Submitted to the Annals of Applied Statistics

DETECTING STRUCTURAL CHANGES IN LONGITUDINAL NETWORK DATA

By Jong Hee Park[*,†], and Yunkyu Sohn[‡]

Seoul National University[†] and Princeton University[‡]

Dynamic modeling of longitudinal networks has been an increasingly important topic in applied research. While longitudinal network data commonly exhibit dramatic changes in its structures, existing methods have largely focused on modeling smooth topological changes over time. In this paper, we develop a hidden Markov multilinear tensor model (HMTM) that combines the multilinear tensor regression model (Hoff, 2011) with a hidden Markov model using Bayesian inference. We model changes in network structure as shifts in discrete states yielding particular sets of network generating parameters. Our simulation results demonstrate that the proposed method correctly detects the number, locations, and types of changes in latent node characteristics. We apply the proposed method to international military alliance networks from 1816 to 2012 and identify structural changes in the coalition structure among nations.

1. Introduction. Dynamic modeling of longitudinal network data has been an increasingly important topic in social, biological, and other fields of science.[1] A longitudinal network dataset can be characterized as a tensor $\mathcal{Y} = \{\mathbf{Y}_t | t \in \{1, \ldots, T\}\} \in \mathbb{R}^{N \times N \times T}$, or a multilayer network with T time-ordered layers, in which an $N \times N$ square matrix $\mathbf{Y}_t = \{y_{ijt} | i, j \in \{1, \ldots, N\}\}$ represents relationships between all pairs of N actors at time t. A generic challenge of dynamic network modeling is to recover the time-varying network generating process of the tensor data $\mathcal{Y}$ in the presence of simultaneous dependence between dyadic observations and dynamic time series.

Hoff (2011, 2015) presented a multilinear tensor regression model (MTRM) for recovering covariate effects and network effects in the generating process of multilayer network data. The multiplicative formulation of network effects within MTRM is particularly useful for dynamic network modeling since the multiplicative formulation characterizes the varying dependence

*Corresponding Author

Keywords and phrases: Bayesian analysis, network change, hidden Markov model, military alliance, WAIC

[1] For a range of examples for longitudinal network analysis, see Holme and Saramäki (2012) and references therein.

structure between dyadic observations over time. Importantly, it allows us to *jointly estimate time-specific network generation rules with time-constant latent node positions*. Let $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_N)^\top \in \mathbb{R}^{N \times R}$ be the R-dimensional latent node positions of N nodes and $\mathbf{v}_t = (v_{1t}, \ldots, v_{Rt}) \in \mathbb{R}^R$ be a vector exhibiting dimension-specific node connection rules at time t. In this formulation, network effects are modeled by the product of latent node traits ($\mathbf{u}_i$ for node i and $\mathbf{u}_j$ for node j) and layer-specific network generation rules ($\mathbf{v}_t$ at time t or tth layer) as follows:

$$
\begin{aligned}
&(1)\ \Pr(y_{i,j,t} = 1 | \mathbf{x}_{i,j,t}, \mathbf{u}_i, \mathbf{u}_j, \mathbf{v}_t) &&= \mathbf{x}_{i,j,t}\boldsymbol{\beta} + \langle \mathbf{u}_i, \mathbf{v}_t, \mathbf{u}_j \rangle + \epsilon_{i,j,t} \\
&(2)\ \mathbf{U} &&\sim \text{matrix normal}(\mathbf{M} = \mathbf{1}\boldsymbol{\mu}_U^T, \mathbf{I}_N, \Psi_U) \\
&(3)\ \mathbf{V} &&\sim \text{matrix normal}(\mathbf{M} = \mathbf{1}\boldsymbol{\mu}_V^T, \mathbf{I}_T, \Psi_V) \\
&(4)\ \epsilon_{i,j,t} &&\sim \mathcal{N}(0, \sigma^2).
\end{aligned}
$$

The resulting estimates of node-specific latent variables recover a specific type of similarity between nodes, so that nodes with similar connections profiles are likely to have similar values (Hoff, 2008). If $\mathbf{u}_i$ and $\mathbf{u}_k$ exhibit similar values, they will have similar inner product outcomes with node j's latent position vector $\mathbf{u}_j$, corresponding to the notion of stochastic equivalence in network theory (Wasserman and Faust, 1994). In addition, the generation rule parameter $\mathbf{v}_t$ contains the information on what the distance relationships on each dimension of the $\mathbf{U}$ space reveal about their connection probability. For example, $v_{rt} > 0$ corresponds to the case when a network generation rule for the rth dimension at time t is homophilous (assortative). In words, $v_{rt} > 0$ indicates that two nodes on rth dimension at time t are more likely to be connected if they are located in the same side of the axis and the magnitude of their product is high. Similarly, $v_{rt} < 0$ corresponds to the case when a network generation rule for rth dimension at time t is heterophilous (dissortative), so that nodes located on the opposite sides are more likely to be connected than the ones with same sign.

In this paper, we develop a hidden Markov multilinear tensor model (HMTM) that extends Hoff (2011)'s MTRM. HMTM combines MTRM with a hidden Markov model (HMM) using Bayesian inference. Conventional approaches to dynamic network modeling usually extend a framework of static networks by assuming smooth topological changes over time (Robins and Pattison, 2001; Hanneke et al., 2010; Desmarais and Cranmer, 2012; Snijders et al., 2006, 2010; Westveld and Hoff, 2011; Ward et al., 2013). However, longitudinal network datasets typically show irregular dynamics, implying multiple changes in their data generating parameters. For example, a human brain network study shows that functional connectivity networks between

brain regions are highly non-stationary, consisting of many latent states (Cribben et al., 2013). In the proposed method, HMTM, latent node positions follow a discrete Markov process in which parameters are allowed to change multiple times whenever there is a symptom of fundamental change in the latent traits. As shown in other applications (Baum et al., 1970; Chib, 1998; Robert et al., 2000; Cappe et al., 2005; Scott et al., 2005; Frühwirth-Schnatter, 2006; Teh et al., 2006), the conditional independence assumption in HMM allows us to model each local regime using a well-known parametric model. Suppose $\mathbf{\Theta}$ be a collection of parameters that represent a network generating process of a longitudinal network $\mathcal{Y}$. Then, the model is decomposed as

(5)
$$p(\mathcal{Y}|\mathbf{\Theta}) = \int p(S_1|\mathbf{\Theta})p(\mathbf{Y}_1|S_1,\mathbf{\Theta})\prod_{t=2}^{T}\sum_{m=1}^{M} p(\mathbf{Y}_t|\mathbf{\Theta}_m)\Pr(S_t=m|S_{t-1},\mathbf{\Theta})d\mathbf{S},$$

where $p(\mathbf{Y}_t|\mathbf{\Theta}_m)$ a generative network model at regime m and $\mathbf{S}=(S_1,\ldots,S_T)$ are discrete hidden states. In our model, the duration of hidden state m follows a geometric distribution of $1-p_{mm}$ where p_{mm} is the mth diagonal element of an $M\times M$ transition matrix. The regime change probability can be easily computed using the posterior draws of hidden states (e.g. $\frac{1}{G}\sum_{g=1}^{G}\mathcal{I}(S_t^{(g)}\neq S_{t-1}^{(g)})$).

We introduce a degree correction formulation (Karrer and Newman, 2011; Chaudhuri et al., 2012) to account for degree heterogeneity that may hinder stochastic equivalence trait recovery. We will show that the degree correction method makes a crucial difference in the recovery of group-structures in longitudinal network data, controlling for factors coming from disturbance other than the data generating parameters. We also discuss model diagnostics of HMTMs using the approximate log marginal likelihood (Chib, 1995) and the Watanabe-Akaike Information Criterion (WAIC) (Watanabe, 2010).

Although the proposed method can be used to analyze longitudinal network datasets from various fields, we would like to highlight that HMTM has a strong substantive appeal to the fields where abrupt changes in the structural properties of networks have important substantive implications. For example, many historical questions in the social sciences are deeply related to the identification of distinct periods or critical events (Abbott, 2001; Mahoney and Rueschemeyer, 2003; Pierson, 2004). For this goal, social scientists increasingly resort to HMMs and related models to identify hidden regimes from historical time series data (Hamilton, 1989; Kim and Nelson, 1998; Western and Kleykamp, 2004; Spirling, 2007; Park, 2011; Pang et al.,

2012). While network is one of the most prevalent forms of dataset in social sciences, the lack of methodological tools for the identification of critical regimes prevents scholars from using longitudinal network dataset to answer such important questions in their fields.

We apply our method to international military alliance networks, which is available from the Correlates of War (COW) project (Gibler, 2008). The dataset contains dyadic alliance relationships among countries from 1816 to 2012. Although military alliance networks have been one of most frequently studied network datasets in the social sciences (Maoz, 2009; Warren, 2010; Cranmer, Desmarais and Menninga, 2012; Cranmer, Desmarais and Kirkland, 2012; Jackson and Nei, 2015; Chiba et al., 2015), there has been no study to investigate structural changes in military alliance networks. Our findings reveal that military alliance networks show strongly clustered structures and these structures have changed from time to time, supporting central claims by historical studies. Among others, the structural change found in the major power alliance network around the 1880s was related to the disappearance of a country that could claim to be a "*courtier honnête* (honest broker)" in the international system.

All elements of the methods introduced in the article are provided in `NetworkChange` which is an open-source `R` package.

1.1. *Related Work.* Recently, a variety of approaches for network change point detection emerged (e.g. Guo et al., 2007; Heard et al., 2010; Wang et al., 2014; Cranmer et al., 2014; Cribben and Yu, 2016; Ridder et al., 2016). For example, Cribben and Yu (2016) propose a two-step approach to network change point detection in which cosine distances for the principal eigenvectors of time-specific graph Laplacian matrix are used to find change points given pre-specified significance thresholds. Similarly, Cranmer et al. (2014) pre-tested the existence of breaks in global network statistics within the framework of the TERGM. Although this type of two-step approaches could be useful in learning specific aspects of network evolution, they are inherently unstable and inefficient by understating uncertainties in each estimation step and hence do not provide a principled means to select the number of breaks.

Guo et al. (2007) and Wang et al. (2014) model network change-points by allowing shifts of parameter values in ERGMs. However, both models exhibit computational inefficiency, in which the maximum size of network analyzed was 11 nodes (Guo et al., 2007) and 6 nodes respectively (Wang et al., 2014). Ridder et al. (2016) use stochastic block models (SBM) as a probabilistic model for network generation of each time-specific network and identify the

existence of a single break by comparing the bootstrapped distribution of the log-likelihood ratio between a null model (a SBM without a break) and an alternative (a SBM with a break). However, the asymptotic distribution of a SBM with a break approaches to a mixture of χ^2-distributions and hence a log-likelihood ratio test statistic used by Ridder et al. (2016) does not meet the regularity condition (Drton, 2009).

2. The Proposed Method. Our task of developing a dynamic network model for structural changes must start from the question of "What constitutes structural changes in networks?" On the one hand, one can think of a change in summary statistics of *macro-scopic* network properties, such as average shortest path length or network density as a structural change (Cranmer et al., 2014). On the other hand, a change in the population statistics of *micro-scopic* network properties, such as transitivity or node degree, can be considered as a structural change (Heard et al., 2010; Lung-Yut-Fong et al., 2012; Kolar and Xing, 2012). But global network statistics and local indices cannot fully represent generative processes of dynamic networks as the granularity of the information entailed in such measures is too limited. Instead, studies in network science have paid an increasing amount of attention to *meso-scopic* features of networks (e.g. community structures, stochastic blocks, core-periphery structures) (Nowicki and Snijders, 2001; Newman, 2006; Fortunato, 2010; Tibély et al., 2011; Sporns, 2014). Technically, various approaches for meso-scopic trait discovery locates nodes in a discrete or continuous latent space on the basis of their similarity. In this paper, a structural change in networks is defined as a change in meso-scopic features of networks. We support this claim by using synthetic examples and show that this perspective is effective enough to recover fundamental aspects of changes in network generation.

2.1. *Latent Space Approach with Degree-correction.* We take the latent space approach in which the proximity of nodes in the latent space implies similar connecting profiles (Hoff et al., 2002; Hoff, 2008). The latent space becomes a stochastic equivalence space in terminology of social network analysis (Wasserman and Faust, 1994). In particular, we use a MTRM as a baseline for dynamic network modeling. However, the formulation of MTRM in Equation (1) entails a critical weakness in uncovering *meso-scopic network features* because there is no treatment to account for degree heterogeneity except exogenous covariates ($\mathbf{x}_{i,j,t}$). The implicit assumption in this latent model is that the expected degree of nodes having a similar role (i.e. proximal in the latent space) is identical while the distribution of degree in empirical networks is highly heterogeneous and skewed following power law or expo-

nential distributions (Clauset et al., 2009). This problem is well known in the network science literature and various degree-correction methods have been proposed in the literature of network block discovery (Newman, 2006, 2010; Karrer and Newman, 2011; Chaudhuri et al., 2012; Zhao et al., 2012).

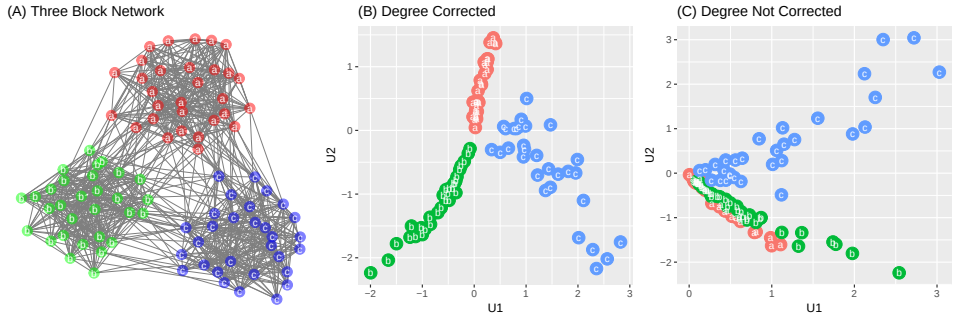


Fig 1: *Degree Correction for Group Structure Discovery: A 90 node undirected network with 3 communities is generated using the within-group link probability of 0.5 and the between-group link probability of 0.05.*

Figure 1 illustrates the problem. We generate an undirected network consisting of 90 nodes and 3 homogeneous groups (*i.e.* blocks) in panel (A). Panel (C) shows recovered latent node positions from an MTRM with a probit link. It is clear that the planted three block structure is not identified in panel (C). In panel (B), the same model is applied after degree-correction. For degree-correction, we used an additive null model (ω_{ijt}), consisting of the principal eigenvalue ($\lambda_t^{princ} = \max(|\lambda(\mathbf{Y}_t)|)$) and its associated eigenvector to control for the baseline expected level of associations among pairs of nodes due to their degrees (Peixoto, 2013):

$$\omega_{ijt} = \lambda_t^{princ} \tilde{\mathbf{u}}_{it} \tilde{\mathbf{u}}_{it}^T, \tag{6}$$

where $\tilde{\mathbf{u}}_{it}$ is the ith row of the associated eigenvector.[2] The planted three group structure is well recovered in panel (B).

[2]Alternatively, one can use a modularity matrix (M_{ijt}):

$$M_{ijt} = y_{ijt} - \frac{k_i k_j}{2m}$$

where $m = \frac{\sum_{i=1}^{N} k_i}{2}$ and k_i is the sum of weights for i (Newman and Girvan, 2004). Both methods are available in `NetworkChange`.

2.2. *Hidden Markov Multilinear Tensor Model.* In a HMTM with k breaks, the probability distribution of a degree-corrected symmetric temporal network data ($\mathbf{B}_t = \mathbf{Y}_t - \mathbf{\Omega}_t$) is modeled as a Markov mixture of k MTRMs in which hidden states move forward (Chib, 1998). As shown by Chib (1998) and Park (2012), multiple change-point problems are equivalent to the estimation of a nonergodic (or forward-moving) HMM, which has advantages in latent state identification and parameter estimation, thanks to the order constraint in latent states. Let us denote S_t as a hidden state variable and $\mathbf{P}$ as a $k+1$ by $k+1$ transition matrix where $p_{i,i}$ is the ith diagonal element of $\mathbf{P}$. Then, a HMTM is

$$\mathbf{B}_t = \mathbf{Y}_t - \mathbf{\Omega}_t \tag{7}$$

$$\mathbf{B}_t = \beta \mathbf{J}_N + \mathbf{U}_{S_t} \mathbf{V}_t \mathbf{U}_{S_t}^T + \mathbf{E}_t \tag{8}$$

$$S_t | S_{t-1}, \mathbf{P} \sim \mathcal{M}arkov(\mathbf{P}, \pi_0)$$

$$p_{i,i} \sim \mathcal{B}eta(a_0, b_0) \tag{9}$$

$$\mathbf{E}_t \sim \mathcal{N}_{N \times N}(\mathbf{0}, \sigma_{S_t}^2 \mathbf{I}_N, \mathbf{I}_N), \tag{10}$$

where π_0 is the initial probability ($\pi_0 = (1, 0, \ldots, 0)$), $\mathbf{J}_N$ is $N \times N$ all-ones matrix. $\beta \mathbf{J}_N$ is added as a constant to stabilize simulation outputs.

For prior distributions and parameter estimation, we follow Hoff (2011)'s hierarchical scheme with two major exceptions. First, we orthogonalize each column of $\mathbf{U}_m$ using the Gram-Schmidt process (Björck, 1996; Guhaniyogi and Dunson, 2015) in each simulation step. Hoff (2011)'s hierarchical scheme centers rows of $\mathbf{U}_m$ around its global mean ($\boldsymbol{\mu}_{u,m}$) using a multivariate normal distribution. This does not guarantee the orthogonality of each latent factor in $\mathbf{U}_m$. The lack of orthogonality makes the model unidentified, causing numerical instability in parameter estimation and model diagnostics (Murphy, 2012; Guhaniyogi and Dunson, 2015).

In addition, we use independent inverse-gamma distributions instead of inverse-Wishart distribution for the prior distribution of a variance parameter ($\Psi_{u,m}, \Psi_{v,m}$):

$$\Psi_u = \begin{pmatrix} \psi_{1,u,m} & \cdots & 0 \\ 0 & \psi_{r,u,m} & 0 \\ 0 & \cdots & \psi_{R,u,m} \end{pmatrix}$$

$$\psi_{r,u,m} \sim \mathcal{IG}\left(\frac{u_0}{2}, \frac{u_1}{2}\right).$$

The use of inverse-Wishart distribution for the prior distribution of a variance parameter ($\Psi_{u,m}, \Psi_{v,m}$) comes at a great cost because choosing informative inverse-Wishart prior distributions for $\Psi_{u,m}$ and $\Psi_{v,m}$ is not easy

(Chung et al., 2015) and a poorly specified inverse-Wishart prior distribution has serious impacts on the marginal likelihood estimation. In our trials, the log posterior inverse-Wishart density of $\Psi_{u,m}$ and $\Psi_{v,m}$ often goes to a negative infinity, failing to impose proper penalties. In HMTM, the off-diagonal covariance of $\mathbf{U}_m$ is constrained to be 0, thanks to the Gram-Schmidt process, and the off-diagonal covariance of $\mathbf{V}$ is close to 0 as $\mathbf{v}_t$ measures time-varying weights of independent $\mathbf{U}_m$. Thus, inverse-gamma distributions resolve a computational issue without a loss of information.

Prior distributions for each row vector of regime specific latent positions ($\mathbf{U}_m$) are multivariate normal distributions conditioned on a variance parameter ($\Psi_{u,m}$): $\{\mathbf{u}_{1,m}, \ldots, \mathbf{u}_{N,m}\} \sim \mathcal{N}_R(\boldsymbol{\mu}_{u,m}, \Psi_{u,m})$ and $\boldsymbol{\mu}_{u,m}|\Psi_{u,m} \sim \mathcal{N}_R(\boldsymbol{\mu}_{0,u^m}, \Psi_u^m)$. Similarly, prior distributions for each row vector of $\mathbf{V}$ are also multivariate normal distributions on a variance parameter (Ψ_v): $\{\mathbf{v}_1, \ldots, \mathbf{v}_T\} \sim \mathcal{N}_R(\boldsymbol{\mu}_v, \Psi_v)$ and $\boldsymbol{\mu}_v|\Psi_v \sim \mathcal{N}_R(\boldsymbol{\mu}_{0,v}, \Psi_v)$. The prior distribution for regime specific error variances (σ_m^2) is an inverse gamma distribution ($\mathcal{IG}\left(\frac{c_0}{2}, \frac{d_0}{2}\right)$). The prior distribution for transition probabilities is the Beta distribution ($\mathcal{B}eta(a, b)$) and the prior distribution for β is a normal distribution ($\mathcal{N}(b_0, B_0)$).

Let $\boldsymbol{\Theta}$ indicate a parameter vector beside hidden states ($\mathbf{S}$) and a transition matrix ($\mathbf{P}$): $\boldsymbol{\Theta} = \{\mathbf{U}, \mathbf{V}, \boldsymbol{\mu}_u, \Psi_u, \boldsymbol{\mu}_v, \Psi_v, \beta, \sigma^2\}$. Let $\boldsymbol{\Theta}_{S_t}$ denote regime-specific $\boldsymbol{\Theta}$ at t. Then, the joint posterior density $p(\boldsymbol{\Theta}, \mathbf{P}, \mathbf{S}|\mathcal{B})$ is

$$
\begin{aligned}
p(\boldsymbol{\Theta}, \mathbf{P}, \mathbf{S}|\mathcal{B}) \quad \propto \quad & \mathcal{N}_{N\times N}(\mathbf{B}_1|\boldsymbol{\Theta}_1) \prod_{t=2}^{T} \Big(\mathcal{N}_{N\times N}(\mathbf{B}_t|\mathcal{B}_{t-1}, \boldsymbol{\Theta}_{S_t}) p(S_t|S_{t-1}, \mathbf{P})\Big) \\
& \prod_{m=1}^{M} \Big(\mathcal{N}_R(\boldsymbol{\mu}_{u,m}|\boldsymbol{\mu}_{0,u^m}, \psi_{.,u,m}) \mathcal{N}_R(\boldsymbol{\mu}_v|\boldsymbol{\mu}_{0,v}, \psi_{.,v,m})\Big) \\
& \prod_{m=1}^{M} \prod_{r=1}^{R} \Big(\mathcal{IG}(\psi_{r,u,m}|u_{0,m}, u_{1,m}) \mathcal{IG}(\psi_{r,v,m}|v_{0,m}, v_{1,m})\Big) \\
& \prod_{m=1}^{M} \Big(\mathcal{IG}(\sigma_m^2|c_0, d_0) \mathcal{B}eta(p_{mm}|a, b)\Big) \mathcal{N}(\beta|b_0, B_0)
\end{aligned}
$$

where $\mathcal{B}_{t-1} = (\mathbf{B}_1, \ldots, \mathbf{B}_{t-1})$. Using the conditional independence we decompose the joint posterior distribution into three blocks and marginalize conditional distributions (Liu et al., 1994; van Dyk and Park, 2008):

$$
p(\boldsymbol{\Theta}, \mathbf{P}, \mathbf{S}|\mathcal{B}) \quad = \quad \underbrace{p(\boldsymbol{\Theta}|\mathcal{B}, \mathbf{P}, \mathbf{S})}_{\text{Part 1}} \underbrace{p(\mathbf{P}|\mathcal{B}, \mathbf{S})}_{\text{Part 2}} \underbrace{p(\mathbf{S}|\mathcal{B})}_{\text{Part 3}}.
$$

We discuss the details of the sampling algorithm in the appendix.

2.3. *Assessing Model Uncertainty using Marginal Likelihood and WAIC.* In this article, we consider two fully Bayesian measures of model uncertainty. Due to the symmetry of undirected temporal networks, we use the upper triangular array of $\mathcal{B}$ ($\mathcal{B}^{\text{upper}}$) for the computation of model uncertainty. The first measure is the approximate log marginal likelihood method using the candidate's estimator (Chib, 1995). Main advantages of the approximate log marginal likelihood are its direct connection with Bayes' theorem and its consistency when models are well identified and MCMC chains converge to the target distribution. A disadvantage of the approximate log marginal likelihood is its computational cost arising from additional MCMC runs at each Gibbs sampling block. Using the Rao-Blackwell approximation, the approximate log marginal likelihood of HMTM with M numbers of latent states ($\mathcal{M}_M$) can be computed as follows:

$$\begin{aligned}\log \hat{p}(\mathcal{B}^{\text{upper}}|\mathcal{M}_M) \quad = \quad & \underbrace{\log p(\mathcal{B}^{\text{upper}}|\boldsymbol{\mu}^*_u, \psi^*_{\cdot,u}, \boldsymbol{\mu}^*_v, \psi^*_{\cdot,v}, \beta^*, \sigma^{2*}, \mathbf{P}^*, \mathcal{M}_M)}_{\text{the log likelihood}} \\ & + \underbrace{\sum_{m=1}^{M} \log p(\boldsymbol{\mu}^*_{u,m}, \psi^*_{\cdot,u,m}, \boldsymbol{\mu}^*_{v,m}, \psi^*_{\cdot,v,m}, \beta^*, \sigma^{2*}_m, p^*_{m,m}|\mathcal{M}_M)}_{\text{the log prior density of posterior means}} \\ & - \underbrace{\sum_{m=1}^{M} \log p(\boldsymbol{\mu}^*_{u,m}, \psi^*_{\cdot,u,m}, \boldsymbol{\mu}^*_{v,m}, \psi^*_{\cdot,v,m}, \beta^*, \sigma^{2*}_m, p^*_{m,m}|\mathcal{B}^{\text{upper}}, \mathcal{M}_M)}_{\text{the log posterior density of posterior means}}\end{aligned}$$

where $\{\boldsymbol{\mu}^*_u, \psi^*_{\cdot,u}, \boldsymbol{\mu}^*_v, \psi^*_{\cdot,v}, \beta^*, \sigma^{2*}, \mathbf{P}^*\}$ are posterior means of MCMC outputs. The log likelihood is computed by summing log predictive density values evaluated at posterior means across all states and over all upper triangular array elements as follows:

$$\begin{aligned}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=i+1}^{N-1}\sum_{m=1}^{M} & p(b_{i,j,t}|\mathcal{B}^{\text{upper}}_{t-1}, \boldsymbol{\mu}^*_{u,m}, \psi^*_{\cdot,u,m}, \boldsymbol{\mu}^*_{v,m}, \psi^*_{\cdot,v,m}, \beta^*, \sigma^{2*}_m, \mathbf{P}^*, S_t = m, \mathcal{M}_M) \\ & p(S_t = m|\mathcal{B}^{\text{upper}}_{t-1}, \boldsymbol{\mu}^*_{u,m}, \psi^*_{\cdot,u,m}, \boldsymbol{\mu}^*_{v,m}, \psi^*_{\cdot,v,m}, \beta^*, \sigma^{2*}_m, \mathbf{P}^*_m, \mathcal{M}_M).\end{aligned}$$

The second measure of model uncertainty is WAIC (Watanabe, 2010). WAIC approximates the expected log pointwise predictive density by subtracting a bias for the effective number of parameters from the sum of log pointwise predictive density. WAIC approximates leave-one-out cross validation (LOO-CV) in singular models and hence can serve as a metric for out-of-sample predictive accuracy of HMTM (Gelman et al., 2014). Predictive accuracy is a good standard for detecting the number of breaks because

overfitting is a major concern in analysis using mixture models and HMMs. Also, the cost of computation is very low as WAIC is computed from MCMC outputs. Note that WAIC of HMTM partitions the data into T pieces of conditional density, and hence the one-step ahead prediction density must be used. Using the formula suggested by Gelman et al. (2014), WAIC for HMTM with M number of latent states ($\mathcal{M}_M$) is

$$\text{WAIC}_{\mathcal{M}_M} = -2\Bigg(\underbrace{\sum_{t=1}^{T}\log\left[\frac{1}{G}\sum_{g=1}^{G}p(\mathcal{B}_t^{\text{upper}}|\boldsymbol{\mu}_u^{(g)},\psi_{.,u}^{(g)},\boldsymbol{\mu}_v^{(g)},\psi_{.,v}^{(g)},\beta^{(g)},\sigma^{2,(g)},\mathbf{P}^{(g)},\mathcal{M}_M)\right]}_{\text{the expected log pointwise predictive density}} - \underbrace{\sum_{t=1}^{T}V_{g=1}^{G}\left[\log p(\mathcal{B}_t^{\text{upper}}|\boldsymbol{\mu}_u^{(g)},\psi_{.,u}^{(g)},\boldsymbol{\mu}_v^{(g)},\psi_{.,v}^{(g)},\beta^{(g)},\sigma^{2,(g)},\mathbf{P}^{(g)},\mathcal{M}_M)\right]}_{\text{bias for the effective number of parameters}}\Bigg)$$

where G is the MCMC simulation size, $V[\cdot]$ indicates a variance, and $\boldsymbol{\Theta}^{(g)}, \mathbf{P}^{(g)}$ are the gth simulated outputs. Throughout the paper, we report the approximate log marginal likelihood in the deviance scale by multiplying -2 to $\log\hat{p}(\mathcal{B}^{\text{upper}}|\mathcal{M}_M)$ for easy comparison with WAIC following the advice of Gelman et al. (2014): The smaller the deviance, the better the accuracy.

3. Simulation Studies. In this section, we check whether the proposed method correctly recovers the number of planted breaks and the type of planted network changes using simulated data. We test the proposed method using five different settings: (1) no break, (2) a single block-merging break, (3) a single block-splitting break, (4) a block-merging break followed by a block-splitting break, and (5) a block-splitting break followed by a block-merging break. We only report the results for (1) no break, (2) a single block-merging break, and (4) a block-merging break followed by a block-splitting break to save space.[3]

Blocks in simulated data were generated by an assortative rule in which nodes belonging to the same block had a higher link probability ($p_{in} = 0.5$) than nodes belonging to different blocks ($p_{out} = 0.05$). In the block merging examples, two blocks were merged so that the tie formation probability between the members of the two blocks changed from p_{out} to p_{in}. In the block splitting examples, an existing block split into two equal size blocks so that the connection probability between the members of the newly generated blocks became p_{out} from p_{in}. The length of time layers was 40. The planted break occurred at $t = 20$ in the case of the single break examples and $t = 10$

[3]The other results are reported in the supplementary information.

and $t = 30$ in the case of two breaks. We fit four different HMTMs from no break ($\mathcal{M}_0$) to three breaks ($\mathcal{M}_3$) and compare their model diagnostics, recovered latent spaces, and time-varying network generation rules. From Table 1 to Table 3, the first column shows the number of imposed breaks and their model-fits. The columns in the middle display estimated regime-specific latent spaces for each model. Note that there are $k + 1$ regimes for k break model: $k = 0, \ldots 3$. The last column shows estimated time-varying network generation rule parameters for $R = 2$.

Table 1 summarizes the results from the no break case. In the no break case, the longitudinal network was generated from a two block structure, consisting of a 10 node block and a 20 node block respectively, and the underlying block structure remains constant. The reading of the results starts from model diagnostics in the first column. While the approximate log marginal likelihood incorrectly favors the one break model ($\mathcal{M}_1$), WAIC correctly shows that the no break model ($\mathcal{M}_0$) fits the data best. The no break model ($\mathcal{M}_0$) in Table 1 correctly recovers the latent block structure. Note that one dimensional information is sufficient to identify the planted block structure and hence the second dimensional network generation rule (v_2) oscillates around zero throughout time.

In fact, the approximate log marginal likelihood consistently favored over-identified HMTMs in our simulation studies. The source of over-identification is the existence of single observation states (latent states consisting of a single observation). If we fit HMTMs with more breaks than necessary, the state sampler often assigns a single observation to redundant states. In high dimensional time series data, these single observation states produce extremely large log likelihoods. A similar problem has been noticed in finite mixture models with singular components (Hartigan, 1985; Bishop, 2006). Chib (1995)'s algorithm is based on the summation of log likelihoods evaluated at posterior means and hence sensitive to the presence of single observation states in high dimensional time series data. In contrast, WAIC relies on the log pointwise predictive density as a measure of the goodness of the fit and its variance as a penalty. Since the log pointwise predictive density is averaged over the entire MCMC scan ($\frac{1}{G}\sum_{g=1}^{G} p(\mathcal{B}_t^{\text{upper}}|\mathbf{\Theta}^{(g)}, \mathbf{P}^{(g)}, \mathcal{M}_M)$), it is less sensitive to singular components in high dimensional mixture models like HMTM. This is why WAIC outperforms in the break number detection in the context of HMTM.[4]

Table 2 summarizes the simulation results from the block-splitting change example. The ground truth is the single break HMTM in the third row. Again, WAIC correctly identifies the single break model as the best-fitting

[4]We discussed this issue in more details in the supplementary information.

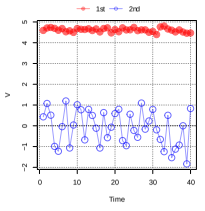
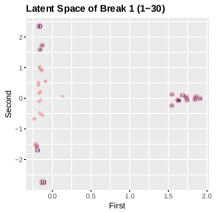
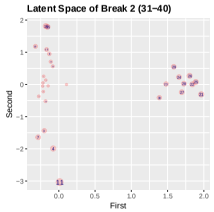
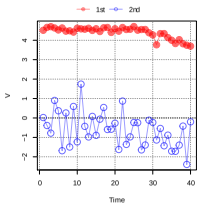
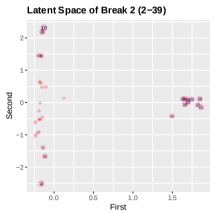
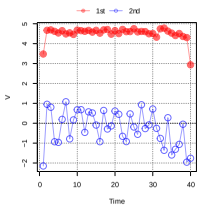
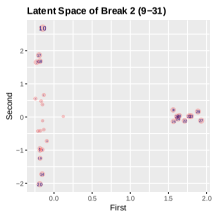
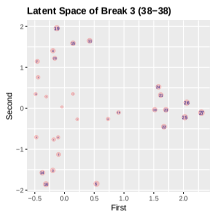
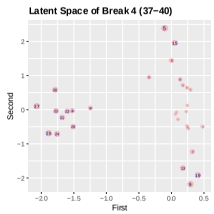
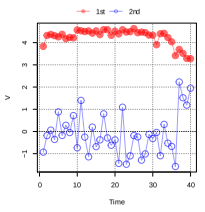
Model Fit	Latent Space ($\mathbf{U}_m$) Changes				Generation Rule ($\mathbf{v}_t$)
	Regime 1	Regime 2	Regime 3	Regime 4	
Break = 0 WAIC = 14573 $-2 \times \log p(\mathcal{B}\|\mathcal{M}_0) = 14435$ log likelihood = -7191					
Break = 1 WAIC = 14589 $-2 \times \log p(\mathcal{B}\|\mathcal{M}_1) = 14404$ log likelihood = -7155					
Break = 2 WAIC = 14653 $-2 \times \log p(\mathcal{B}\|\mathcal{M}_2) = 14514$ log likelihood = -7204					
Break = 3 WAIC = 14739 $-2 \times \log p(\mathcal{B}\|\mathcal{M}_3) = 14564$ log likelihood = -7229					

Table 1: Simulation Analysis of Longitudinal Network Data with a Constant Block Structure. The ground truth is no break and the underlying block structure is a two-block network.

model while the approximate log marginal likelihood favors the three break HMTM. The one break model ($\mathcal{M}_1$) in Table 2 correctly recovers the block-splitting latent change and the second dimensional network generation rule (v_2) jumps from 0 to a positive number in the middle as the number of blocks increases from 2 to 3.

Next, we check whether HMTM correctly recovers more complicated planted network changes. In this example, we generated a synthetic longitudinal network dataset with two different changes. The first break is defined by a block-merging change at $t = 11$ and the second break is defined by a block-splitting change at $t = 31$. Table 3 compares the simulation results from four competing models ($\mathcal{M}_0$ to $\mathcal{M}_3$). Again, WAIC correctly detects $\mathcal{M}_2$ as the best-fitting model while the approximate log likelihood favors $\mathcal{M}_3$. The two break HMTM correctly recovers the two underlying changes at $t = 11$ (block merging) and $t = 31$ (block splitting). Changes in the relative size of network generation rules (the last column) inform us the types of changes underlying network structures go through. For example, when the number of blocks changes from 2 to 3 in the transition to Regime 3, v_2 returns to its previous level at Regime 1.

Overall, our simulation results show that the network break number detection using WAIC is highly effective for HMTM in correctly identifying the number of true breaks. Also, the simulation results show that HMTM successfully recovers various types of block structure changes via shifts in the state-specific latent node positions.

4. Applications. The structure of military alliance networks reflects the distribution of military power among coalitions of states, which changes over time in response to exogenous shocks to the international system or endogenous network dynamics. However, there has been no study that investigates changes in coalition structures of military alliance networks over time. A main reason is the lack of statistical methods that model structural changes in longitudinal network data. Another reason is the changing node set across time. That is, states often disappear, newly emerge, or are split into different states. To resolve the changing node set problem, we limit our observations to "major powers," a subset of countries that have played a dominant role in shaping the structure of the international system and tend to last for a long time. We follow COW dataset's coding of "major powers" and include nine countries (Austria-Hungary, China, France, Germany, Italy, Japan, Russia, the United Kingdom, and the United States) in the

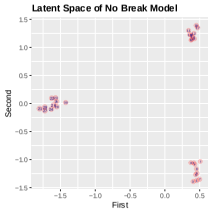
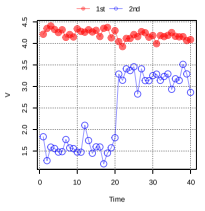
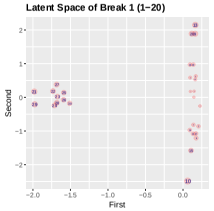
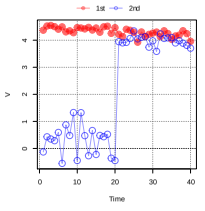
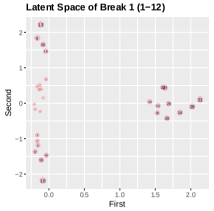
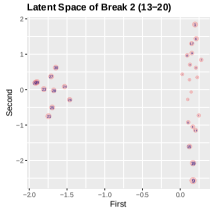
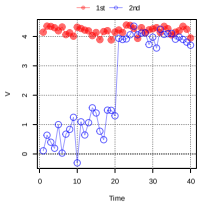
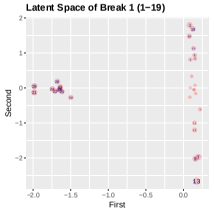
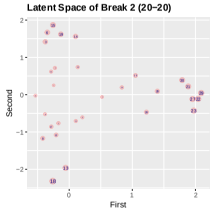
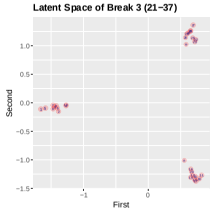
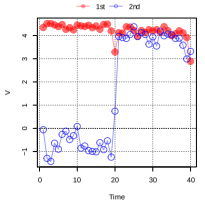
Model Fit	Latent Space ($\mathbf{U}_m$) Changes				Generation Rule ($\mathbf{v}_t$)
	Regime 1	Regime 2	Regime 3	Regime 4	
Break = 0 WAIC = 13744 $-2 \times \log p(\mathcal{B}\|\mathcal{M}_0) = 13648$ log likelihood = -6796					
Break = 1 WAIC = 13016 $-2 \times \log p(\mathcal{B}\|\mathcal{M}_1) = 12922$ log likelihood = -6405					
Break = 2 WAIC = 13067 $-2 \times \log p(\mathcal{B}\|\mathcal{M}_2) = 12929$ log likelihood = -6393					
Break = 3 WAIC = 13064 $-2 \times \log p(\mathcal{B}\|\mathcal{M}_3) = 12877$ log likelihood = -6369					

Table 2: Simulation Analysis of Longitudinal Network Data with a Block-splitting Structure. The ground truth is one break and the underlying block structure changes from a two block structure to a three block structure in the middle.

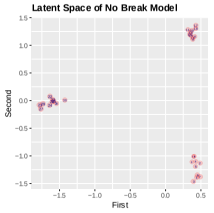
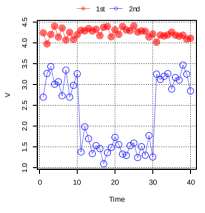
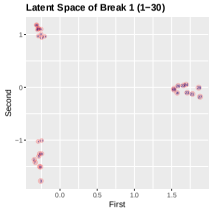
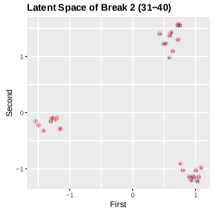
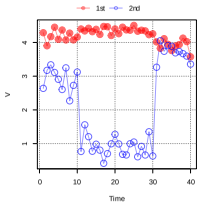
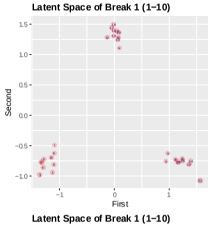
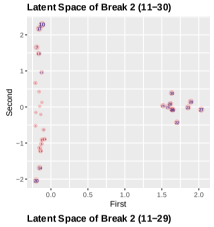
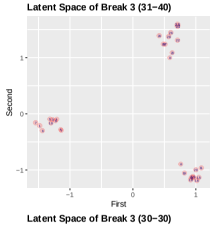
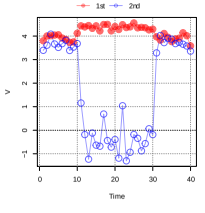
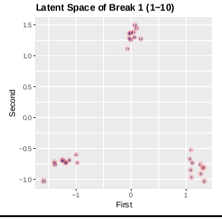
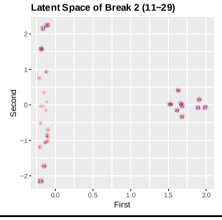
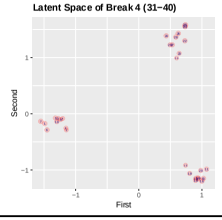
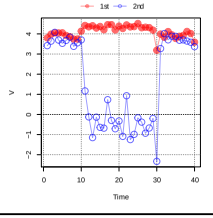
Model Fit	Latent Space ($\mathbf{U}_m$) Changes				Generation Rule ($\mathbf{v}_t$)
	Regime 1	Regime 2	Regime 3	Regime 4	
Break = 0 WAIC = 13727 $-2 \times \log p(\mathcal{B}\|\mathcal{M}_0) = 13630$ log likelihood = -6787					
Break = 1 WAIC = 13583 $-2 \times \log p(\mathcal{B}\|\mathcal{M}_1) = 13431$ log likelihood = -6662					
Break = 2 WAIC = 13202 $-2 \times \log p(\mathcal{B}\|\mathcal{M}_2) = 13043$ log likelihood = -6446					
Break = 3 WAIC = 13203 $-2 \times \log p(\mathcal{B}\|\mathcal{M}_3) = 13028$ log likelihood = -6428					

Table 3: Simulation Analysis of Longitudinal Network Data with a Block-splitting Structure. The ground truth is two breaks and the underlying block structure changes from a two block structure to a three block structure right after $t = 10$ and from a three block structure to a two block structure right after $t = 30$. The ground truth is two breaks.

analysis.[5]

TABLE 4
Bayesian Model Diagnostics of Military Alliance Network Changes Among Major Powers, 1816 - 2012: Geweke's z-score is the test for equality of the means of regime-dependent σ^2 for the first 10 % and last 50% part of MCMC samples. Reported results are from 1,000 MCMC runs after 1,000 burn-in.

	Model					
	$\mathcal{M}_0$	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_5$
$-2 * \log p(\mathcal{B}\|\mathcal{M}_k)$	2673	1339	966	842	475	755
WAIC	2819	1714	1312	1325	1224	1827
Geweke's z-score (σ^2_m)	1.03	0.20	0.68	2.31	1.23	39.19

Table 4 shows the model diagnostic results for the MTRM and five HMTMs fitted to the major power alliance data set. We report Geweke's z-score for σ^2_m for convergence diagnostics in addition to WAIC and the approximate log marginal likelihood. HMTM with four breaks is favored both by WAIC and the approximate log marginal likelihood. Models with more than four breaks show strong signs of nonconvergence due to the existence of redundant states. HMTM with three breaks also shows a sign of nonconvergene (Geweke's z-score = 2.31), which is caused by a fewer number of hidden states than necessary.

Figure 2 visualizes changes in latent node positions of major powers (top) and changing patterns of the major-power network topology (bottom) from the four break HMTM.[6] Regime-specific network generation rule parameters v_{rt} are reported in axis labels. Several substantive findings are noteworthy.

First, the findings of the first two regimes clearly demonstrate the centrality of Austria-Hungary, connecting groups of major powers. Historians call 33 years after the end of Napoleonic Wars "the age of Metternich," highlighting the central role played by Chancellor of Austria-Hungary in the European balance of power system after Napoleonic Wars (Rothenberg, 1968). The second panel of Figure 2 clearly shows how Austria-Hungary, a weaker state than other major powers, was able to hold a diplomatic power over other major powers. The position of Austria-Hungary in the major power alliance network is highly critical in the sense that the removal of Austria-Hungary would have made the major power alliance network completely

[5]In supplementary material, we provide the results of analysis for the extended set of 104 countries during the post-war period.

[6]All network diagrams are drawn using a Fruchterman-Reingold layout, which locates nodes with more connections and short topological distance in proximal locations, for the better visibility of the state labels.

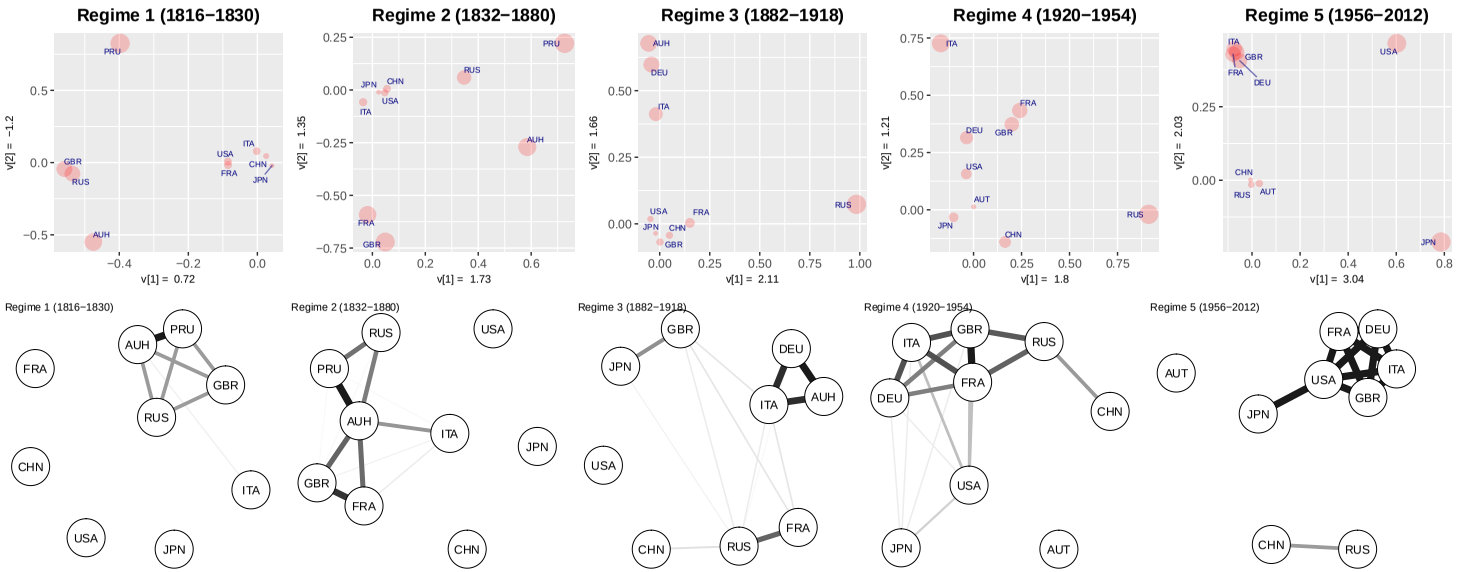


Fig 2: *Changing Node Positions and Network Topology of Military Alliance Networks Among Major Powers, 1816 - 2012: Regime averages of v_t values for each dimension are reported in the axis (top panel). Line widths (bottom panel) are proportional to the duration of alliance links.*

disconnected. In the language of social network analysis, Austria-Hungary filled a "structural hole" in the major power alliance network at the time, playing the role of broker (Burt, 2005, 2009; Stovel and Shaw, 2012). However, the distinct role of Austria-Hungary as broker appeared most clearly during Regime 2 (1832-1880), which only partially overlaps with "the age of Metternich" (1815-1848).

Second, the findings of the HMTM clearly indicates that the social space of the major power alliance network has gone through a structural change around 1880. Reading from historical texts, the origin of the structural change can be attributed to the Dual Alliance between Germany and Austria-Hungary in 1879 and a sequence of alliances that followed it (Snyder, 1997; Vermeiren, 2016). First, Russia formed alliances with Germany and Austria-Hungary (Three Emperors' Alliance) in 1881. Then, Italy joined Germany and Austria-Hungary (Triple Alliance) in 1882. France, a long-time rival of Germany, formed an alliance with Russia in 1894 to check Germany and Austria-Hungary. In this process, an important cleavage in the alliance networks emerged. The network diagram of the third regime (bottom-center) shows two emerging clusters (Austria-Hungary, Germany, Italy vs. France, Russia, and the United Kingdom), which led these countries to World War I and World War II. As a result, after 1882, there has been no country that could serve as a "*courtier honnête* (honest broker)."[7]

Last, both v_{1t} and v_{2t} continue to be positive except for Regime 1, indicating a strong homophily in major power alliance networks.

The analysis of the major power alliance networks over two centuries using HMTMs clearly demonstrates several advantages of HMTMs in longitudinal network data analysis. We strongly believe that HMTMs can help applied researchers in various fields detect and analyze structural changes and related quantities of interest (e.g. regime-dependent block structures and time-varying network generation rules) in many substantively important questions.

[7] A shrewd European historian succinctly summarizes it as follows:

> It was not until Metternich that a statesman appeared who had not only internalized the concept [the balance of power] but was given the opportunity to create a new international structure that explicitly embodied it. His less perceptive successors allowed it to collapse. Bismarck recreated it, although on a far less stable basis. Again his successors allowed it to collapse. The First World War came about not because of the unstable power balance created by competing alliances ... but because the German Empire was no longer interested in maintaining a power balance (Howard, 1994, 134).

5. Concluding Remarks. In this article, we presented HMTM as a statistical method to detect and analyze changes in structural properties of longitudinal network data. The proposed method has several advantages over existing dynamic network models.

First, the proposed method combines a highly flexible generative model of multilayer networks (MTRM) with a HMM, which has proved to be an effective tool to model irregular dynamics in temporal data. This formulation is flexible enough to accommodate a variety of network representations such as graph Laplacian (Rohe et al., 2011) and motif Laplacian (Benson et al., 2016) as an input data format. Our simulation studies showed that our generative approach is a powerful tool to detect and analyze diverse types of network changes.

Second, the Bayesian inference of HMTM enables researchers to identify the number of network changes in a principled way. Our simulation studies show that WAIC correctly identifies the number of breaks and the type of network changes in all tests while the approximate log marginal likelihood consistently favor overfitted models.

Finally, HMTM provides an important tool to investigate changes in meso-scale structures of longitudinal network data. Meso-scale structural changes are important quantities that reflect fundamental changes in the network generating process, that are unable to be captured by local network properties or global summary indices.

While we only consider undirected networks, our model can be extended to analyze other types of longitudinal network data consisting of directed networks or bipartite networks using a singular value decomposition-based framework (De Lathauwer et al., 2000; Hoff, 2007) and the hierarchical multilinear framework Hoff (2011) in general. Also, a hierarchical Dirichlet process prior can be used to endogenously detect the number of breaks (Beal et al., 2002; Ko et al., 2015; Teh et al., 2006; Fox et al., 2011). Another interesting extension of HMTM is the inclusion of nodal covariates (Volfovsky and Hoff, 2015) or covariates for network effects, where the latent space formulation may serve as an instrument to control for unobserved heterogeneous effects on tie formation.

Appendix 1. MCMC Algorithm for Hidden Markov Tensor Model.

For each t layer, generate $\mathbf{B}_t = \mathbf{Y}_t - \mathbf{\Omega}_t$ by choosing a null model ($\mathbf{\Omega}_t$).
Set the total number of changepoints M and initialize $(\mathbf{U}, \boldsymbol{\mu}_u, \Psi_v, \mathbf{V}, \boldsymbol{\mu}_v, \Psi_v, \beta, \sigma^2, \mathbf{S}, \mathbf{P})$.

Part 1

Step 1 The sampling of regime specific $\mathbf{U}, \boldsymbol{\mu}, \Psi_u$ consists of the following three steps for

each regime m. Let $\Psi_u = \begin{pmatrix} \psi_{1,u,m} & \cdots & 0 \\ 0 & \psi_{r,u,m} & 0 \\ 0 & \cdots & \psi_{R,u,m} \end{pmatrix}$.

1. $p(\psi_{r,u,m}|\mathcal{B},\mathbf{P},\mathbf{S},\boldsymbol{\Theta}^{-\Psi_{u,m}}) \propto \mathcal{IG}\left(\frac{u_0+N}{2}, \frac{\mathbf{U}_{r,m}^T\mathbf{U}_{r,m}+u_1}{2}\right)$.
2. $p(\boldsymbol{\mu}_{u,m}|\mathcal{B},\mathbf{P},\mathbf{S},\boldsymbol{\Theta}^{-\boldsymbol{\mu}_{u,m}}) \propto$ multivariate normal$(\mathbf{U}_m^T\mathbf{1}/(N+1), \Psi_{u,m}/(N+1))$.
3. $p(\mathbf{U}_m|\mathcal{B},\mathbf{P},\mathbf{S},\boldsymbol{\Theta}^{-\mathbf{U}_m}) \propto$ matrix normal$_{N\times R}(\tilde{\mathbf{M}}_{u,m}, \mathbf{I}_N, \tilde{\Psi}_{u,m})$ where

$$\begin{aligned} \tilde{\Psi}_{u,m} &= (\mathbf{Q}_{u,m}/\sigma_m^2 + \Psi_{u,m}^{-1})^{-1} \\ \tilde{\mathbf{M}}_{u,m} &= (\mathbf{L}_{u,m}/\sigma_m^2 + \mathbf{1}\boldsymbol{\mu}_{u,m}^T\Psi_{u,m}^{-1})\tilde{\Psi}_{u,m} \\ \mathbf{Q}_{u,m} &= (\mathbf{U}_m^T\mathbf{U}_m)\circ(\mathbf{V}_m^T\mathbf{V}_m) \\ \mathbf{L}_{u,m} &= \sum_{j,t:\ t\in S_t=m} b_{\cdot,j,t}\otimes(\mathbf{U}_{m,j,\cdot}\circ\mathbf{V}_{m,t,\cdot}) \end{aligned}$$

4. Orthogonalize $\mathbf{U}_m$ using the Gram-Schmidt algorithm.

Step 2 The sampling of $\mathbf{V}, \boldsymbol{\mu}_v, \Psi_v$ is done for each regime. Let $\Psi_v = \begin{pmatrix} \psi_{1,v,m} & \cdots & 0 \\ 0 & \psi_{r,v,m} & 0 \\ 0 & \cdots & \psi_{R,v,m} \end{pmatrix}$.

1. $p(\psi_{r,v,m}|\mathcal{B},\mathbf{P},\mathbf{S},\boldsymbol{\Theta}^{-\Psi_{v,m}}) \propto \mathcal{IG}\left(\frac{v_0+T}{2}, \frac{\mathbf{V}_{r,m}^T\mathbf{V}_{r,m}+v_1}{2}\right)$.
2. $p(\boldsymbol{\mu}_{v,m}|\mathcal{B},\mathbf{P},\mathbf{S},\boldsymbol{\Theta}^{-\boldsymbol{\mu}_{v,m}}) \propto$ multivariate normal$(\mathbf{V}_m^T\mathbf{1}/(T_m+1), \Psi_{v,m}/(T_m+1))$.
3. $p(\mathbf{V}_m|\mathcal{B},\mathbf{P},\mathbf{S},\boldsymbol{\Theta}^{-\mathbf{V}_m}) \propto$ matrix normal$_{T_m\times R}(\tilde{\mathbf{M}}_{v,m}, \mathbf{I}_{T_m}, \tilde{\Psi}_{v,m})$ where

$$\begin{aligned} \tilde{\Psi}_{v,m} &= (\mathbf{Q}_{v,m}/\sigma_m^2 + \Psi_{v,m}^{-1})^{-1} \\ \tilde{\mathbf{M}}_{v,m} &= (\mathbf{L}_{v,m}/\sigma_m^2 + \mathbf{1}\boldsymbol{\mu}_{v,m}^{T_m}\Psi_{v,m}^{-1})\tilde{\Psi}_{v,m} \\ \mathbf{Q}_{v,m} &= (\mathbf{U}_m^T\mathbf{U}_m)\circ(\mathbf{U}_m^T\mathbf{U}_m) \\ \mathbf{L}_{v,m} &= \sum_{i,j} b_{i,j,\cdot}\otimes(\mathbf{U}_{m,i,\cdot}\circ\mathbf{U}_{m,j,\cdot}) \end{aligned}$$

Step 3 The sampling of β from $\mathcal{N}(b_1, B_1)$ where

$$\begin{aligned} B_1 &= (B_0^{-1} + \sum_{m=1}^{M}\sigma_m^{-2}N^2\mathbf{1}(\mathbf{S}=m))^{-1} \\ b_1 &= B_1\times\left(B_0^{-1}b_0 + \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} b_{i,j,t} - \mu_{i,j,t}\right). \end{aligned}$$

$\mathbf{1}(\mathbf{S}=m)$ is the number of time units allocated to state m and $\mu_{i,j,t}$ is an element of $\mathbf{U}_{S_t}\boldsymbol{\Lambda}_t\mathbf{U}_{S_t}^T$.

Step 4 The sampling of σ_m^2 from $\mathcal{IG}\left(\frac{c_0+N_m\cdot N_m\cdot T_m}{2}, \frac{d_0+\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} b_{i,j,t}-\beta-\mu_{i,j,t}}{2}\right)$.

Part 2

Step 5 Sample $\mathbf{S}$ recursively using Chib (1998)'s algorithm. The joint conditional distribution of the latent states $p(S_0, \ldots, S_T|\mathbf{\Theta}, \mathcal{B}, \mathbf{P})$ can be written as the product of T numbers of independent conditional distributions:

$$p(S_0, \ldots, S_T|\mathbf{\Theta}, \mathcal{B}, \mathbf{P}) = p(S_T|\mathbf{\Theta}, \mathcal{B}, \mathbf{P}) \ldots p(S_t|\mathbf{S}^{t+1}, \mathbf{\Theta}, \mathcal{B}, \mathbf{P}) \ldots p(S_0|\mathbf{S}^1, \mathbf{\Theta}, \mathcal{B}, \mathbf{P}).$$

Using Bayes' Theorem, Chib (1998) shows that

$$p(S_t|\mathbf{S}^{t+1}, \mathbf{\Theta}, \mathcal{B}, \mathbf{P}) \quad \propto \quad \underbrace{p(S_t|\mathbf{\Theta}, \mathbf{B}_{1:t}, \mathbf{P})}_{\text{State probabilities given all data}} \qquad \underbrace{p(S_{t+1}|S_t, \mathbf{P})}_{\text{Transition probability at } t} \quad .$$

The second part on the right hand side is a one-step ahead transition probability at t, which can be obtained from a sampled transition matrix ($\mathbf{P}$). The first part on the right hand side is state probabilities given all data, which can be simulated via a forward-filtering-backward-sampling algorithm as shown in Chib (1998).

Step 5-1 During the burn-in iterations, if sampled $\mathbf{S}$ has a state with single observation, randomly sample $\mathbf{S}$ with replacement using a pre-chosen perturbation weight ($\mathbf{w}_{\text{perturb}} = (w_1, \ldots, w_M)$).

Part 3: $p(\mathbf{P}|\mathcal{B}, \mathbf{S}, \mathbf{\Theta})$

Step 6 Sample each row of $\mathbf{P}$ from the following Beta distribution:

$$p_{kk} \sim \mathcal{B}eta(a_0 + j_{k,k} - 1, b_0 + j_{k,k+1})$$

where p_{kk} is the probability of staying when the state is k, and $j_{k,k}$ is the number of jumps from state k to k, and $j_{k,k+1}$ is the number of jumps from state k to $k+1$.

Appendix 2. The Approximate Log Marginal Likelihood of a Hidden Markov Tensor Model. The computation of the log posterior density of posterior means requires a careful blocking in a highly parameterized model. In our HMTM, the log posterior density of posterior means is decomposed into seven blocs:

$$\begin{aligned}
\log p(\boldsymbol{\mu}_u^*, \psi_{\cdot,u}^*, \boldsymbol{\mu}_v^*, \psi_{\cdot,v}^*, \beta^*, \sigma^{2*}, \mathbf{P}^*|\mathcal{B}) \quad = \quad & \log p(\boldsymbol{\mu}_u^*|\mathcal{B}) + \sum_{r=1}^{R} \log p(\psi_{r,u}^*|\mathcal{B}, \boldsymbol{\mu}_u^*) \\
& + \log p(\boldsymbol{\mu}_v^*|\mathcal{B}, \boldsymbol{\mu}_u^*, \psi_{\cdot,u}^*) + \sum_{r=1}^{R} \log p(\psi_{r,v}^*|\mathcal{B}, \boldsymbol{\mu}_u^*, \psi_{\cdot,u}^*, \boldsymbol{\mu}_v^*) \\
& + \log p(\beta^*|\mathcal{B}, \boldsymbol{\mu}_u^*, \psi_{\cdot,u}^*, \boldsymbol{\mu}_v^*, \psi_{\cdot,v}^*) \\
& + \log p(\sigma^{2*}|\mathcal{B}, \boldsymbol{\mu}_u^*, \psi_{\cdot,u}^*, \boldsymbol{\mu}_v^*, \psi_{\cdot,v}^*, \beta^*) \\
& + \log p(\mathbf{P}^*|\mathcal{B}, \boldsymbol{\mu}_u^*, \psi_{\cdot,u}^*, \boldsymbol{\mu}_v^*, \psi_{\cdot,v}^*, \beta^*, \sigma^{2*}).
\end{aligned}$$

Let $\mathbf{\Theta}$ indicate a parameter vector beside hidden states ($\mathbf{S}$) and a transition matrix ($\mathbf{P}$): $\mathbf{\Theta} = \{\boldsymbol{\mu}_u, \psi_{\cdot,u}, \boldsymbol{\mu}_v, \psi_{\cdot,v}, \beta, \sigma^2\}$. Let $(\mathbf{\Theta}^*, \mathbf{P}^*)$ be posterior means of $(\mathbf{\Theta}, \mathbf{P})$. Using Chib (1995)'s formula to compute the approximate

log marginal likelihood,

$$\begin{aligned} p(\boldsymbol{\Theta}^*, \mathbf{P}^* | \mathcal{B}) &= \frac{p(\mathcal{B}|\boldsymbol{\Theta}^*, \mathbf{P}^*) p(\boldsymbol{\Theta}^*, \mathbf{P}^*)}{m(\mathcal{B})} \\ m(\mathcal{B}) &= \frac{p(\mathcal{B}|\boldsymbol{\Theta}^*, \mathbf{P}^*) p(\boldsymbol{\Theta}^*, \mathbf{P}^*)}{p(\boldsymbol{\Theta}^*, \mathbf{P}^*|\mathcal{B})} \\ \log m(\mathcal{B}) &= \log p(\mathcal{B}|\boldsymbol{\Theta}^*, \mathbf{P}^*) + \log p(\boldsymbol{\Theta}^*, \mathbf{P}^*) - \log p(\boldsymbol{\Theta}^*, \mathbf{P}^*|\mathcal{B}). \end{aligned}$$

The quantities in the right hand side of Equation (11) can be computed by Chib (1995)'s candidate formula:

Step 1

$$p(\boldsymbol{\mu}_u^*|\mathcal{B}) \approx \int p(\boldsymbol{\mu}_u^*|\mathcal{B}, \psi_{\cdot,u}, \boldsymbol{\mu}_v, \psi_{\cdot,v}, \beta, \sigma^2, \mathbf{P}, \mathbf{S}) dp(\psi_{\cdot,u}, \boldsymbol{\mu}_v, \psi_{\cdot,v}, \beta, \sigma^2, \mathbf{P}, \mathbf{S}|\mathcal{B})$$

Step 2

$$p(\psi_{\cdot,u}^*|\mathcal{B}, \boldsymbol{\mu}_u^*) \approx \int p(\psi_{\cdot,u}^*|\mathcal{B}, \boldsymbol{\mu}_u^*, \boldsymbol{\mu}_v, \psi_{\cdot,v}, \beta, \sigma^2, \mathbf{P}, \mathbf{S}) dp(\boldsymbol{\mu}_v, \psi_{\cdot,v}, \beta, \sigma^2, \mathbf{P}, \mathbf{S}|\mathcal{B})$$

Step 3

$$p(\boldsymbol{\mu}_v^*|\mathcal{B}, \boldsymbol{\mu}_u^*, \psi_{\cdot,u}^*) \approx \int p(\boldsymbol{\mu}_v^*|\mathcal{B}, \boldsymbol{\mu}_u^*, \psi_{\cdot,u}^*, \psi_{\cdot,v}, \beta, \sigma^2, \mathbf{P}, \mathbf{S}) dp(\psi_{\cdot,v}, \beta, \sigma^2, \mathbf{P}, \mathbf{S}|\mathcal{B})$$

Step 4

$$p(\psi_{\cdot,v}^*|\mathcal{B}, \boldsymbol{\mu}_u^*, \psi_{\cdot,u}^*, \boldsymbol{\mu}_v^*) \approx \int p(\psi_{\cdot,v}^*|\mathcal{B}, \boldsymbol{\mu}_u^*, \psi_{\cdot,u}^*, \boldsymbol{\mu}_v^*, \beta, \sigma^2, \mathbf{P}, \mathbf{S}) dp(\beta, \sigma^2, \mathbf{P}, \mathbf{S}|\mathcal{B})$$

Step 5

$$p(\beta^*|\mathcal{B}, \boldsymbol{\mu}_u^*, \psi_{\cdot,u}^*, \boldsymbol{\mu}_v^*, \psi_{\cdot,v}^*) \approx \int p(\beta^*|\mathcal{B}, \boldsymbol{\mu}_u^*, \psi_{\cdot,u}^*, \boldsymbol{\mu}_v^*, \psi_{\cdot,v}^*, \sigma^2, \mathbf{P}, \mathbf{S}) dp(\sigma^2, \mathbf{P}, \mathbf{S}|\mathcal{B})$$

Step 6

$$p(\sigma^{2*}|\mathcal{B}, \boldsymbol{\mu}_u^*, \psi_{\cdot,u}^*, \boldsymbol{\mu}_v^*, \psi_{\cdot,v}^*, \beta^*) \approx \int p(\sigma^{2*}|\mathcal{B}, \boldsymbol{\mu}_u^*, \psi_{\cdot,u}^*, \boldsymbol{\mu}_v^*, \psi_{\cdot,v}^*, \beta^*, \mathbf{P}, \mathbf{S}) dp(\mathbf{P}, \mathbf{S}|\mathcal{B})$$

Step 7

$$p(\mathbf{P}^*|\mathcal{B}, \boldsymbol{\mu}_u^*, \psi_{\cdot,u}^*, \boldsymbol{\mu}_v^*, \psi_{\cdot,v}^*, \beta^*, \sigma^{2*}) \approx \int p(\mathbf{P}^*|\mathcal{B}, \boldsymbol{\mu}_u^*, \psi_{\cdot,u}^*, \boldsymbol{\mu}_v^*, \psi_{\cdot,v}^*, \beta^*, \sigma^{2*}, \mathbf{S}) dp(\mathbf{S}|\mathcal{B})$$

References.

Abbott, A. (2001), *Time Matters: On Theory and Method*, Chicago: University of Chicago Press.

Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970), "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains," *The Annals of Mathematical Statistics*, 41(1), 164–171.

Beal, M. J., Ghahramani, Z., and Rasmussen, C. E. (2002), "The infinite Hidden Markov Model," *Neural Information Processing Systems*, .

Benson, A. R., Gleich, D. F., and Leskovec, J. (2016), "Higher-order organization of complex networks," *Science*, 353(6295), 163–166.

Bishop, C. M. (2006), *Pattern Recognition and Machine Learning* Springer.

Björck, A. (1996), *Numerical Methods for Least Squares Problems* SIAM.

Burt, R. S. (2005), *Brokerage and Closure* Oxford University Press.

Burt, R. S. (2009), *Structural Holes: The Social Structure of Competition* Harvard University Press.

Cappe, O., Moulines, E., and Ryden, T. (2005), *Inference in Hidden Markov Models* Springer-Verlag.

Chaudhuri, K., Graham, F. C., and Tsiatas, A. (2012), Spectral Clustering of Graphs with General Degrees in the Extended Planted Partition Model.„ in *COLT*, Vol. 23, pp. 35–1.

Chib, S. (1995), "Marginal Likelihood From the Gibbs Output," *Journal of the American Statistical Association*, 90(432), 1313–1321.

Chib, S. (1998), "Estimation and comparison of multiple change-point models," *Journal of econometrics*, 86(2), 221–241.

Chiba, D., Johnson, J. C., and Leeds, B. A. (2015), "Careful Commitments: Democratic States and Alliance Design," *Journal of Politics*, 77(4), 968–982.

Chung, Y., Gelman, A., Rabe-Hesketh, S., Liu, J., and Dorie, V. (2015), "Weakly Informative Prior for Point Estimation of Covariance Matrices in Hierarchical Models," *Journal of Educational and Behavioral Statistics*, 40(2), 136–157.

Clauset, A., Shalizi, C. R., and Newman, M. E. (2009), "Power-law distributions in empirical data," *SIAM review*, 51(4), 661–703.

Cranmer, S. J., Desmarais, B., and Kirkland, J. H. (2012), "Toward a Network Theory of Alliance Formation," *International Interactions*, 38, 295–324.

Cranmer, S. J., Desmarais, B., and Menninga, E. J. (2012), "Complex Dependencies in the Alliance Network," *Conflict Management and Peace Science*, 29, 279–313.

Cranmer, S. J., Heinrich, T., and Desmarais, B. A. (2014), "Reciprocity and the Structural Determinants of the International Sanctions Network," *Social Networks*, 36(January), 5–22.

Cribben, I., Wager, T. D., and Lindquist, M. A. (2013), "Detecting functional connectivity change points for single-subject fMRI data," *Frontiers in Computational Neuroscience*, 7, 143.

Cribben, I., and Yu, Y. (2016), "Estimating whole-brain dynamics by using spectral clustering," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, .

De Lathauwer, L., De Moor, B., and Vandewalle, J. (2000), "A multilinear singular value decomposition," *SIAM journal on Matrix Analysis and Applications*, 21(4), 1253–1278.

Desmarais, B. A., and Cranmer, S. J. (2012), "Statistical Mechanics of Networks: Estimation and Uncertainty," *Physica A*, 391(4), 1865–1876.

Drton, M. (2009), "Likelihood ratio tests and singularities," *Ann. Statist.*, 37(2), 979–1012.

Fortunato, S. (2010), "Community detection in graphs," *Physics Reports*, 486(3), 75–174.

Fox, E. B., Sudderth, E. B., Jordan, M. I., and Willsky, A. S. (2011), "A sticky HDP-HMM with application to speaker diarization," *Annals of Applied Statistics*, 5(2 A), 1020–1056.

Frühwirth-Schnatter, S. (2006), *Finite Mixture and Markov Switching Models*, Heidelberg: Springer Verlag.

Gelman, A., Hwang, J., and Vehtari, A. (2014), "Understanding predictive information criteria for Bayesian models," *Statistics and Computing*, 24(6), 997–1016.

Gibler, D. (2008), *International Military Alliances, 1648-2008* CQ Press.

Guhaniyogi, R., and Dunson, D. B. (2015), "Bayesian Compressed Regression," *Journal of the American Statistical Association*, 110(512), 1500–1514.

Guo, F., Hanneke, S., Fu, W., and Xing, E. P. (2007), "Recovering Temporally Rewiring Networks: A Model-based Approach," *Proceedings of the 24 th International Conference on Machine Learning*, pp. 321–328.

Hamilton, J. D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57(2), 357–384.

Hanneke, S., Fu, W., and Xing, E. P. (2010), "Discrete Temporal Models of Social Networks," *Electronic Journal of Statistics*, 4, 585–605.

Hartigan, J. A. (1985), A failure of likelihood asymptotics for normal mixtures„ in *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, eds. L. LeCam, and R. A. Olshen, Vol. 2, Wadsworth Statistics/Probability Series, Belmont, California, pp. 807–810.

Heard, N. A., Weston, D. J., Platanioti, K., and Hand, D. J. (2010), "Bayesian Anomaly Detection Methods for Social Networks," *Annals of Applied Statistics*, 4(2), 645–662.

Hoff, P. (2007), "Model averaging and dimension selection for the singular value decomposition," *Journal of the American Statistical Association*, 102(478), 674–685.

Hoff, P. D. (2008), "Modeling homophily and stochastic equivalence in symmetric relational data," in *Advances in Neural Information Processing Systems 20*, eds. J. Platt, D. Koller, Y. Singer, and S. Roweis Cambridge University Press, pp. 657 – 664.

Hoff, P. D. (2011), "Hierarchical Multilinear Models for Multiway Data," *Computational Statistics & Data Analysis*, 55, 530 – 543.

Hoff, P. D. (2015), "Multilinear tensor regression for longitudinal relational data," *The Annals of Applied Statistics*, 9(3), 1169–1193.

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.

Holme, P., and Saramäki, J. (2012), "Temporal networks," *Physics reports*, 519(3), 97–125.

Howard, M. (1994), "The World According to Henry: From Metternich to Me," *Foreign Affairs*, .

Jackson, M. O., and Nei, S. (2015), "Networks of military alliances, wars, and international trade," *Proceedings of the National Academy of Sciences*, 112(50), 15277–15284.

Karrer, B., and Newman, M. E. (2011), "Stochastic blockmodels and community structure in networks," *Physical Review E*, 83(1), 016107.

Kim, C.-J., and Nelson, C. R. (1998), "Business Cycle Turning Points, A New Coincident Index, and Tests of Duration Dependence Based on a Dynamic Factor Model with Regime-Switching," *The Review of Economics and Statistics*, 80(2), 188–201.

Ko, S. I. M., Chong, T. T. L., and Ghosh, P. (2015), "Dirichlet Process Hidden Markov Multiple Change-point Model," *Bayesian Analysis*, 10(2), 275–296.

Kolar, M., and Xing, E. P. (2012), "Estimating networks with jumps," *Electronic journal of statistics*, 6, 2069.

Liu, J. S., Wong, W. H., and Kong, A. (1994), "Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes," *Biometrika*, 81(1), 27.

Lung-Yut-Fong, A., Lévy-Leduc, C., and Cappé, O. (2012), "Distributed detection/localization of change-points in high-dimensional network traffic data," *Statistics and Computing*, 22(2), 485–496.

Mahoney, J., and Rueschemeyer, D., eds (2003), *Comparative Historical Analysis in the Social Sciences* Cambridge University Press.

Maoz, Z. (2009), "The Effects of Strategic and Economic Interdependence on International Conflict across Levels of Analysis," *American Journal of Political Science*, 53(1), 223–240.

Murphy, K. P. (2012), *Machine Learning: A Probabilistic Perspective* MIT press.

Newman, M. E. (2006), "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, 103(23), 8577–8582.

Newman, M. E. (2010), *Networks: An Introduction* Oxford University Press.

Newman, M. E., and Girvan, M. (2004), "Finding and evaluating community structure in networks," *Physical review E*, 69(2), 026113.

Nowicki, K., and Snijders, T. A. B. (2001), "Estimation and prediction for stochastic blockstructures," *Journal of the American Statistical Association*, 96(455), 1077–1087.

Pang, X., Friedman, B., Martin, A. D., and Quinn, K. M. (2012), "Endogenous Jurisprudential Regimes," *Political Analysis*, 20(4), 417–436.

Park, J. H. (2011), "Analyzing Preference Changes using Hidden Markov Item Response Theory Models," in *Handbook of Markov Chain Monte Carlo; Methods and Applications*, eds. G. Jones, S. Brooks, A. Gelman, and X.-L. Meng CRC Press.

Park, J. H. (2012), "A Unified Method for Dynamic and Cross-Sectional Heterogeneity: Introducing Hidden Markov Panel Models," *American Journal of Political Science*, 56(4), 1040–1054.

Peixoto, T. P. (2013), "Eigenvalue Spectra of Modular Networks," *Physical Review Letters*, 111(9), 098701–5.

Pierson, P. (2004), *Politics in Time: History, Institutions, and Social Analysis*, New Jersey: Princeton University Press.

Ridder, S. D., Vandermarliere, B., and Ryckebusch, J. (2016), "Detection and localization of change points in temporal networks with the aid of stochastic block models," *Journal of Statistical Mechanics: Theory and Experiment*, 2016(11), 113302.

Robert, C. P., Ryden, T., and Titterington, D. M. (2000), "Bayesian Inference in Hidden Markov Models through the Reversible Jump Markov Chain Monte Carlo Method," *Journal of the Royal Statistical Society, Ser. B*, 62(1), 57–75.

Robins, G. L., and Pattison, P. E. (2001), "Random Graph Models for Temporal Processes in Social Networks," *Journal of Mathematical Sociology*, 25(5–41).

Rohe, K., Chatterjee, S., and Yu, B. (2011), "Spectral clustering and the high-dimensional stochastic blockmodel," *The Annals of Statistics*, 39(4), 1878–1915.

Rothenberg, G. E. (1968), "The Austrian Army in the Age of Metternich," *Journal of Modern History*, 40(2), 156–165.

Scott, S. L., James, G. M., and Sugar, C. A. (2005), "Hidden Markov Models for Longitudinal Comparisons," *Journal of the American Statistical Association*, 100(470), 359–369.

Snijders, T. A. B., Steglich, C. E. G., and Schweinberger, M. (2006), "Longitudinal Models in the Behavioral and Related Sciences,".

Snijders, T. A. B., van de Bunt, G. G., and Steglich, C. E. G. (2010), "Introduction to stochastic actor-based models for network dynamics," *Social Networks*, 32(1), 44–60.

Snyder, G. H. (1997), *Alliance Politics* Cornell University Press.

Spirling, A. (2007), ""Turning Points" in Iraq: Reversible Jump Markov Chain Monte Carlo in Political Science," *The American Statistician*, 61(4), 315–320.

Sporns, O. (2014), "Contributions and challenges for network models in cognitive neuro-

science," *Nature Neuroscience*, 17(5), 652–660.
Stovel, K., and Shaw, L. (2012), "Brokerage," *Annual Review of Sociology*, 38, 139–158.
Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006), "Hierarchical Dirichlet Processes," *Journal of the American Statistical Association*, 101(476), 1566–1581.
Tibély, G., Kovanen, L., Karsai, M., Kaski, K., Kertész, J., and Saramäki, J. (2011), "Communities and beyond: Mesoscopic analysis of a large social network with complementary methods," *Phys. Rev. E*, 83, 056125.
van Dyk, D. A., and Park, T. (2008), "Partially Collapsed Gibbs Samplers," *Journal of the American Statistical Association*, 103(482), 790–796.
Vermeiren, J. (2016), *The First World War and German National Identity: The Dual Alliance at War* Cambridge University Press.
Volfovsky, A., and Hoff, P. (2015), "Testing for Nodal Dependence in Relational Data Matrices," *Journal of the American Statistical Association*, 110(511), 1037–1046.
Wang, X., Yuan, K., Hellmayr, C., Liu, W., and Markowetz, F. (2014), "Reconstructing evolving signalling networks by hidden Markov nested effects models," *Annals of Applied Statistics*, 8(1), 448–480.
Ward, M. D., Ahlquist, J. S., and Rozenas, A. (2013), "Gravity's Rainbow: A Dynamic Latent Space Model for the World Trade Network," *Network Science*, 1(1), 95–118.
Warren, C. T. (2010), "The Geometry of Security: Modeling Interstate Alliances as Evolving Networks," *Journal of Peace Research*, 47, 697–709.
Wasserman, S., and Faust, K. (1994), *Social network analysis: Methods and applications*, Vol. 8 Cambridge university press.
Watanabe, S. (2010), "Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory," *Journal of Machine Learning Research*, 11, 3571–3594.
Western, B., and Kleykamp, M. (2004), "A Bayesian Change Point Model for Historical Time Series Analysis," *Political Analysis*, 12(4).
Westveld, A. H., and Hoff, P. D. (2011), "A Mixed Effects Model for Longitudinal Relational and Network Data, with Applications to International Trade and Conflict," *Annals of Applied Statistics*, 5, 843–872.
Zhao, Y., Levina, E., Zhu, J. et al. (2012), "Consistency of community detection in networks under degree-corrected stochastic block models," *The Annals of Statistics*, 40(4), 2266–2292.

Jong Hee Park
Seoul National University

Yunkyu Sohn
Princeton University