

Joint Modeling of Longitudinal Relational Data and Exogenous Variables

Rajarshi Guhaniyogi* and Abel Rodriguez*

Abstract. This article proposes a framework based on shared, time varying stochastic latent factor models for modeling relational data in which network and node-attributes co-evolve over time. Our proposed framework is flexible enough to handle both categorical and continuous attributes, allows us to estimate the dimension of the latent social space, and automatically yields Bayesian hypothesis tests for the association between network structure and nodal attributes. Additionally, the model is easy to compute and readily yields inference and prediction for missing link between nodes. We employ our model framework to study co-evolution of international relations between 22 countries and the country specific indicators over a period of 11 years.

Keywords: latent factor model, nodal attribute, social network, spike and slab prior, systemic dimensions.

1 Introduction

Understanding the coevolution of relational and nodal attributes is a common problem in fields as diverse as public health (Christakis and Fowler, 2007; Fowler and Christakis, 2008), finance (Kalyagin et al., 2014) and genomics (Butland et al., 2005). In these types of applications, data consists of two parts: a time series of dyadic relationships among a common set of n actors, which is encoded in the sociomatrix $\mathbf{Y}(t) = \{y_{i,j}(t) : i, j = 1, \dots, n; t \in \mathbb{N}\}$, and a collection of p time series of attributes associated with each of the actors, which is encoded by the matrix $\mathbf{Z}(t) = \{z_{i,k}(t) : i = 1, \dots, n; k = 1, \dots, p\}$.

Broadly speaking, the literature approaches the study of the association between nodal and network attributes through two different approaches. One of these approaches focuses on modeling the structure of the network conditional of the nodal attributes. The goal in that case is to understand how social relationships are formed based on the attributes of individuals, a process known as “selection”. The other approach consists of models of the nodal attributes and their association conditional on the network structure. Those models are used to understand how relationships affect attributes of the individuals in a network, a process referred to as “influence” or “contagion”.

Models of selection are typically built by regressing $y_{ij}(t)$ on node- or dyad-specific regressors using Exponential Random Graph Models (ERGMs) (Holland and Leinhardt, 1981; Robins et al., 2007) or mixed-effects generalized linear models (Wasserman and Anderson, 1987; Holland et al., 1983; Hoff et al., 2002; Hoff, 2005). ERGMs, also known

*Department of Applied Mathematics and Statistics, Baskin School of Engineering, 1156 High Street Santa Cruz, CA 95060 rguhaniy@ucsc.edu abel@soe.ucsc.edu

as p^* models, assume that the conditional distribution of the network given the covariates follows an exponential family model in which the sufficient statistics are functions of a few summary metrics of the network and nodal attributes (such as the number of edges or the number of triangles in the networks). Temporal Exponential Random Graph Models (TERGMs) are extensions of ERGMs that build dependencies between graphs over time by incorporating summary statistics that involve counts of the number links among groups of nodes at various points in time (Robins and Pattison, 2001; Hanneke and Xing, 2007; Hanneke et al., 2010). Uncertainty in TERGMs is usually assessed using resampling techniques such as the bootstrap (Desmarais and Cranmer, 2010, 2012). In a similar spirit, stochastic actor oriented models (SOAMs) (Snijders et al., 2010; Brandes et al., 2012; Snijders, 2013) have been recently proposed to model the evolution of dynamic networks. SOAMs use network-specific rules or network effects such as reciprocity, triadic closure, or homophily and identify network effects to explain the observed evolution of the networks. Though naturally appealing, the performance of ERGMs and SOAMs heavily depends on the choice of sufficient statistics or networks-specific rules that are incorporated into the formulation. ERGMs in particular can be weak at capturing local features of the network, often resulting in less than satisfactory performance in many real world problems (Snijders, 2002; Handcock et al., 2003). Alternatively, latent variable models regress the dyadic responses on covariates and/or unobserved latent variables. These models are computationally efficient and useful in modeling transitivity and homophily. On the other hand, models of contagion are usually constructed by regressing nodal attributes against those of other nodes in their network (for example, see Christakis and Fowler, 2007; Fowler and Christakis, 2008; Shoham et al., 2015 and references therein). Common methodological approaches include simultaneous autoregressive (SAR) models (e.g., see Lin, 2010) and threshold models (e.g., see Watts and Dodds, 2009).

Determining whether selection or contagion are at play (i.e., the direction of the causal relationship) is typically a difficult problem (Doreian, 2001). Instead, we focus here on jointly modeling the co-evolution of network and nodal attributes through shared latent variables. The goal of our model is twofold. First, we are interested in developing tests of association between structural features of the network and individual nodal attributes. In particular, although our approach does not allow us to distinguish between contagion and selection, it does allow us to carry out tests of independence between the network and the nodal attributes. Second, we are interested in developing predictive models that can be used to jointly predict both future links and future nodal attributes. Joint models of network and nodal attributes have started to receive increasing attention over the last few years. In a static setting, Fosdick and Hoff (2015) recently proposed an extension of the bilinear model of Hoff (2005) in which the nodal attributes and the latent factors used to explain transitivity in the network are jointly modeled using a multivariate normal distribution. In their model, testing for association between the network and nodal attributes reduces to testing whether the cross-covariance matrix between latent factors and the nodal attributes is the null matrix using a likelihood ratio test. On the other hand, Durante and Dunson (2017) proposes joint modeling of a binary/categorical response and a network using latent variable tensor factorization of the joint probability model. This framework finds its application in clustering networks

into multiple groups and thus has a different focus than ours. In a dynamic setting, [De la Haye et al. \(2010\)](#) proposed time varying joint models for network and attributes when the attributes are binary or categorical, while [Niezink et al. \(2017\)](#) extend the framework to accommodate continuous nodal attributes. These models rely on extensions of the ERGM and TERGM approaches described above, and therefore inherit the same drawbacks.

In this paper we propose a fully Bayesian approach to inference, testing and prediction for co-evolving networks and nodal attributes. The development of this model is motivated by the study of the relationship between international relationships and country-specific economic performance. Our approach is related to, but distinct from, the one presented in [Fosdick and Hoff \(2015\)](#). In addition to accommodating both discrete and continuous attributes and considering the more general case of time series data, our approach uses a common set of latent factors to explain network transitivity and covariation among attributes and network structure, and provides a fully Bayesian test of association that can be used to study individual nodal attributes. When the nodal attributes are assumed to follow conditional Gaussian distribution, our model can be interpreted as a dynamic version of the model presented in [Fosdick and Hoff \(2015\)](#), but with a structured (and potentially more parsimonious) prior on the covariance matrix between the latent traits and the nodal attributes. When the attributes are non-Gaussian, modeling through shared latent traits allows us to include them in the model in a straightforward fashion, something that the formulation in [Fosdick and Hoff \(2015\)](#) does not permit. Shared latent factors have been recently employed, among others, by [Rodríguez and Moser \(2015\)](#) to jointly model voting outcomes and abstentions in roll-call data in the U.S. Congress, and by [Durante et al. \(2017\)](#) to generate multilayer extension to the bilinear model. Our approach captures the dynamics of the system using autoregressive priors for the shared latent parameters, in an approach reminiscent to [Sarkar and Moore \(2006\)](#), [Durante and Dunson \(2014\)](#) and [Sewell and Chen \(2015\)](#). We refer to our proposed joint modeling framework as Joint Latent Factor Model (JLAFAC).

The remainder of the article flows as follows. Section 2 describes the joint modeling framework with latent factors and highlights some features of the proposed joint model. This section details our prior distribution on the model parameters and latent factors. Section 3 explains posterior computation and explains how MCMC samples are used for link and attribute prediction and assessments of association between structural features of the network and individual nodal attributes. Sections 4 and 5 demonstrate performance of the proposed framework along with the competing models in simulation studies and in a international relationships network data respectively. Finally, Section 6 concludes the article. Details of MCMC updates are described in the online supplement.

2 Model formulation

2.1 A joint model for nodal attributes and networks

Assume that a group of n actors is followed over time, and let $\mathbf{Y}(t) = [y_{i,j}(t)]$ denote the $n \times n$ binary matrix capturing dyadic interactions between these actors at time

t , and $\mathbf{Z}(t) = [z_{i,k}(t)]$ be the $n \times p$ matrix of continuous or discrete attributes for these actors at time t . Although we concentrate here in binary directed relationships between actors, the model can be easily extended to continuous and other categorical relationships, as well as to undirected relationships (please see Section 6). The network and actor attributes are observed at a finite number of time points $t_1 < t_2 < \dots < t_T$ resulting in realizations $\mathbf{Y}(t_1), \dots, \mathbf{Y}(t_T)$ of the stochastic network process $\mathbf{Y}(t)$, and $\mathbf{Z}(t_1), \dots, \mathbf{Z}(t_T)$ of the stochastic attributes $\mathbf{Z}(t)$.

We propose conditionally random effect models for $\mathbf{Y}(t)$ and $\mathbf{Z}(t)$ with shared latent factors to accommodate co-evolution. To be more precise, for the entries of the sociomatrix $\mathbf{Y}(t)$ we consider a bilinear model,

$$y_{i,j}(t) \sim \text{Ber}(\theta_{i,j}(t)), \quad \theta_{i,j}(t) = G\left(\mu(t) + \sum_{r=1}^R \lambda_r u_{i,r}(t) v_{j,r}(t)\right), \quad (2.1)$$

where $\mathbf{u}_i(t) = (u_{i,1}(t), \dots, u_{i,R}(t))'$ and $\mathbf{v}_j(t) = (v_{j,1}(t), \dots, v_{j,R}(t))'$ are both time-varying, R -dimensional latent variables, and G is an appropriate link function. In the sequel we focus our discussion on the probit link where $G(\cdot)$ corresponds to the cumulative distribution function of the standard normal distribution, but more general links such as the logistic can be easily accommodated with relatively minor changes to our computational approach.

Equation (2.1) corresponds to the bilinear model proposed in Hoff (2009). The time varying latent vectors $\mathbf{u}_1(t), \dots, \mathbf{u}_n(t)$ and $\mathbf{v}_1(t), \dots, \mathbf{v}_n(t)$ capture transitivity and reciprocity in the dyadic relationship (Hoff, 2009). The eigenvalues $\lambda_1, \dots, \lambda_R$ play a crucial role in determining the specific form of that relationship, in the sense that similar values $u_{i,r}(t)$, $u_{j,r}(t)$ and $v_{i,r}(t)$, $v_{j,r}(t)$ contribute positively or negatively to the relationships $i \rightarrow j$ and $j \rightarrow i$, depending on whether $\lambda_r > 0$ or $\lambda_r < 0$. Furthermore, note that if $\lambda_r = 0$ then the r -th dimension of the latent variables has no impact on the network structure.

The time varying attributes attached to each node can either be continuous, binary or categorical. To model them, we propose to use a set of conditionally independent generalized linear models $E\{z_{i,k}(t)\} = \delta_{i,k}(t)$ where

$$\delta_{i,k}(t) = H_k\left(\eta_k(t) + \sum_{r=1}^R \psi_{k,r} \alpha_{k,r}(t) u_{i,r}(t) + \sum_{r=1}^R \xi_{k,r} \beta_{k,r}(t) v_{i,r}(t)\right), \quad (2.2)$$

$\boldsymbol{\alpha}_k(t) = (\alpha_{k,1}(t), \dots, \alpha_{k,R}(t))'$ and $\boldsymbol{\beta}_k(t) = (\beta_{k,1}(t), \dots, \beta_{k,R}(t))'$ are $R \times 1$ vectors, and $H_k(\cdot)$ is an appropriate link function for the k th outcome of interest. For example, when $z_{i,k}(t)$ corresponds to a continuous attribute we might set

$$z_{i,k}(t) = \eta_k(t) + \sum_{r=1}^R \psi_{k,r} \alpha_{k,r}(t) u_{i,r}(t) + \sum_{r=1}^R \xi_{k,r} \beta_{k,r}(t) v_{i,r}(t) + \zeta_{i,k}(t), \quad (2.3)$$

where $\zeta_{i,k}(t)$ corresponds to realizations from a white-noise process with variance ϕ_k^{-1} .

Note that the latent factors $\mathbf{u}_1(t), \dots, \mathbf{u}_n(t)$ and $\mathbf{v}_1(t), \dots, \mathbf{v}_n(t)$ we introduced in (2.1) reappear as predictors in (2.2). The static coefficients $\psi_{k,r}$ and $\xi_{k,r}$ capture the “average” effect of the r th component of the latent factors on the corresponding nodal attribute, and the time varying coefficients $\alpha_{k,r}(t)$ and $\beta_{k,r}(t)$ control how those baseline effects vary over time. Hence, $\mathbf{u}_1(t), \dots, \mathbf{u}_n(t)$ and $\mathbf{v}_1(t), \dots, \mathbf{v}_n(t)$ not only determine the structure of the network but also control the level of association between nodal attributes, as well as the co-evolution between the network and the nodal attributes. Furthermore, just like in the case of λ_r , note that setting $\psi_{k,r} = 0$ and $\xi_{k,r} = 0$ simultaneously implies that the r th dimension of the latent factors has no effect in the evolution of $z_{i,k}(t)$.

To motivate the shared latent factor formulation discussed above it is convenient to think of the vectors \mathbf{u}_i and \mathbf{v}_i as describing the position of actor i in a latent “social” space (e.g., see Hoff et al., 2002; Hoff, 2005). From that point of view, our model makes the intuitively appealing assumption that the relative position of the nodes in social space determines both the likelihood of a link between nodes (i.e., the network structure) and the behavior of the attributes. The shared latent factor formulation can also be motivated as a generalization of the model introduced in Fosdick and Hoff (2015), which we discuss in more detail in Section 2.4. Alternatively, \mathbf{u}_i and \mathbf{v}_i can be viewed as i th node-specific scores with respect to a set of R latent features/attributes. These features regulate latent similarities among nodes (via the joint models (2.1) and (2.2)) and the evolution of the associated external covariates.

Although the previous formulation does not incorporate known covariates that might impact the evolution structure of the network or the nodal attributes, extending it in this direction is trivial by including additional regression terms in (2.1) and (2.2).

2.2 Modeling the evolution of the nodal attributes and networks

In order to borrow information across time, we propose to model the stochastic processes $\{\mu_k(t) : t \in \mathcal{T} \subseteq \mathbb{R}^+\}$, $\{u_{i,r}(t) : t \in \mathcal{T} \subseteq \mathbb{R}^+\}$, $\{v_{i,r}(t) : t \in \mathcal{T} \subseteq \mathbb{R}^+\}$, $\{\eta_k(t) : t \in \mathcal{T} \subseteq \mathbb{R}^+\}$, $\{\alpha_{k,r}(t) : t \in \mathcal{T} \subseteq \mathbb{R}^+\}$ and $\{\beta_{k,r}(t) : t \in \mathcal{T} \subseteq \mathbb{R}^+\}$ for $i, j = 1, \dots, n$, $k = 1, \dots, p$ and $r = 1, \dots, R$ using independent Gaussian process priors. Within the general class of Gaussian processes, we find it particularly convenient to work with stationary Ornstein-Uhlenbeck (OU) processes. The OU process is a Gaussian process with an exponentially damping covariance kernel over time. A key feature of the OU process is that its discretization leads to the well known first order autoregressive process, for which fast computational algorithms exist (see discussion below). Another key advantage of this specification is that this prior has “full support,” in the sense that it has positive probability of generating joint functions $\{(\Theta(t), \Delta(t)) : t \in \mathcal{T}\}$ arbitrarily close to any true joint function $\{(\Theta_0(t), \Delta_0(t)) : t \in \mathcal{T}\}$ generating the data, where $\Theta(t) = ((\theta_{i,j}(t)))_{i,j=1}^n$ and $\Delta(t) = ((\delta_{i,k}(t)))_{i,k=1}^{n,p}$. Details of theoretical results on full support, including the proofs, are given in the supplementary material Section 0.2.

For ease of presentation we assume that the times t_1, t_2, \dots, t_T at which the data is observed are equally spaced, in which case an Euler discretization of these Ornstein-Uhlenbeck processes leads to first-order autoregressive priors as described in the next

few paragraphs. This discretization can be easily modified in the case of irregularly spaced observations to provide a time-consistent model.

Let $u_{i,r}(t_l) = u_{i,r,l}$ and $v_{i,r}(t_l) = v_{i,r,l}$ for all $l = 1, \dots, T$. Assuming that t_1, t_2, \dots, t_T are equally spaced, the Ornstein-Uhlenbeck prior implies that

$$\begin{aligned} u_{i,r,l} &= \rho_u u_{i,r,l-1} + \epsilon_{i,r,l}^u, & \epsilon_{i,r,l}^u &\sim \mathbf{N}(0, \sigma_u^2), \\ v_{i,r,l} &= \rho_v v_{i,r,l-1} + \epsilon_{i,r,l}^v, & \epsilon_{i,r,l}^v &\sim \mathbf{N}(0, \sigma_v^2), \end{aligned}$$

where $|\rho_u| < 1$ and $|\rho_v| < 1$ are autocorrelation coefficients. The initial states $u_{i,r,0}$ and $v_{i,r,0}$ are assigned the respective stationary distribution, $u_{i,r,0} \sim \mathbf{N}\left(0, \frac{\sigma_u^2}{1-\rho_u^2}\right)$ and $v_{i,r,0} \sim \mathbf{N}\left(0, \frac{\sigma_v^2}{1-\rho_v^2}\right)$, which implies that the marginal prior distributions for $u_{i,r,l}$ and $v_{i,r,l}$ have zero means and the constant variances. Furthermore, setting $\rho_u = \rho_v = 0$ leads to independent priors at every point in time.

We use a similar argument for each of the other four sets of dynamic parameters in our model. More specifically, letting $\alpha_{k,r}(t_l) = \alpha_{k,r,l}$, $\beta_{k,r}(t_l) = \beta_{k,r,l}$, $\mu(t_l) = \mu_l$ and $\eta_k(t_l) = \eta_{k,l}$, we have

$$\begin{aligned} \alpha_{k,r,l} &= \rho_\alpha \alpha_{k,r,l-1} + \epsilon_{k,r,l}^\alpha, & \epsilon_{k,r,l}^\alpha &\sim \mathbf{N}(0, \sigma_\alpha^2), \\ \beta_{k,r,l} &= \rho_\beta \beta_{k,r,l-1} + \epsilon_{k,r,l}^\beta, & \epsilon_{k,r,l}^\beta &\sim \mathbf{N}(0, \sigma_\beta^2), \\ \eta_{k,l} &= \rho_\eta \eta_{k,l-1} + \epsilon_{k,l}^\eta, & \epsilon_{k,l}^\eta &\sim \mathbf{N}(0, \sigma_\eta^2), \\ \mu_l &= \rho_\mu \mu_{l-1} + \epsilon_l^\mu, & \epsilon_l^\mu &\sim \mathbf{N}(0, \sigma_\mu^2), \end{aligned}$$

where, as before, the processes are assumed to start at their respective stationary distribution, $\alpha_{k,r,0} \sim \mathbf{N}\left(0, \frac{\sigma_\alpha^2}{1-\rho_\alpha^2}\right)$, $\beta_{k,r,0} \sim \mathbf{N}\left(0, \frac{\sigma_\beta^2}{1-\rho_\beta^2}\right)$, $\eta_{k,0} \sim \mathbf{N}\left(0, \frac{\sigma_\eta^2}{1-\rho_\eta^2}\right)$ and $\mu_0 \sim \mathbf{N}\left(0, \frac{\sigma_\mu^2}{1-\rho_\mu^2}\right)$. As before, $\rho_\alpha, \rho_\beta, \rho_\eta, \rho_\mu$ are autocorrelation coefficients with $|\rho_\alpha| < 1, |\rho_\beta| < 1, |\rho_\eta| < 1$ and $|\rho_\mu| < 1$ to ensure stationarity.

To help address identifiability issues, the parameters $\sigma_u^2, \sigma_v^2, \sigma_\alpha^2$ and σ_β^2 all are fixed at 1 (please see Section 3.1). Additionally, σ_μ^2 and σ_η^2 are also kept fixed at 1 without loss of generality. The rest of the variance parameters $\phi_k^{-1}, k = 1, \dots, p$ follow proper inverse gamma distributions with infinite means, $\text{IGam}(1, 2)$. Finally, since lag parameters $\rho_u, \rho_v, \rho_\alpha, \rho_\beta, \rho_\eta, \rho_\mu$ lie in $[-1, 1]$ interval, we assign the standard normal prior distributions on transformed $\rho_u, \rho_v, \rho_\alpha, \rho_\beta, \rho_\eta, \rho_\mu$'s with full range in \mathbb{R} . This choice of priors for ρ_u, \dots, ρ_μ is meant to facilitate the use of elliptical slice sampling algorithms Nishihara et al. (2014) in order to improve the mixing of the algorithm (see Section 3).

2.3 Dimension selection and association testing

As mentioned in Section 3.3, the coefficients $\lambda_1, \dots, \lambda_R$ can potentially be used to determine which components of the latent factor vector affect the structure of the network. To exploit this property of the model we propose to use independent Gaussian mixture

priors,

$$\lambda_r \mid \gamma_r \sim \begin{cases} \mathbf{N}\left(0, \frac{1}{\tau_r}\right) & \gamma_r = 1 \\ \mathbf{N}\left(0, \frac{1}{v_0 \tau_r}\right) & \gamma_r = 0 \end{cases} \quad (2.4)$$

where $\gamma_r \mid \pi_\lambda \sim \text{Ber}(\pi_\lambda)$, $\pi_\lambda \sim \text{Beta}(a_\lambda, b_\lambda)$, and $\tau_r \sim \text{Gam}(q^{3(r-1)}, q^{2(r-1)})$, $q > 1$. This prior specification on τ_r ensures fast decay of the variance parameter $\frac{1}{\tau_r}$, so as to safeguard the implied prior on the $\theta_{i,j}(t)$ s from degenerating as R becomes large. In particular, $\sum_{r=1}^R E\{1/\tau_r\} = \sum_{r=1}^R q^{2(r-1)} / \{q^{3(r-1)} - 1\}$, so that $\sum_{r=1}^R E\{1/\tau_r\}$ converges as $R \rightarrow \infty$, which implies that the prior variance of $\theta_{i,j}$ also converges as $R \rightarrow \infty$. In our experiments we work with $q = 1.5$. On the other hand, allowing the prior inclusion probability π_λ to be random allows us to automatically adjust for multiple comparisons (Scott et al., 2010).

Note that, as $v_0 \rightarrow \infty$, the mixture component associated with $\gamma_r = 0$ converges to a Dirac delta function at 0, δ_0 . Though using a degenerate measure at zero as one component of mixture distribution is suitable for selecting unimportant dimensions, we instead use a mixture of two normal distributions with large value of v_0 in (2.4) to mimic the effect of δ_0 . This is a very common practice in the variable selection literature to enhance better mixing of the parameters, see (George and McCulloch, 1993; Ishwaran et al., 2005). Sensitivity of the procedure to the choice of v_0 is explored and discussion is offered in Section 4.

We use a similar approach to identify components of the latent factors that do not affect the different nodal attributes. In particular, we let

$$\psi_{k,r} \mid \omega_{k,r} \sim \begin{cases} \mathbf{N}\left(0, \frac{1}{\nu_{k,r}}\right) & \omega_{k,r} = 1 \\ \mathbf{N}\left(0, \frac{1}{v_0 \nu_{k,r}}\right) & \omega_{k,r} = 0 \end{cases}, \quad \xi_{k,r} \mid \varsigma_{k,r} \sim \begin{cases} \mathbf{N}\left(0, \frac{1}{\kappa_{k,r}}\right) & \varsigma_{k,r} = 1 \\ \mathbf{N}\left(0, \frac{1}{v_0 \kappa_{k,r}}\right) & \varsigma_{k,r} = 0 \end{cases}, \quad (2.5)$$

where

$$\begin{aligned} \omega_{k,r} \mid \pi_{\psi,k} &\sim \text{Ber}(\pi_{\psi,k}), & \pi_{\psi,k} &\sim \text{Beta}(a_\psi, b_\psi), & \nu_{k,r} &\sim \text{Gam}(q^{3(r-1)}, q^{2(r-1)}), \\ \varsigma_{k,r} \mid \pi_{\xi,k} &\sim \text{Ber}(\pi_{\xi,k}), & \pi_{\xi,k} &\sim \text{Beta}(a_\xi, b_\xi), & \kappa_{k,r} &\sim \text{Gam}(q^{3(r-1)}, q^{2(r-1)}). \end{aligned}$$

The indicator variables $\gamma_1, \dots, \gamma_R$, $\omega_{1,1}, \dots, \omega_{p,R}$ and $\varsigma_{1,1}, \dots, \varsigma_{p,R}$ can be used to investigate the pattern of association between the network structure and the nodal attributes, as well as to estimate the effective dimension of the latent space (see Section 3.3 for more details, as well as for a discussion on the elicitation of the parameters a_λ , b_λ , a_ψ , b_ψ , a_ξ and b_ξ).

2.4 Relationship with other models in the literature

Fosdick and Hoff (2015), which focuses in situations in which the network and nodal

attributes are observed only once, builds a joint model through a prior of the form

$$\begin{pmatrix} \mathbf{u}_i \\ \mathbf{v}_i \\ \mathbf{z}_i \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \boldsymbol{\eta} \end{pmatrix}, \begin{pmatrix} \mathbf{I} & \mathbf{0} & \boldsymbol{\Omega}_{u,z} \\ \mathbf{0} & \mathbf{I} & \boldsymbol{\Omega}_{v,z} \\ \boldsymbol{\Omega}'_{u,z} & \boldsymbol{\Omega}'_{v,z} & \boldsymbol{\Omega}_{z,z} \end{pmatrix} \right),$$

i.e., they assume independence between the latent factors but allow for general cross-covariance matrices $\boldsymbol{\Omega}_{u,z}$, $\boldsymbol{\Omega}_{v,z}$ between latent factors and attributes, as well as for an unrestricted covariance structure $\boldsymbol{\Omega}_{z,z}$ for the nodal attributes. The authors are then careful to enforce constraints on $\boldsymbol{\Omega}_{u,z}$, $\boldsymbol{\Omega}_{v,z}$ and $\boldsymbol{\Omega}_{z,z}$ so that the full covariance matrix is well defined. In this formulation, the nodal attributes and the network structure are independent if and only if both $\boldsymbol{\Omega}_{u,z} = \mathbf{0}$ and $\boldsymbol{\Omega}_{v,z} = \mathbf{0}$, which the authors check using a likelihood ratio test.

Contrast this structure with that induced by our model in the simplified case in which the data is static (i.e., $T = 1$) and the attributes are conditionally Gaussian. For our model, the joint distribution for $(\mathbf{u}'_i, \mathbf{v}'_i, \mathbf{z}'_i)'$ is also normal and the covariance structure has a very similar structure. In particular, the latent factors are all independent, and we have $\text{cov}(u_{i,r}, z_{i,k}) = \psi_{k,r} \alpha_{k,r} \frac{\sigma_u^2}{1 - \rho_u^2}$ and $\text{cov}(v_{i,r}, z_{i,k}) = \xi_{k,r} \beta_{k,r} \frac{\sigma_v^2}{1 - \rho_v^2}$. Hence, our model allows for cross-covariance matrices $\boldsymbol{\Omega}_{u,z}$ and $\boldsymbol{\Omega}_{v,z}$ that are as general as those induced by the model discussed in [Fosdick and Hoff \(2015\)](#). However, in our case

$$\boldsymbol{\Omega}_{zz} = \frac{\sigma_u^2}{1 - \rho_u^2} \mathbf{A} \mathbf{A}' + \frac{\sigma_v^2}{1 - \rho_v^2} \mathbf{B} \mathbf{B}' + \boldsymbol{\Phi}^{-1},$$

where $[\mathbf{A}]_{k,r} = \psi_{k,r} \alpha_{k,r}$, $[\mathbf{B}]_{k,r} = \xi_{k,r} \beta_{k,r}$ and $\boldsymbol{\Phi}^{-1}$ is a diagonal matrix with entries ϕ_1, \dots, ϕ_p (recall [Section 3.3](#)). Therefore, when $R \ll p$, our model induces a sparse, “low rank + diagonal” decomposition for the marginal covariance matrix for the attributes $\boldsymbol{\Omega}_{zz}$, while for $R \geq p$ the model allows for a general (unrestricted) covariance matrix.

This discussion shows that the model discussed in [Fosdick and Hoff \(2015\)](#) can be recovered as a special case of our shared latent factor model, and also makes the advantages of the more general specification clear. First, if the attributes were being modeled independently, it would be natural to assume some sort of sparse structure for the matrix $\boldsymbol{\Omega}_{zz}$. Our latent factor formulation yields that for free. Secondly, the formulation automatically leads to a well-defined joint covariance matrix without any need for a careful prior specification for the cross-covariance matrices. Thirdly, the latent factor formulation can easily accommodate categorical attributes with minimal modifications. Finally, the latent factor formulation allows for an easy extension to the time-varying case that is the focus of this paper.

3 Posterior Inference

Under a Bayesian framework, parameter estimation can be achieved via Markov chain Monte Carlo (MCMC) algorithms, in which posterior distributions for the unknown quantities are approximated with empirical distributions of samples from a Markov chain. To streamline computation, we follow [Albert and Chib \(1993\)](#) and introduce

latent variables $w_{i,j,l}$ for the network data such that $w_{i,j,l} > 0$ if $y_{i,j}(t_l) = 1$ and $w_{i,j,l} < 0$ otherwise. Equation (2.1) can now be written in terms of $w_{i,j,l}$ as

$$w_{i,j,l} = \mu_l + \sum_{r=1}^R \lambda_r u_{i,r,l} v_{j,r,l} + \epsilon_{i,j,l}, \quad \epsilon_{i,j,l} \sim N(0, 1). \quad (3.1)$$

For binary or ordinal nodal attributes we follow a similar latent variable augmentation, while for attributes that follow Gaussian distributions no data augmentation is required. This data augmentations lead to updates that mostly use Gibbs sampling steps (see supplementary material Section 0.1 for details). One exception is the correlation coefficients ρ_u, \dots, ρ_μ , for which we employ elliptical slice samplers Nishihara et al. (2014) in order to improve the mixing of the MCMC algorithm, particularly for large values of p and T (see Section 4 for an illustrations).

3.1 Identifiability of Parameters

One important issue related to the proposed model is the identifiability of the unknown parameters $\mathbf{u}_{i,l} = (u_{i,1,l}, \dots, u_{i,R,l})'$, $\mathbf{v}_{j,l} = (v_{j,1,l}, \dots, v_{j,R,l})'$, $\boldsymbol{\alpha}_{k,l} = (\alpha_{k,1,l}, \dots, \alpha_{k,R,l})'$, $\boldsymbol{\beta}_{k,l} = (\beta_{k,1,l}, \dots, \beta_{k,R,l})'$, $\lambda_1, \dots, \lambda_R$, $\psi_{k,1}, \dots, \psi_{k,R}$ and $\xi_{k,1}, \dots, \xi_{k,R}$. Note that the parameters in the proposed model could have potentially faced following identifiability issue.

Scale indeterminacy: For any constant c , $\sum_{r=1}^R \lambda_r u_{i,r,l} v_{j,r,l} = \sum_{r=1}^R \frac{\lambda_r}{c^2} (c u_{i,r,l})(c v_{j,r,l})$, $\sum_{r=1}^R \psi_{k,r} \alpha_{k,r,l} u_{i,r,l} = \sum_{r=1}^R \frac{\psi_{k,r}}{c^2} (c \alpha_{k,r,l})(c u_{i,r,l})$, and $\sum_{r=1}^R \xi_{k,r} \beta_{k,r,l} v_{i,r,l} = \sum_{r=1}^R \frac{\xi_{k,r}}{c^2} (c \beta_{k,r,l})(c v_{i,r,l})$.

Orthogonal indeterminacy: If all $\lambda_r = \xi_{k,r} = \psi_{k,r}$, then for any orthogonal matrix \mathbf{A} , $\mathbf{u}'_{i,l} \mathbf{v}_{j,l} = (\mathbf{A} \mathbf{u}_{i,l})' (\mathbf{A} \mathbf{v}_{j,l})$, $\mathbf{u}'_{i,l} \boldsymbol{\alpha}_{k,l} = (\mathbf{A} \mathbf{u}_{i,l})' (\mathbf{A} \boldsymbol{\alpha}_{k,l})$ and $\mathbf{v}'_{i,l} \boldsymbol{\beta}_{k,l} = (\mathbf{A} \mathbf{v}_{i,l})' (\mathbf{A} \boldsymbol{\beta}_{k,l})$.

Permutation indeterminacy: Again, if all $\lambda_1 = \dots = \lambda_R$, then for any permutation P of $\{1, \dots, R\}$, $\sum_{r=1}^R \lambda_r u_{i,r,l} v_{j,r,l} = \sum_{r=1}^R \lambda_r u_{i,P(r),l} v_{j,P(r),l}$. Similar identifiability issue can be faced for $\sum_{r=1}^R \psi_{k,r} \alpha_{k,r,l} u_{i,r,l}$ and $\sum_{r=1}^R \xi_{k,r} \beta_{k,r,l} v_{i,r,l}$ if all $\psi_{k,r}$'s and all $\xi_{k,r}$'s are equal respectively.

Note that λ_r 's, $\xi_{k,r}$'s and $\psi_{k,r}$'s are assigned normal prior distributions which imposes increasing shrinkage towards zero as R increases, so that both orthogonal and permutation indeterminacies are ameliorated. Moreover, $\sigma_u^2 = \sigma_v^2 = \sigma_\alpha^2 = \sigma_\beta^2 = 1$ partially solves the scale indeterminacy issue, namely λ_r 's, $\xi_{k,r}$'s and $\psi_{k,r}$'s are identifiable up to sign.

3.2 Link and nodal attribute prediction

We can approximate the posterior probability of a directed dyad from node k_1 to node k_2 at time t_l as an average of M post burn-in, suitably thinned, MCMC samples as

$$P(y_{k_1, k_2}(t_l) = 1 \mid \text{data}) \approx \frac{1}{M} \sum_{s=1}^M G \left(\mu_l^{(s)} + \sum_{r=1}^R \lambda_r^{(s)} u_{k_1, r, l}^{(s)} v_{k_2, r, l}^{(s)} \right),$$

where the superscript (s) denotes the s -th post burn-in MCMC sample for a parameter after suitable thinning. Note that this estimated link probability can be used to infer missing links within observed networks (under the additional assumption of ignorable missingness), or to predict the structure of the network at a future time. To decide whether a directed dyad between nodes k_1 and k_2 is present, one can choose a cut-off $c \in (0, 1)$ so that $\theta_{k_1, k_2}(t_l) > c$ implies a directed link. By varying c we can construct a receiver characteristic curve for our network prediction algorithm.

A similar approach can be used to predict the value of nodal attributes. For example, for the purpose of point prediction,

$$E\{z_{i, k}(t_l) \mid \text{data}\} \approx \frac{1}{M} \sum_{s=1}^M H_k \left(\eta_{k, l}^{(s)} + \sum_{r=1}^R \psi_{k, r}^{(s)} \alpha_{k, r, l}^{(s)} u_{i, r, l}^{(s)} + \sum_{r=1}^R \xi_{k, r}^{(s)} \beta_{k, r, l}^{(s)} v_{i, r, l}^{(s)} \right).$$

3.3 Patterns of association between network and nodal attributes

The latent space provides us understanding of the association between network and nodal attributes. The R dimensions of the latent space can be divided into four groups: *systemic* dimensions, which influence both the network structure and at least one of the nodal attributes, two groups of *idiosyncratic* dimensions, one group that is associated only with the network structure, and a second group that solely impacts the association between nodal attributes, and a set of inactive dimensions that have no effect on either of the outcomes of interest. More specifically, a given dimension r is systemic if and only if it significantly affects the structure of the network (i.e., $\gamma_r = 1$) and it also affects at least one of the nodal attributes (i.e., if there is at least one k for which $\omega_{k, r} = 1$ or $\varsigma_{k, r} = 1$, or both). Hence, q_S , the number of systemic dimensions in the latent space is given by

$$q_S = \sum_{r=1}^R \gamma_r \left\{ 1 - \prod_{k=1}^p (1 - \omega_{k, r}) (1 - \varsigma_{k, r}) \right\}.$$

It is argued in the supplementary material Section 0.3 that q_S is an identifiable quantity. Furthermore, $q_S = 0$ if and only if the sequence of networks and the time series of nodal attributes are mutually independent. Hence,

$$P(q_S = 0 \mid \text{data}) \approx \frac{1}{M} \sum_{s=1}^M I(q_S^{(s)} = 0),$$

where $I(\cdot)$ denotes the indicator function. $P(q_S = 0 \mid \text{data})$ provides us with a mechanism to evaluate global association between the network structure and nodal attributes, with high values of this probability indicating that these are independent. This is a Bayesian alternative to the likelihood ratio test discussed in [Fosdick and Hoff \(2015\)](#), one that also takes into account the dynamic nature of our data.

The aforesaid procedure to understand general association between network and nodal attributes can be slightly modified to investigate whether the network structure is associated with a particular nodal attribute. For this purpose, define

$$q_{S,k} = \sum_{r=1}^R \gamma_r \{1 - (1 - \omega_{k,r})(1 - \varsigma_{k,r})\}.$$

Again, an estimate of $P(q_{S,k} = 0 \mid \text{data})$ can be obtained from the samples generated by the MCMC algorithm. High values for this probability indicate that the network structure and the k th nodal attribute are (marginally) independent from each other.

A similar approach can be used to define the number of idiosyncratic dimensions associated only with the network structure,

$$q_N = \sum_{r=1}^R \gamma_r \prod_{k=1}^p (1 - \omega_{k,r})(1 - \varsigma_{k,r}),$$

and the number of idiosyncratic dimensions associated with the nodal attributes,

$$q_A = \sum_{r=1}^R (1 - \gamma_r) \left\{ 1 - \prod_{k=1}^p (1 - \omega_{k,r})(1 - \varsigma_{k,r}) \right\}.$$

Note that if $q_A = 0$, then all correlations among the nodal attributes is explained by some of the same factors that explain the network structure. Furthermore, although the model allows the dimension of the latent space to be as high as R , the effective dimensionality of the space is $R^* = q_S + q_N + q_A \leq R$, with the number of inactive dimensions being $q_O = R - q_S - q_N - q_A$. Supplementary material Section 0.3 show all of q_N , q_A and q_S and R^* are identifiable. Hence, a posteriori our model is able to learn the dimension of the latent space.

Hyperparameter selection

The prior distribution on the summaries q_S , q_N and q_A depends critically on the hyperparameters a_λ , b_λ , a_ψ , b_ψ , a_ξ and b_ξ , which must be carefully chosen. For example, a priori, $q_N \mid \Upsilon_N \sim \text{Bin}(R, \Upsilon_N)$, where $\Upsilon_N = \pi_\lambda \prod_{k=1}^p (1 - \pi_{\psi,k})(1 - \pi_{\varsigma,k})$ is a random variable with expectation,

$$\text{E}\{\Upsilon_N\} = \frac{a_\lambda}{a_\lambda + b_\lambda} \left(\frac{a_\psi}{a_\psi + b_\psi} \right)^p \left(\frac{a_\xi}{a_\xi + b_\xi} \right)^p.$$

For even moderate p , the usual choice of $a_\psi = b_\psi = a_\xi = b_\xi = 1$ (i.e., using uniform distributions for the inclusion probabilities) will lead to a prior distribution on q_N that is

heavily skewed towards 0. Such prior distribution would typically be unappealing since it would make it difficult to identify components that are idiosyncratic to the network, and therefore lead to an overestimation of the number of systemic components.

To address this issue, in our own data analysis we set $a_\lambda = b_\lambda = 1$, which ensures that the $\sum_{r=1}^R \gamma_r$ follows a uniform distribution on $\{0, \dots, R\}$, and then set the values of a_ψ , b_ψ , a_ξ and b_ξ so that $a_\psi = a_\xi = b_\psi = b_\xi$ and the marginal distributions for q_S , q_N and q_A are (approximately) identical. We evaluate the sensitivity to various choices of a_ψ , b_ψ , a_λ , b_λ in Section 4.4.

4 Simulation study

We evaluate the performance of our model using five simulated datasets, each generated under a different scenario. The purpose of this simulation study is fourfold: (a) to assess predictive performance of JLAFAC in terms of link and nodal attribute predictions, (b) to investigate the ability of our model to identify dependence and independence relationships in the data, (c) to assess the impact of hyperparameters (and, in particular, of the constant v_0 , a_ψ and a_λ used in our variable selection prior) on model performance, and (d) to evaluate the model performance under a misspecified setting. For each of datasets 1, 2 and 5 we fitted JLAFAC using $R = 5$ latent dimensions, JLAFAC is fitted with $R = 3$ for dataset 3, while for dataset 4 JLAFAC was fitted with $R = 10$ latent dimensions. All posterior inferences are based on 4,000 samples from the MCMC iterations obtained after a burn-in period of 10,000 iterations and thinning the chain every 10 samples. Convergence is assessed by monitoring the behavior of the likelihood function and the posterior distributions of q_N , q_S , q_A and R^* . On the other hand, overall mixing, monitor the effective sample size out of the 4,000 thinned post burn-in samples for all parameters. Figure 1 presents boxplots of effective sample sizes (ESS) for $\psi_{k,r}$, $\xi_{k,r}$'s over $k = 1, \dots, p$; $r = 1, \dots, R$. Since in each simulation there are only a few λ_r 's, $r = 1, \dots, R$, boxplots for effective sample sizes of λ_r 's are not presented. In fact, the average ESS for λ_r 's for scenario 1-5 are given by 1851, 1650, 2915, 1436, 3203 respectively. These results suggests a good performance of the algorithm in terms of mixing.

In our first simulation scenario the true dimension of the latent space is $R^* = 5$, the number of attributes is $p = 5$, sample size $n = 40$ and number of time points $T = 10$. The true vector of eigenvalues takes the form $\boldsymbol{\lambda} = (0, 0, 0, \lambda_4, \lambda_5)$, while the matrices of selection coefficients take the form

$$\boldsymbol{\Psi} = \begin{pmatrix} \psi_{1,1} & \psi_{1,2} & 0 & 0 & 0 \\ \psi_{2,1} & \psi_{2,2} & 0 & 0 & 0 \\ 0 & 0 & \psi_{3,3} & \psi_{3,4} & 0 \\ 0 & 0 & \psi_{4,3} & \psi_{4,4} & 0 \\ \psi_{5,1} & \psi_{5,2} & \psi_{5,3} & \psi_{5,4} & \psi_{5,5} \end{pmatrix}, \quad \boldsymbol{\Xi} = \begin{pmatrix} \xi_{1,1} & \xi_{1,2} & 0 & 0 & 0 \\ \xi_{2,1} & \xi_{2,2} & 0 & 0 & 0 \\ 0 & 0 & \xi_{3,3} & \xi_{3,4} & 0 \\ 0 & 0 & \xi_{4,3} & \xi_{4,4} & 0 \\ \xi_{5,1} & \xi_{5,2} & \xi_{5,3} & \xi_{5,4} & \xi_{5,5} \end{pmatrix},$$

Note that, in this case, we have $q_S = 2$, $q_N = 0$ and $q_A = 3$. For our second scenario we again set the true dimension of the latent space to $R^* = 5$, the number of attributes is

$p = 5$, $n = 40$ and $T = 10$. However, in this case, the true vector of eigenvalues takes the form $\boldsymbol{\lambda} = (0, 0, 0, 0, \lambda_5)$ and the selection matrices are of the form

$$\boldsymbol{\Psi} = \begin{pmatrix} \psi_{1,1} & \psi_{1,2} & 0 & 0 & 0 \\ \psi_{2,1} & \psi_{2,2} & 0 & 0 & 0 \\ 0 & 0 & \psi_{3,3} & \psi_{3,4} & 0 \\ 0 & \psi_{4,2} & \psi_{4,3} & 0 & 0 \\ \psi_{5,1} & \psi_{5,2} & \psi_{5,3} & \psi_{5,4} & 0 \end{pmatrix}, \quad \boldsymbol{\Xi} = \begin{pmatrix} \xi_{1,1} & \xi_{1,2} & 0 & 0 & 0 \\ \xi_{2,1} & \xi_{2,2} & 0 & 0 & 0 \\ 0 & 0 & \xi_{3,3} & \xi_{3,4} & 0 \\ 0 & 0 & \xi_{4,3} & \xi_{4,4} & 0 \\ \xi_{5,1} & \xi_{5,2} & \xi_{5,3} & \xi_{5,4} & 0 \end{pmatrix},$$

In this case $q_S = 0$, $q_N = 1$ and $q_A = 4$, i.e., the network and the nodal attributes are independent in this dataset. Note that the second simulation study allows structural mismatch in terms of positioning of nonzero entries between $\boldsymbol{\Psi}$ and $\boldsymbol{\Xi}$. We set up the third simulation study to allow for a large mismatch between the fitted dimension $R = 10$ and the true dimension of latent variables. Specifically, we set the true dimension of the latent space to $R^* = 3$, the number of attributes to $p = 5$, $n = 40$ and $T = 10$. In this case the true vector of eigenvalues is $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$ and we set

$$\boldsymbol{\Psi} = \begin{pmatrix} 0 & \psi_{1,2} & \psi_{1,3} \\ 0 & \psi_{2,2} & \psi_{2,3} \\ 0 & \psi_{3,2} & \psi_{3,3} \\ 0 & \psi_{4,2} & \psi_{4,3} \\ 0 & \psi_{5,2} & \psi_{5,3} \end{pmatrix}, \quad \boldsymbol{\Xi} = \begin{pmatrix} \xi_{1,1} & \xi_{1,2} & 0 \\ \xi_{2,1} & \xi_{2,2} & 0 \\ \xi_{3,1} & \xi_{3,2} & 0 \\ \xi_{4,1} & \xi_{4,2} & 0 \\ \xi_{5,1} & \xi_{5,2} & 0 \end{pmatrix},$$

so that $q_S = 3$, $q_N = 0$ and $q_A = 0$.

The fourth simulation evaluates the performance of the JLAFAC model when both the sample size and the number of time points is moderately large. Specifically, both the true dimension R^* and the fitted dimension R is set to be 3. The number of attributes is $p = 5$, the number of network nodes $n = 100$ and the number of time points $T = 100$. In this case the true vector of eigenvalues is $\boldsymbol{\lambda} = (0, \lambda_2, \lambda_3)$ and we set

$$\boldsymbol{\Psi} = \begin{pmatrix} \psi_{1,1} & 0 & 0 \\ \psi_{2,1} & 0 & 0 \\ 0 & \psi_{3,2} & \psi_{3,3} \\ 0 & \psi_{4,2} & \psi_{4,3} \\ \psi_{5,1} & \psi_{5,2} & \psi_{5,3} \end{pmatrix}, \quad \boldsymbol{\Xi} = \begin{pmatrix} \xi_{1,1} & 0 & 0 \\ \xi_{2,1} & 0 & 0 \\ 0 & \xi_{3,2} & 0 \\ 0 & \xi_{4,2} & \xi_{4,3} \\ \xi_{5,1} & \xi_{5,2} & \xi_{5,3} \end{pmatrix},$$

so that $q_S = 2$, $q_N = 0$ and $q_A = 1$. The fifth and final simulation scenario aims at addressing performance of the JLAFAC model under a misspecified setting. In this simulation, the nodal attributes are simulated from (2.3) with the same selection coefficient matrices $\boldsymbol{\Psi}$ and $\boldsymbol{\Xi}$ used to simulate dataset 1. However, in this case the networks is simulated conditionally on the nodal attributes using a temporal exponential random graph model (TERGM) that includes two sufficient statistics, $\mathcal{S}_1(\mathbf{Y}_l, \mathbf{Z}_l) = \sum_{i=1}^n (\sum_{j=1}^n y_{i,j,l}) z_{i,1,l} z_{i,2,l}$ and $\mathcal{S}_2(\mathbf{Y}_l, \mathbf{Z}_l) = \sum_{i,j=1}^n y_{i,j,l}$ with the corresponding coefficients $\theta_1 = l/T$, $\theta_2 = 0.5$. The network data is thus simulated from a totally misspecified TERGM with complex higher order dependence between the network and nodal attributes. This simulation sets $p = 5$, $n = 40$ and $T = 10$. We fit the JLAFAC model with $R = 5$ in this simulated data. In scenario 1-4 we generate the non-zero entries of $\boldsymbol{\lambda}$,

ρ_u	σ_u^2	ρ_v	σ_v^2	ρ_α	σ_α^2	ρ_β	σ_β^2	ρ_η	σ_η^2	ρ_μ	σ_μ^2
0.4	1.0	0.25	1.0	0.35	1.0	0.7	0.5	0.1	0.25	0.3	0.6

Table 1: True value of the hyperparameters used to generate data for our simulation study.

Ψ and Ξ from normal distributions with means 2, 2 and -3 and variances 2, 0.5 and 0.5, respectively. Scenario 5 simulates only Ψ and Ξ from normal distributions with means 1 and -0.5 and variances 0.5 and 0.5, respectively. On the other hand, the true value of the random processes $\{\eta_k(t)\}$, $\{\alpha_{k,r}(t)\}$, $\{\beta_{k,r}(t)\}$, $\mu(t)$, $\{u_{i,r}(t)\}$ and $\{v_{i,r}(t)\}$ are generated using the hyperparameters presented in Table 1. Finally, all the attribute values are generated from Gaussian distributions according to Equation (2.3), with variances $\phi_1^{-1}, \dots, \phi_p^{-1}$ drawn i.i.d from $U(0.1, 0.5)$.

4.1 Link and attribute prediction

In order to evaluate the ability of the model to predict links, we carry out an out-of-sample cross-validation exercise. More specifically, we randomly select 400 dyads to hold-out as a validation set and then estimate our model treating these dyads as if they were missing at random. Figure 1 shows the area under the ROC curve (AUC) of JLAFAC for the observations in the validation set. In order to assess the value of jointly modeling nodal attributes and network features, we also present the results generated by fitting only the components of the model associated with the network data, referred to as the marginal model. It is evident that in simulations in which network and attributes are associated in the true data generation mechanism (scenarios 1, 3, 4 and 5), joint modeling has clear advantages in terms of predicting missing links. This can be attributed to the fact that the shared latent factors in such cases are estimated based on both the relational and nodal attribute data, leading to more accurate estimation. This is true even in the case in which the true data generating mechanism is different from our model (scenario 5). On the other hand, the predictive performance of both models is essentially identical when $q_S = 0$, as would be expected.

In addition to the marginal model as a competitor to JLAFAC, we also implement logistic regression of $y_{i,j}(t_l)$ on the $(p+1)^2$ covariates obtained by taking the nodal attributes corresponding to the i th and j th node along with the all possible interactions between these nodal attributes, $[\mathbf{1}, \mathbf{z}_{i,l}, \mathbf{z}_{j,l}, \mathbf{z}_{i,l} \otimes \mathbf{z}_{j,l}]$, where $\mathbf{z}_{i,l} = (z_{i,1}(t_l), \dots, z_{i,p}(t_l))'$ and $\mathbf{z}_{j,l} = (z_{j,1}(t_l), \dots, z_{j,p}(t_l))'$. This approach exhibits performance similar to a random classifier in all three simulation studies and is thus omitted from the comparison.

We also carry out prediction of attribute values based on both JLAFAC and a dynamic factor model that ignores the relational information. Table 2 presents the mean squared error (MSE) values obtained under each model and for each scenario. Again, the results suggest that jointly modeling both sources of information leads to much improved predictions of nodal attributes in scenarios 1, 3, 4 and 5, and that those advantages disappear in scenario 2 where the nodal attributes evolve independently from the relational data.

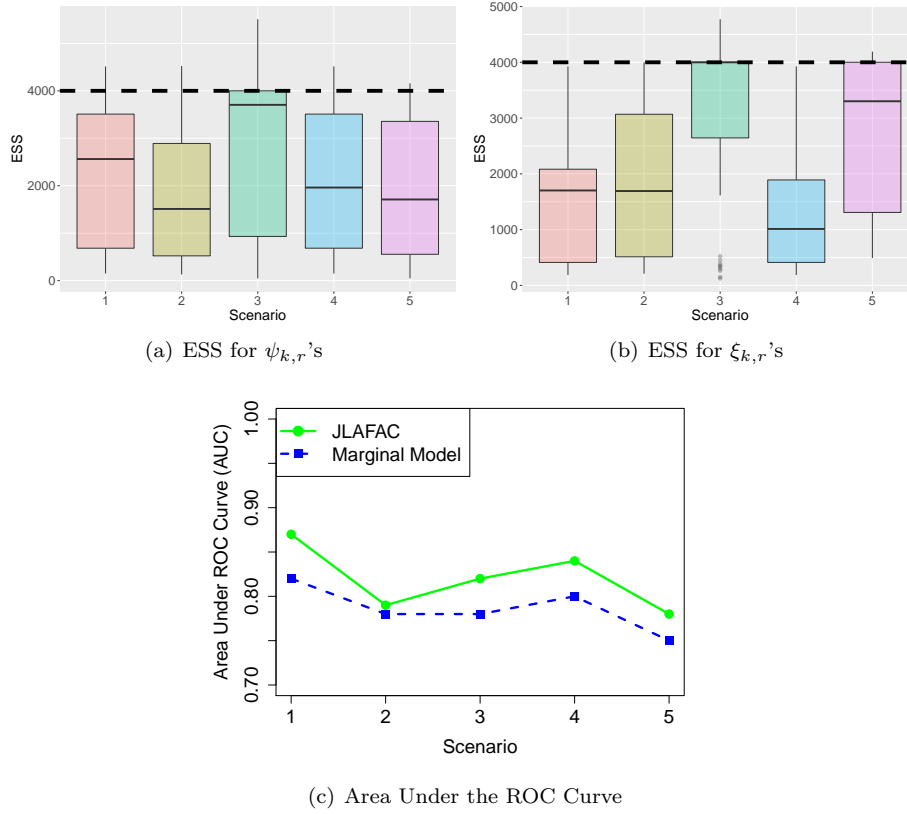


Figure 1: Figures in the first row presents boxplots of Effective Sample Sizes (ESS) for scenario 1-5 for $\psi_{k,r}$'s and $\xi_{k,r}$'s. The figures have horizontal dashed lines at 4000 to indicate that the ESS are calculated on 4000 post burn-in thinned MCMC samples. Second row presents Area Under the ROC Curve (AUC) for our out-of-sample cross-validation in five different simulation scenarios. Continuous curves correspond to the results under our JLA FAC model, while the dashed curves come from a marginal model that does not incorporate the attribute data.

Scenario	MSE				
	1	2	3	4	5
JLA FAC	1.57	1.17	0.09	0.08	0.66
Dynamic Factor	3.38	1.04	0.13	0.59	0.74

Table 2: Mean squared error (MSE) for the two competitors in simulation cases 1 to 5. MSE is calculated as the mean absolute squared deviation of predicted and true values of all the predictors.

4.2 Patterns of association between network and nodal attributes

Prior and posterior distributions for q_S , q_N , q_A and $R^* = q_S + q_N + q_A$ under JLAFAC for each of the three simulation scenarios are recorded in Table 3. In scenario 1 and 4 the model is capable of identifying both the right dimension of the latent space, the breakdown into systemic and idiosyncratic components, and thereby the lack of independence between the network and attribute data. In scenario 2 the model also is able to correctly assess the dimension of dimensions of the latent space, and thereby the independence of network and attribute data, In scenario 3 the model is able to estimate q_S and q_N accurately, although q_A (and therefore R^*) is slightly overestimated. In spite of the misspecification in scenario 5, our model is still able to provide moderate evidence of association between the network structure and the attributes ($P(q_S > 0) \approx 0.65$), as well as evidence of residual correlation between the attributes ($P(q_A > 0) \approx 1$).

Finally, Figure 2 shows the posterior distributions for $q_{S,1}, \dots, q_{S,p}$ on each of the three simulation scenarios. Again, the model is capable of correctly identifying the number of systemic components associated with each attribute for scenarios 1, 2, 3 and 4. The results are also presented for scenario 5, which shows $P(q_{S,1} > 0) \approx 0.27, P(q_{S,2} > 0) \approx 0.51, P(q_{S,3} > 0) \approx 0.15, P(q_{S,4} > 0) \approx 0.12, P(q_{S,5} > 0) \approx 0.21$. Thus the model correctly identifies the association between attribute 2 and the network and no association between nodal attributes 3,4,5 with the network. For nodal attribute 1, the model detects a weak association with the network. Overall, it does a fair job even under a complex mis-specified setting.

4.3 Computational Complexity

The computational complexity of the proposed model is dominated by the complexity of updating $\mathbf{u}_{i,1}, \dots, \mathbf{u}_{i,T}$, $\mathbf{v}_{i,1}, \dots, \mathbf{v}'_{i,T}$, $\boldsymbol{\alpha}_{k,1}, \dots, \boldsymbol{\alpha}'_{k,T}$ and $\boldsymbol{\beta}_{k,1}, \dots, \boldsymbol{\beta}'_{k,T}$ for $k = 1, \dots, p$ and $i = 1, \dots, n$. However, we note that \mathbf{u}_i and \mathbf{v}_i , $i = 1, \dots, n$ can be sampled in parallel over all network nodes. Similarly, $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$, $k = 1, \dots, p$ can be updated in parallel over nodal attributes. Hence, parallelization allows us to supports data with moderately large. Furthermore, the choice of an OU process for modeling the dynamics of the latent factors allows us to use a forward-filtering, backward-sampling algorithm to jointly sample parameters over time (e.g., see Frühwirth-Schnatter, 1994). Hence, the Gibbs sampling algorithm can support inference for a moderately-sized dataset. To illustrate this, per iteration run times are provided in Table 4 for different choices of n and T assuming parallelizing computation over multiple servers. In Section 6 we offer more discussion on several other strategies to ease computation for very large n and T and for general Gaussian process priors on the latent functions.

4.4 Sensitivity Analysis

All of the results presented above were obtained after fixing $v_0 = 100$. Also as mentioned before, we choose a_ψ and a_λ so as to ensure approximately same distribution for q_S, q_A, q_N, q_O a priori. We investigated the sensitivity of the posterior means of q_S, q_N, q_A by rerunning the model with $v_0 = 1000, 10000$, $a_\psi = b_\psi = 2, 6, 10$ and $a_\lambda = b_\lambda = 1, 3, 6$.

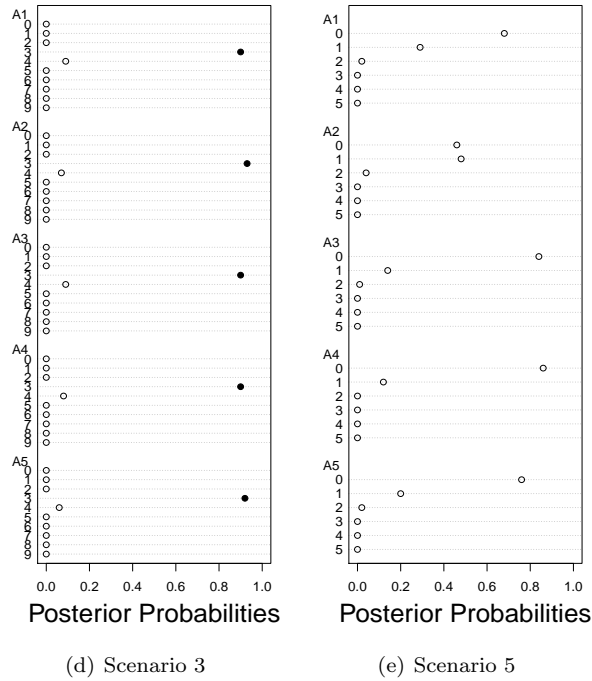
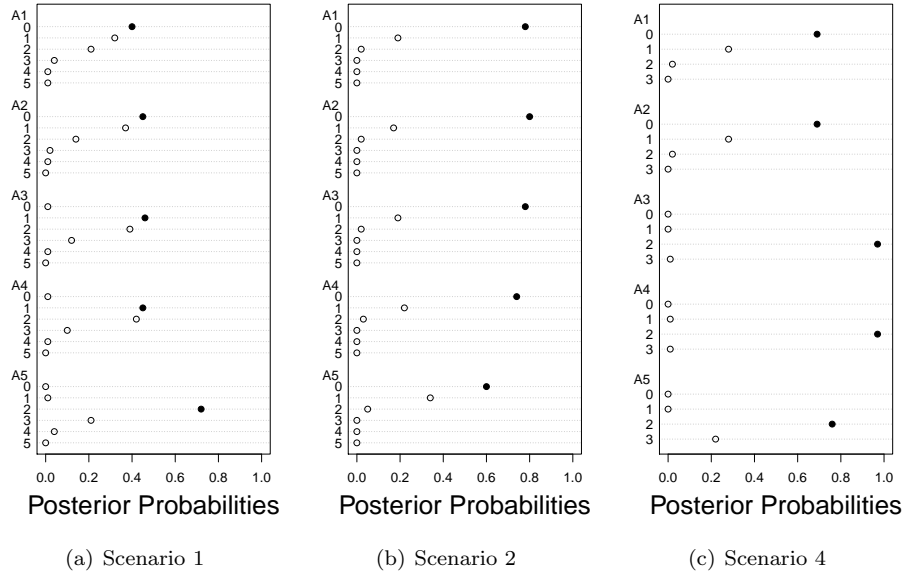


Figure 2: Posterior distribution for $q_{S,1}, \dots, q_{S,p}$ under each of our five simulation scenarios. Filled bullets indicate the true value of the summary. A_k denotes the k th attribute. For simulation 5, the fitted model is different from the true model. Thus the truth is not specified.

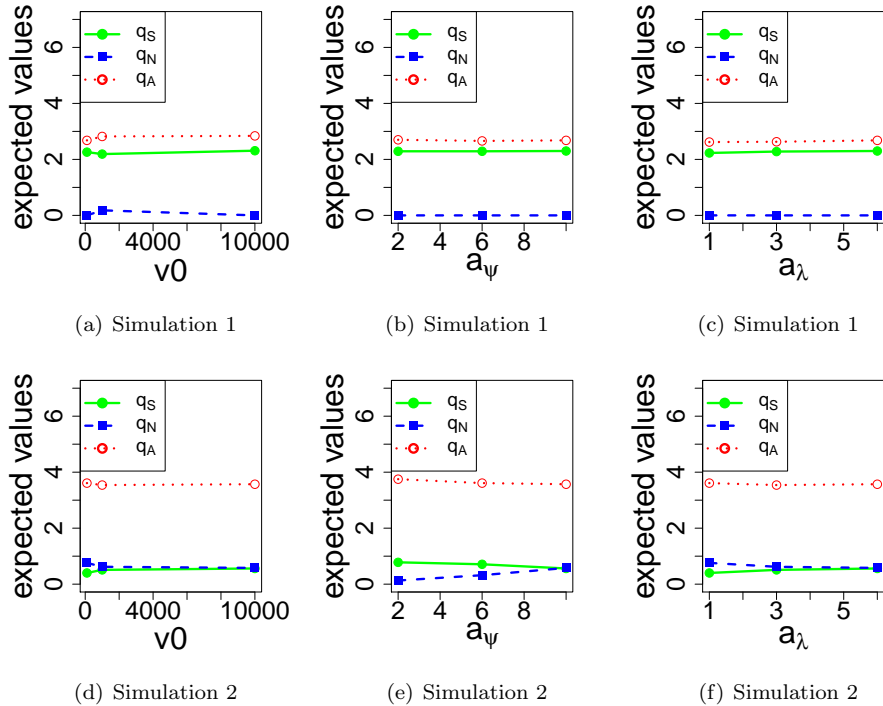


Figure 3: Posterior means of q_S , q_N and q_A for different choices of a_λ , a_ψ and v_0 .

To save space, we just present in Figure 3 the results for two representative simulations, Scenario 1 and 2. The figures show that the results are quite robust to the choice of these hyperparameters.

5 Analysis of Longitudinal International Relationship Data with Country Specific Indicators of Economic Performance

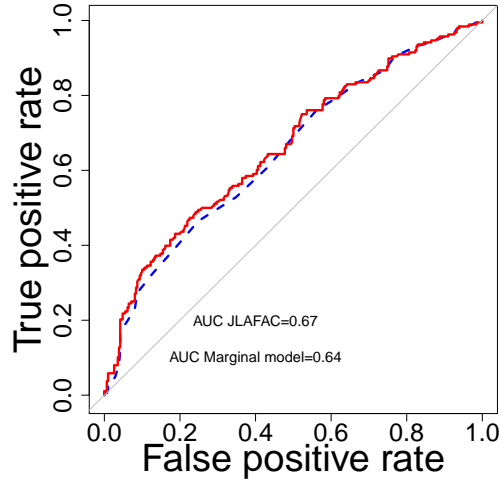
In this section we investigate the association between international conflicts and various economic indicators in a set of 22 countries. The list of countries includes all five permanent members of the United Nations security council (China, France, Russia, USA and the UK), a few members of the G-20 group (Australia, Germany, India, Japan, South Korea, Turkey), as well as various countries from the middle East (Egypt, Iran, Iraq, Israel, Lebanon, Palestine, Sudan), eastern Europe (Ukraine, Georgia) and south Asia (Afghanistan, Pakistan).

Annual networks of interactions among countries between 2004 and 2014 were de-

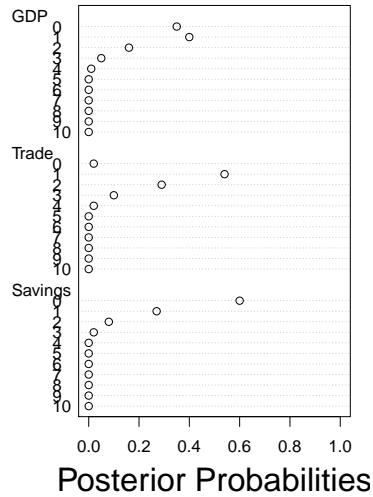
fined by linking two countries if there was at least one “positive verbal action” during the first week of the year (Goldstein, 1992). Examples of positive verbal actions include granting diplomatic recognition and positive verbal support in international forums. These data on international relations is available from <http://www.stat.washington.edu/people/pdhoff/Code> and was previously studied in Hoff (2015). In addition to the relational data, we consider the evolution of three country-specific economic indicators, which are modeled using a conditionally Gaussian model: the growth rate of Gross Domestic Product (GDP) Per Capita, the level of National Savings as a percentage of GDP, and the amount of International Trade as a percentage of GDP. Data on these indicators are available from the World Development Indicators Database, which is published annually by the World Bank. For this dataset we estimate JLAFAC using $R = 10$ latent dimensions. All inferences presented below are based on 50,000 samples from our MCMC algorithm obtained after a burn-in period of 10,000 samples and thinning the chain every 10 iterations.

As in Section 4, we first evaluate the ability of JLAFAC to make out-of-sample predictions. In this case we hold out 500 randomly-chosen dyads, whose values are predicted using the rest of the data. Figure 4 presents the ROC curve and the AUC value under both JLAFAC and a dynamic network model that ignores the attribute data (this comparison is similar to the one we carried out in Section 4). As in the simulation study, jointly modeling the relational and nodal attribute data provides some improvement in the predictions, although the improvement in this case is relatively minor. We also calculate the RMSE of estimating the predictors which turns out to be 0.065 as compared to the value of 0.147 for the marginal model.

Next we investigate the pattern of association between the network and country specific attributes in the data. Table 5 presents the prior and posterior distributions for q_S , q_N , q_A and $R^* = q_S + q_N + q_A$. It is evident from the Table that the data favors the use of 7 latent dimensions. Furthermore, the model provides strong evidence for the presence of at least one systemic component ($P(q_S > 0 \mid \text{data}) \approx 0.95$), suggesting that the network structure and the nodal attributes are indeed associated. To investigate the pattern of association in more detail we present in Figure 4 the posterior distributions for $q_{S,1}$, $q_{S,2}$ and $q_{S,3}$, where subscripts 1,2,3 correspond to growth of GDP per capita, ratio of trade to GDP and ratio of savings to GDP respectively. Note that while there is strong evidence that the ratio of trade to GDP and GDP growth per capita are associated with the structure of international relations ($P(q_{S,1} > 0 \mid \text{data}) \approx 0.65$ and $P(q_{S,2} > 0 \mid \text{data}) \approx 0.98$), the evidence for an association with national savings is weak ($P(q_{S,3} > 0 \mid \text{data}) \approx 0.40$, respectively). A review of the literature suggests that our results are consistent with previous findings. Indeed, there is a well established relationship between international trade and conflict, which was first highlighted in the classic paper by Polachek (1980) (see also Reuveny and Kang, 1996, Morrow, 1999 and Reuveny, 2000). Similarly, there is some evidence in the literature for an association between GDP growth and international conflict (e.g., see Rodrik, 1999, Anderton et al., 2003, Miguel et al., 2004 and Polachek and Sevastianova, 2012), although the evidence is highly disputed and the direction of the causal effect often unclear. On the other hand, as far as we could find, an association between international conflict and the level aggregate national savings has not been proposed or reported in the literature.



(a) ROC curves



(b) Systemic dimensions

Figure 4: Figure in the top presents ROC curves and corresponding AUC values for JLAFACT and a dynamic network model that ignores information about network attributes in the international relations data. Solid and dashed curves represent ROC curves for JLAFACT and dynamic network model without network attributes respectively. Figure in the bottom shows posterior distributions for $q_{S,1}$, $q_{S,2}$ and $q_{S,3}$ for the international relations data.

6 Conclusion

This article introduces the idea of jointly modeling of network and associated nodal attributes over time. Our proposed framework relied on modeling the network and nodal attributes jointly through latent factor representations, with some latent factors shared across both models to introduce dependence of directional dyad between nodes on nodal attributes. Evolution of both network and nodal attributes over time is modeled by allowing the latent factors to vary over time, and inference is carried out using a Bayesian approach.

Because of the application that motivated our work we have assumed that the network process $\mathbf{Y}(t)$ is directed and binary. The model can easily be reformulated to accommodate undirected networks by replacing equations (2.1) and (2.2) with

$$y_{i,j}(t) \sim \text{Ber}(\theta_{i,j}(t)), \quad \theta_{i,j}(t) = G\left(\mu(t) + \sum_{r=1}^R \lambda_r u_{i,r}(t) u_{j,r}(t)\right),$$

and

$$E\{z_{i,k}(t)\} = H_k\left(\eta_k(t) + \sum_{r=1}^R \psi_{k,r} \alpha_{k,r}(t) u_{i,r}(t)\right),$$

respectively. Similarly, situation in which $y_{i,j}(t)$ belong to other members of the exponential family can be easily accommodated through appropriate generalized linear factor analysis models.

An important feature of our model is its ability to yield dependence / independence among network and attributes. While our simulations suggests that these test might be somewhat robust to model misspecification (recall the results from scenario 5 in our simulation study), it is important to remember that all of our tests are conditioned on the specific model structure we have assumed. In particular, while the eigenvalue model for networks that we employ here is quite flexible, there is a possibility that it might not be able to capture all structural features of the network that affect the attributes. Similarly, our model does not incorporate lags of the latent process. While it is important to keep these issues in mind, we should note the dependence on model assumptions is a common limitation of most statistical approaches. For example, tests of conditional independence for Gaussian graphical models are valid only if the data is actually Gaussian.

Although the article presents an simulation study of the proposed approach for moderately large network and time points, one important extension of the proposed model would be to scale it for larger datasets with millions of individuals connected under social networking and observed over a large number of time points. Since OU process is a variant of the Gaussian process and there is an emerging literature on modeling Gaussian processes for big data, we intend to employ such computationally convenient models to specify correlations across large number of time points. On the other hand, a potential issue in extending these models for massive networks would be to estimate a large number of latent variables. We propose to adapt strategies described

in Guhaniyogi et al. (2017); Guhaniyogi and Banerjee (2018); Heaton et al. (2017) for efficient estimation of large number of correlated latent variables. It is also interesting to extend our methodology for online social networks. Some of these constitutes our current work.

Supplementary Material

Supplementary Material: Joint Modeling of Longitudinal Relational Data and Exogenous Variables (). Supplementary material consists of three sections. Section 1 presents the full conditional posteriors for the MCMC algorithm. Section 2 states and proves theorems related to large support of the prior. Section 3 argues identifiability of q_S , q_A , q_N and R^* .

References

- Albert, J. H. and Chib, S. (1993). “Bayesian analysis of binary and polychotomous response data.” *Journal of the American statistical Association*, 88(422): 669–679. [8](#)
- Anderton, C. H., Beck, N., Carter, J. R., Dorussen, H., Gartzke, E., Gissinger, R., Gleditsch, N., Hegre, H., Levy, J. S., Li, Q., et al. (2003). *Globalization and armed conflict*. Rowman & Littlefield Publishers. [19](#)
- Brandes, U., Indlekofer, N., and Mader, M. (2012). “Visualization methods for longitudinal social networks and stochastic actor-oriented modeling.” *Social Networks*, 34(3): 291–308. [2](#)
- Butland, G., Peregrín-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., et al. (2005). “Interaction network containing conserved and essential protein complexes in *Escherichia coli*.” *Nature*, 433(7025): 531–537. [1](#)
- Christakis, N. A. and Fowler, J. H. (2007). “The spread of obesity in a large social network over 32 years.” *n engl j med*, 2007(357): 370–379. [1](#), [2](#)
- De la Haye, K., Robins, G., Mohr, P., and Wilson, C. (2010). “Obesity-related behaviors in adolescent friendship networks.” *Social Networks*, 32(3): 161–167. [3](#)
- Desmarais, B. A. and Cranmer, S. J. (2010). “Consistent confidence intervals for maximum pseudolikelihood estimators.” In *Proceedings of the Neural Information Processing Systems 2010 Workshop on Computational Social Science and the Wisdom of Crowds*. Citeseer. [2](#)
- (2012). “Statistical mechanics of networks: Estimation and uncertainty.” *Physica A: Statistical Mechanics and its Applications*, 391(4): 1865–1876. [2](#)
- Doreian, P. (2001). “Causality in social network analysis.” *Sociological Methods & Research*, 30(1): 81–114. [2](#)
- Durante, D. and Dunson, D. B. (2014). “Nonparametric Bayes dynamic modelling of relational data.” *Biometrika*, 101(4): 883–898. [3](#)

- (2017). “Bayesian inference and testing of group differences in brain networks.” *Bayesian Analysis*. 2
- Durante, D., Mukherjee, N., and Steorts, R. C. (2017). “Bayesian learning of dynamic multilayer networks.” *Journal of Machine Learning Research*, 18(43): 1–29. 3
- Fosdick, B. K. and Hoff, P. D. (2015). “Testing and modeling dependencies between a network and nodal attributes.” *Journal of the American Statistical Association*, 110(511): 1047–1056. 2, 3, 5, 7, 8, 11
- Fowler, J. H. and Christakis, N. A. (2008). “Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study.” *Bmj*, 337: a2338. 1, 2
- Frühwirth-Schnatter, S. (1994). “Data augmentation and dynamic linear models.” *Journal of time series analysis*, 15(2): 183–202. 16
- George, E. I. and McCulloch, R. E. (1993). “Variable selection via Gibbs sampling.” *Journal of the American Statistical Association*, 88(423): 881–889. 7
- Goldstein, J. S. (1992). “A conflict-cooperation scale for WEIS events data.” *Journal of Conflict Resolution*, 36(2): 369–385. 19
- Guhaniyogi, R. and Banerjee, S. (2018). “Meta-kriging: Scalable Bayesian modeling and inference for massive spatial datasets.” *Technometrics*, (just-accepted). 22
- Guhaniyogi, R., Li, C., Savitsky, T. D., and Srivastava, S. (2017). “A Divide-and-Conquer Bayesian Approach to Large-Scale Kriging.” *arXiv preprint arXiv:1712.09767*. 22
- Handcock, M. S., Robins, G., Snijders, T., Moody, J., and Besag, J. (2003). “Assessing degeneracy in statistical models of social networks.” Technical report, Working paper. 2
- Hanneke, S., Fu, W., Xing, E. P., et al. (2010). “Discrete temporal models of social networks.” *Electronic Journal of Statistics*, 4: 585–605. 2
- Hanneke, S. and Xing, E. P. (2007). “Discrete temporal models of social networks.” In *Statistical network analysis: Models, issues, and new directions*, 115–125. Springer. 2
- Heaton, M. J., Datta, A., Finley, A., Furrer, R., Guhaniyogi, R., Gerber, F., Gramacy, R. B., Hammerling, D., Katzfuss, M., Lindgren, F., et al. (2017). “Methods for analyzing large spatial data: A review and comparison.” *arXiv preprint arXiv:1710.05013*. 22
- Hoff, P. D. (2005). “Bilinear mixed-effects models for dyadic data.” *Journal of the American Statistical Association*, 100(469): 286–295. 1, 2, 5
- (2009). “Multiplicative latent factor models for description and prediction of social networks.” *Computational and mathematical organization theory*, 15(4): 261. 4
- (2015). “Multilinear tensor regression for longitudinal relational data.” *The annals of applied statistics*, 9(3): 1169. 19

- Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). “Latent space approaches to social network analysis.” *Journal of the American Statistical Association*, 97(460): 1090–1098. [1](#), [5](#)
- Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983). “Stochastic blockmodels: First steps.” *Social Networks*, 5(2): 109–137. [1](#)
- Holland, P. W. and Leinhardt, S. (1981). “An exponential family of probability distributions for directed graphs.” *Journal of the American Statistical Association*, 76(373): 33–50. [1](#)
- Ishwaran, H., Rao, J. S., et al. (2005). “Spike and slab variable selection: frequentist and Bayesian strategies.” *The Annals of Statistics*, 33(2): 730–773. [7](#)
- Kalyagin, V., Koldanov, A., Koldanov, P., Pardalos, P., and Zamaraev, V. (2014). “Measures of uncertainty in market network analysis.” *Physica A: Statistical Mechanics and its Applications*, 413: 59–70. [1](#)
- Lin, X. (2010). “Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables.” *Journal of Labor Economics*, 28(4): 825–860. [2](#)
- Miguel, E., Satyanath, S., and Sergenti, E. (2004). “Economic shocks and civil conflict: An instrumental variables approach.” *Journal of Political Economy*, 112(4): 725–753. [19](#)
- Morrow, J. D. (1999). “How could trade affect conflict?” *Journal of Peace Research*, 36(4): 481–489. [19](#)
- Niezink, N. M., Snijders, T. A., et al. (2017). “Co-evolution of social networks and continuous actor attributes.” *The Annals of Applied Statistics*, 11(4): 1948–1973. [3](#)
- Nishihara, R., Murray, I., and Adams, R. P. (2014). “Parallel MCMC with generalized elliptical slice sampling.” *The Journal of Machine Learning Research*, 15(1): 2087–2112. [6](#), [9](#)
- Polachek, S. W. (1980). “Conflict and trade.” *Journal of Conflict Resolution*, 24(1): 55–78. [19](#)
- Polachek, S. W. and Sevastianova, D. (2012). “Does conflict disrupt growth? Evidence of the relationship between political instability and national economic performance.” *The Journal of International Trade & Economic Development*, 21(3): 361–388. [19](#)
- Reuveny, R. (2000). “The trade and conflict debate: A survey of theory, evidence and future research.” *Peace Economics, Peace Science and Public Policy*, 6(1). [19](#)
- Reuveny, R. and Kang, H. (1996). “International trade, political conflict/cooperation, and Granger causality.” *American Journal of Political Science*, 943–970. [19](#)
- Robins, G. and Pattison, P. (2001). “Random graph models for temporal processes in social networks.” *Journal of Mathematical Sociology*, 25(1): 5–41. [2](#)
- Robins, G., Pattison, P., Kalish, Y., and Lusher, D. (2007). “An introduction to ex-

- ponential random graph (p^*) models for social networks.” *Social networks*, 29(2): 173–191. [1](#)
- Rodríguez, A. and Moser, S. (2015). “Measuring and accounting for strategic abstentions in the US Senate, 1989–2012.” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64(5): 779–797. [3](#)
- Rodrik, D. (1999). “Where did all the growth go? External shocks, social conflict, and growth collapses.” *Journal of economic growth*, 4(4): 385–412. [19](#)
- Sarkar, P. and Moore, A. W. (2006). “Dynamic social network analysis using latent space models.” In *Advances in Neural Information Processing Systems*, 1145–1152. [3](#)
- Scott, J. G., Berger, J. O., et al. (2010). “Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem.” *The Annals of Statistics*, 38(5): 2587–2619. [7](#)
- Sewell, D. K. and Chen, Y. (2015). “Latent space models for dynamic networks.” *Journal of the American Statistical Association*, 110(512): 1646–1657. [3](#)
- Shoham, D. A., Hammond, R., Rahmandad, H., Wang, Y., and Hovmand, P. (2015). “Modeling social norms and social influence in obesity.” *Current epidemiology reports*, 2(1): 71–79. [2](#)
- Snijders, T. A. (2002). “Markov chain Monte Carlo estimation of exponential random graph models.” *Journal of Social Structure*, 3(2): 1–40. [2](#)
- (2013). “Stochastic actor-oriented models for network change.” In *Evolution of social networks*, 193–216. Routledge. [2](#)
- Snijders, T. A., Van de Bunt, G. G., and Steglich, C. E. (2010). “Introduction to stochastic actor-based models for network dynamics.” *Social networks*, 32(1): 44–60. [2](#)
- Wasserman, S. and Anderson, C. (1987). “Stochastic a posteriori blockmodels: Construction and assessment.” *Social Networks*, 9(1): 1–36. [1](#)
- Watts, D. J. and Dodds, P. (2009). “Threshold models of social influence.” *The Oxford handbook of analytical sociology*, 475–497. [2](#)

		Scenario 1		Scenario 2		Scenario 3		Scenario 4		Scenario 5	
		Prior	Post	Prior	Post	Prior	Post	Prior	Post	Prior	Post
q_S	0	0.31	0.00	0.31	0.49	0.18	0.00	0.42	0.00	0.31	0.35
	1	0.28	0.01	0.28	0.38	0.16	0.00	0.30	0.00	0.28	0.52
	2	0.21	0.70	0.21	0.10	0.14	0.00	0.18	0.76	0.21	0.10
	3	0.13	0.22	0.13	0.01	0.12	0.79	0.08	0.22	0.13	0.01
	4	0.05	0.04	0.05	0.00	0.10	0.18	-	-	0.05	0.00
	5	0.01	0.00	0.01	0.00	0.08	0.02	-	-	0.01	0.00
	6	-	-	-	-	0.06	0.00	-	-	-	-
	7	-	-	-	-	0.05	0.00	-	-	-	-
	8	-	-	-	-	0.03	0.00	-	-	-	-
	9	-	-	-	-	0.02	0.00	-	-	-	-
	10	-	-	-	-	0.01	0.00	-	-	-	-
q_N	0	0.37	1	0.37	0.40	0.32	0.80	0.59	1.00	0.37	0.78
	1	0.30	0.00	0.30	0.60	0.21	0.17	0.26	0.00	0.30	0.21
	2	0.19	0.00	0.19	0.00	0.15	0.02	0.10	0.00	0.19	0.00
	3	0.09	0.00	0.09	0.00	0.10	0.00	0.03	0.00	0.09	0.00
	4	0.03	0.00	0.03	0.00	0.07	0.00	-	-	0.03	0.00
	5	0.00	0.00	0.00	0.00	0.05	0.00	-	-	0.00	0.00
	6	-	-	-	-	0.03	0.00	-	-	-	-
	7	-	-	-	-	0.02	0.00	-	-	-	-
	8	-	-	-	-	0.01	0.00	-	-	-	-
	9	-	-	-	-	0.00	0.00	-	-	-	-
	10	-	-	-	-	0.00	0.00	-	-	-	-
q_A	0	0.31	0.00	0.31	0.00	0.18	0.14	0.42	0.22	0.31	0.00
	1	0.27	0.04	0.27	0.00	0.16	0.30	0.30	0.77	0.27	0.00
	2	0.21	0.22	0.21	0.05	0.14	0.26	0.18	0.00	0.21	0.02
	3	0.12	0.70	0.12	0.28	0.12	0.18	0.08	0.00	0.12	0.14
	4	0.05	0.01	0.05	0.62	0.10	0.05	-	-	0.05	0.70
	5	0.01	0.00	0.01	0.04	0.08	0.04	-	-	0.01	0.12
	6	-	-	-	-	0.07	0.02	-	-	-	-
	7	-	-	-	-	0.05	0.00	-	-	-	-
	8	-	-	-	-	0.03	0.00	-	-	-	-
	9	-	-	-	-	0.02	0.00	-	-	-	-
	10	-	-	-	-	0.01	0.00	-	-	-	-
R^*	0	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00
	1	0.03	0.00	0.03	0.00	0.00	0.00	0.10	0.00	0.03	0.00
	2	0.09	0.00	0.09	0.00	0.01	0.00	0.26	0.01	0.09	0.00
	3	0.19	0.00	0.19	0.00	0.02	0.12	0.59	0.98	0.19	0.00
	4	0.29	0.00	0.29	0.20	0.03	0.38	-	-	0.29	0.04
	5	0.37	1.00	0.37	0.79	0.05	0.25	-	-	0.37	0.95
	6	-	-	-	-	0.07	0.19	-	-	-	-
	7	-	-	-	-	0.10	0.05	-	-	-	-
	8	-	-	-	-	0.15	0.00	-	-	-	-
	9	-	-	-	-	0.21	0.00	-	-	-	-
	10	-	-	-	-	0.32	0.00	-	-	-	-

Table 3: Prior and posterior distributions on q_S , q_N , q_A and $R^* = q_S + q_N + q_A$ under JLAFAC for the five datasets in our simulation study. Grey backgrounds indicate the true value under each scenario. Note that Scenario 5 has no ground truth.

	$n = 40$	$n = 70$	$n = 100$
$T = 10$	0.09	0.16	0.34
$T = 50$	1.28	2.09	2.85
$T = 100$	5.02	11.58	18.23

Table 4: Computation time in seconds of JLA FAC per iteration. $R = 5$ is assumed for model fitting.

	q_S		q_N		q_A		R^*	
	Prior	Posterior	Prior	Posterior	Prior	Posterior	Prior	Posterior
0	0.16	0.01	0.38	0.04	0.16	0.05	0.00	0.00
1	0.14	0.24	0.21	0.22	0.14	0.21	0.00	0.00
2	0.13	0.33	0.14	0.39	0.13	0.30	0.00	0.00
3	0.12	0.25	0.09	0.21	0.12	0.24	0.01	0.00
4	0.10	0.10	0.06	0.08	0.10	0.12	0.02	0.01
5	0.09	0.03	0.04	0.02	0.09	0.04	0.04	0.11
6	0.07	0.00	0.02	0.00	0.07	0.01	0.06	0.25
7	0.06	0.00	0.01	0.00	0.06	0.00	0.09	0.31
8	0.04	0.00	0.00	0.00	0.04	0.00	0.14	0.20
9	0.03	0.00	0.00	0.00	0.03	0.00	0.22	0.07
10	0.01	0.00	0.00	0.00	0.01	0.00	0.38	0.01

Table 5: Prior and posterior distributions for q_S , q_N , q_A and $R^* = q_S + q_N + q_A$ for the international relations data. The bolded values are the probabilities of the mode.