\mathrm{Cov}(\mathbf{a} \otimes \mathbb{1}) = \mathrm{Cov}(\mathbf{a}) \otimes J_n$, and
4. $\mathrm{Cov}(\mathbb{1} \otimes \mathbf{a}, \mathbf{b} \otimes \mathbb{1}) = \mathbb{1} \otimes \mathrm{Cov}(\mathbf{a}, \mathbf{b}) \otimes \mathbb{1}'$,

where J_n is the $n \times n$ matrix of 1's. We may then write the covariance of the A^*_{ijt}'s as given in (III) as

$$\begin{aligned}
\mathrm{Cov}(\mathscr{A}_t) &= \mathrm{Cov}(\mathrm{vec}(\boldsymbol{s}_t \mathbb{1}') + \mathrm{vec}(\mathbb{1}\boldsymbol{r}_t') + \mathrm{vec}(E_t)) \\
&= \mathrm{Cov}(\mathbb{1} \otimes \boldsymbol{s}_t + \boldsymbol{r}_t \otimes \mathbb{1} + \mathrm{vec}(E_t)) \\
&= J_n \otimes \Sigma_{st} + \Sigma_{rt} \otimes J_n + \mathbb{1} \otimes \Sigma_{srt} \otimes \mathbb{1}' + \mathbb{1}' \otimes \Sigma_{srt}' \otimes \mathbb{1} + \sigma_R^2 M_R + (\sigma_\epsilon^2 + \sigma_R^2) I_{n^2}.
\end{aligned}$$

Hence (I), (II), and (III) have the same covariance structure.

Finally, we have from (III)

$$\begin{aligned}
\mathscr{A}_t &= \mathrm{vec}(\langle \boldsymbol{\beta}, \mathscr{X}_t \rangle + \langle \boldsymbol{\theta}, \mathscr{G}_t \rangle) + \mathbb{1} \otimes \boldsymbol{s}_t + \boldsymbol{r}_t \otimes \mathbb{1} + \mathrm{vec}(E_t) \\
&= \mathrm{vec}(\langle \boldsymbol{\beta}, \mathscr{X}_t \rangle + \langle \boldsymbol{\theta}, \mathscr{G}_t \rangle) + \big(\mathbb{1} \otimes I_n\big) \boldsymbol{s}_t + \big(I_n \otimes \mathbb{1}\big) \boldsymbol{r}_t + \mathrm{vec}(E_t) \\
&\overset{\mathscr{D}}{=} \mathrm{vec}(\langle \boldsymbol{\beta}, \mathscr{X}_t \rangle + \langle \boldsymbol{\theta}, \mathscr{G}_t \rangle) + \Big(\big(\mathbb{1} \otimes I_n, I_n \otimes \mathbb{1}\big) \Sigma_t^{\frac{1}{2}}, \big(\sigma_R^2 M_R + (\sigma_\epsilon^2 + \sigma_R^2) I_{n^2}\big)^{\frac{1}{2}} \Big) \mathbf{z}
\end{aligned}$$

where $\mathbf{z}$ is a $(2n + n^2) \times 1$ vector of independent standard normal random variables, and

$$\Sigma_t := \begin{pmatrix} \Sigma_{st} & \Sigma_{srt} \\ \Sigma_{srt}' & \Sigma_{rt} \end{pmatrix},$$

Since $\mathrm{vec}(\mathscr{A}_t)$ is an affine transformation of $\mathbf{z}$, we have that the A^*_{ijt}'s are jointly normal, indicating that (I), (II), and (III) are equivalent. $\square$