Detecting Structural Changes in Longitudinal Network Data: An Application to Changes in Military Alliance Networks, 1816-2012

Jong Hee Park
Seoul National University
email: jongheepark@snu.ac.kr

Yunkyu Sohn
Princeton University
email: ysohn@princeton.edu

May 25, 2017

Abstract

Dynamic modeling of longitudinal network data has been an increasingly important topic in applied research. However, existing methods to dynamic network modeling assume smooth topological changes over time and little work has been done for short, dramatic topological changes. 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. Our simulation results demonstrate that the proposed method correctly detects the number, location, and type of changes in subgroup structures that are often the most important units for understanding the organizations and dynamics of networks. By applying the proposed method to military alliance networks from 1816 to 2012, we identify structural changes in military alliance coalition among nations.

KEYWORDS: Bayesian analysis, network change, hidden Markov model, military alliance, WAIC

1 Introduction

Dynamic modeling of longitudinal network data has been an increasingly important topic in social, biological, and other fields of science as more and more network data are collected for the same node set over time. Inferring a dynamic process from longitudinal network data is a challenging problem because longitudinal network data take a non-conventional form of multidimensional array (or tensor). For example, if we observe relationships between pairs of actors $i = 1, \ldots, N, j = 1, \ldots, N$ for $t = 1, \ldots, T$, we have an $N \times N \times T$ dimension of tensor $\mathcal{Y}$. Dyadic relationships between i and j across temporal layers, $\mathbf{y}_{i,j,\cdot} = (y_{i,j,1}, \ldots, y_{i,j,t})$, are dependent upon the history of both their own and other relationships in $\mathcal{Y}$. The goal of dynamic network modeling is to model the history of these interdependent relationships.

Existing approaches to dynamic network modeling rely on statistical methods that either extend static networks over discrete time intervals or assume smooth topological changes over time. For example, Robins and Pattison (2001), Hanneke et al. (2010), and Desmarais and Cranmer (2012) extend exponential random graph models into the time domain, which are called temporal exponential random graph models (TERGMs). Snijders et al. (2006, 2010) develop stochastic actor-oriented models for network dynamics where tie formation dynamics follow a user-defined objective function reflecting various network effects. Although these methods have some advantages to model network dependence over time, the generative process of dynamic networks in these methods remains as a network, instead of a tensor. One consequence of the matrix representation of dynamic networks is that the type of network relationships that can be represented is limited to one pattern of tie formation, mostly that of homophily (or assortativity). It is difficult to detect changes of network patterns in these

dynamic methods.

The second limitation of the existing methods is related to a temporal mode. Existing approaches to dynamic network modeling assume smooth topological changes over time, failing to account for abrupt, dramatic topological changes in network dynamics. For example, based on the latent space approach to network data, Westveld and Hoff (2011), Ward et al. (2013), and Hoff (2015) develop latent space models for longitudinal network data in which parameters of the latent space follow autoregressive processes. These autoregressive network models assume mean-reverting changes over time, ignoring the possibility of mean-shifting changes in structural properties of networks. In many cases, mean-shifting network changes are substantively more interesting than mean-reverting network changes to applied researchers. For example, studies of brain network show that functional connectivity networks between brain regions are highly non-stationary, including many latent states (Cribben et al., 2013). So far little work has been done for developing a generative network model for short, dramatic changes in structural properties of networks. Guo et al. (2007) and Wang et al. (2014) are two exceptions although these methods are either computationally inefficient (Guo et al., 2007) or limited to genetic data (Wang et al., 2014).

Recently, a group of hybrid methods emerged for the network change point detection (NCPD) (e.g. Heard et al., 2010; Lung-Yut-Fong et al., 2012; Akoglu et al., 2014; Araujo et al., 2014; Koutra et al., 2012; Kolar et al., 2010; Cribben and Yu, 2016; Ridder et al., 2016). Most of NCPD methods combine techniques of network analysis with time series methods to detect change points in longitudinal network data. For example, Cribben and Yu (2016) proposed a two-step approach to network change point detection in which communities are identified by spectral clustering of the Laplacian matrix and then rank-reduced

community information is used to find change points given pre-specified significance thresholds. Ridder et al. (2016) use stochastic block models (SBM) as a probabilistic model for network generation process 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 a SBM with a break. Although these types of two-step approach are useful in some settings, they are inherently unstable and inefficient by understating uncertainties in each estimation step. As a result, one important limitation of these two-step methods is that determining a break number must be done by heuristic methods. For example, 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) for NCPD does not meet the regularity condition (Drton, 2009).

One effective strategy to model network dynamics is to assume that observed networks are independent conditional upon *discrete hidden states*: $\mathbf{S} = (S_1, \ldots, S_T)$, the approach of which, called hidden Markov models (HMMs) or Markov mixture models, has been successfully applied 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). Advantages of HMMs in dynamic network modeling are clear. First, any generative model of networks can be directly used to model local regimes in the HMM framework. Suppose $\mathbf{\Theta}$ be a collection of parameters that represent a network generating process of $\mathbf{Y}_t$. Then, we can denote the network dynamics as

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

where $\mathbf{\Theta}_m$ is parameters for hidden state m.

Second, Bayesian inference of HMMs allows us to make probabilistic inference of many quantities of interest in dynamic network analysis such as the number of changes, the timing of changes, and the characteristics of changes. Most importantly, methods of Bayesian model diagnostics can be used to diagnose network change point numbers and hence applied researchers do not need to resort to heuristic methods or arbitrary thresholds to learn the number of network change points. Also, 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. $\frac{1}{G}\sum_{g=1}^{G}\mathcal{I}(S_t^{(g)} \neq S_{t-1}^{(g)})$).

Last, dynamic network modeling using HMMs have 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).

In this paper, we develop a hidden Markov multilinear tensor model (HMTM) that models change-point processes in longitudinal network data. HMTM combines the multilinear tensor regression model (MTRM) (Hoff, 2011) with a HMM using Bayesian inference. HMTM extends Hoff (2011)'s multilinear tensor regression model into three different directions. First, we transform temporal network data using the degree correction method to apply

the latent space approach to the majority of empirical networks with highly skewed degree distributions. As shown in earlier works (e.g. Karrer and Newman, 2011; Chaudhuri et al., 2012), the degree correction method makes a crucial difference in the recovery of group-structures in network data. Second, HMTM combines MTRM with a nonergodic (or forward-moving) hidden Markov model (Chib, 1998). In HMTM, latent node positions follow a discrete Markov process in which parameters are allowed to change multiple times whenever there is a fundamental change in the latent traits of temporal network data. Third, we provide a method to detect a break number by assessing model uncertainty of HMTMs with a varying number of breaks using the approximate log marginal likelihood (Chib, 1995) and the Watanabe-Akaike Information Criterion (WAIC) (Watanabe, 2010). All elements of the methods introduced in the article are provided in **`NetworkChange`**, which is an open-source **`R`** package.

We applied our method to military alliance networks from 1816 to 2012, which is available from the Correlates of War (COW) project (Gibler, 2008).[1] y_{ijt} is coded 1 if country i and country j have a defense pact, the highest level of military commitment, at year t and $y_{ijt} = 0$ otherwise. Scholars of political science have long emphasized that "how states choose alliance partners will shape the evolution of the international system as a whole" (Walt, 1987, 1). Not surprisingly, military alliance networks have been one of most frequently studied network data sets 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). The findings of HMTM show that military alliance networks show strongly clustered structures and these structures have changed from time to time. The structural

[1]The raw data are available at `http://www.correlatesofwar.org/data-sets/formal-alliances`.

change found in the major power alliance network around the 1880s was related with the disappearance of a country that could claim to be a "*courtier honnête* (honest broker)" in the international system. The structural change found in the postwar alliance network in the 1950s is the emergence of the second dimension by the Non-Aligned Movement countries. The connectivity among the Non-Aligned Movement countries kept growing, replacing the East-West division as early as in the 1950s.

In the next section, we explain the proposed method. Next, we test the proposed method using simulated data. Then, we apply our method to interstate military alliance network data from 1816 to 2012 (Gibler, 2008), to investigate changes in the overall patterns of alliance formation. We conclude this article by summarizing our main findings and discussing future directions.

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 on *macro-scopic* network properties, such as average shortest path length or network density as a structural network change (Cranmer et al., 2014; Cribben and Yu, 2016). 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 network 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 pay increasing attention to meso-scopic features of networks such as community structures, block structures, and core-periphery structures to better understand topological characteristics of networks and their dynamics (Nowicki and Snijders, 2001; Newman, 2006; Fortunato, 2010; Tibély et al., 2011; Sporns, 2014). In this paper, a structural network change is defined as a change in meso-scopic features of networks.

2.1 Latent Space Approach to Meso-scopic Network Features

To represent the network data generating process, we take the latent space approach to network analysis in which the proximity of nodes in the latent space implies that these nodes have similar connected partners in social relationships (Hoff et al., 2002). In social network terminology, the latent space becomes a structural equivalence space (Wasserman and Faust, 1994).

Longitudinal network data are a tensor $\mathcal{Y} \in \mathbb{R}^{N \times N \times T}$, in which an $N \times N$ square matrix $\mathbf{Y}_t$ represents relationships between all pairs of actors $i = 1, \ldots, N, j = 1, \ldots, N$ at t. Hoff (2011) presents a multiplicative latent space model for tensor data where 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). For undirected, unweighted

temporal network data of size $N \times N \times T$, the model is written as follows:

$$\begin{aligned}
\Pr(y_{i,j,t} = 1 | \mathbf{x}_{i,j,t}, \mathbf{u}_i, \mathbf{u}_j, \mathbf{v}_t) &= \mathbf{x}_{i,j,t}\boldsymbol{\beta} + \langle \mathbf{u}_i, \mathbf{v}_t, \mathbf{u}_j \rangle + \epsilon_{i,j,t} && (2) \\
\mathbf{U} &\sim \text{matrix normal}(\mathbf{M} = \mathbf{1}\boldsymbol{\mu}_U^T, \mathbf{I}_N, \Psi_U) && (3) \\
\mathbf{V} &\sim \text{matrix normal}(\mathbf{M} = \mathbf{1}\boldsymbol{\mu}_V^T, \mathbf{I}_T, \Psi_V) && (4) \\
\epsilon_{i,j,t} &\sim \mathcal{N}(0, \sigma^2), && (5)
\end{aligned}$$

where $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_N) \in \mathbb{R}^{R \times N}$ are the R-dimensional latent node positions of N nodes and $\mathbf{v}_t = (v_{1t}, \ldots, v_{Rt}) \in \mathbb{R}^R$ is a vector exhibiting dimension-specific node connection rules at time t.

The multilinear tensor regression model (MTRM) is highly useful to model network dynamics. First, the multiplicative formulation, $\langle \mathbf{u}_i, \mathbf{v}_t, \mathbf{u}_j \rangle$, allows researchers to estimate time-specific network generation rules ($\mathbf{v}_t$) with time-constant latent node positions ($\mathbf{u}_i$). For example, $v_{rt} < 0$ corresponds to the case when a network generation rule for the rth dimension is heterophilic (dissortative) while $v_{rt} > 0$ corresponds to the case when a network generation rule for the rth dimension is homophilic (assortative). This property is easily obtained because of the multiplicative formulation of node distance. That is, two nodes on rth dimension with positive (negative) v_{rt} are more likely (less) to breed connections if they are located close to each other due to the inner product term $\langle \mathbf{u}_i, \mathbf{v}_t, \mathbf{u}_j \rangle$.

Second, the MTRM takes a Bayesian hierarchical form in which parameters are estimated by an efficient Gibbs sampling algorithm. Additional complexity can be easily built on the MTRM, taking advantage of the hierarchical model structure and the modularity of an Gibbs sampling algorithm.

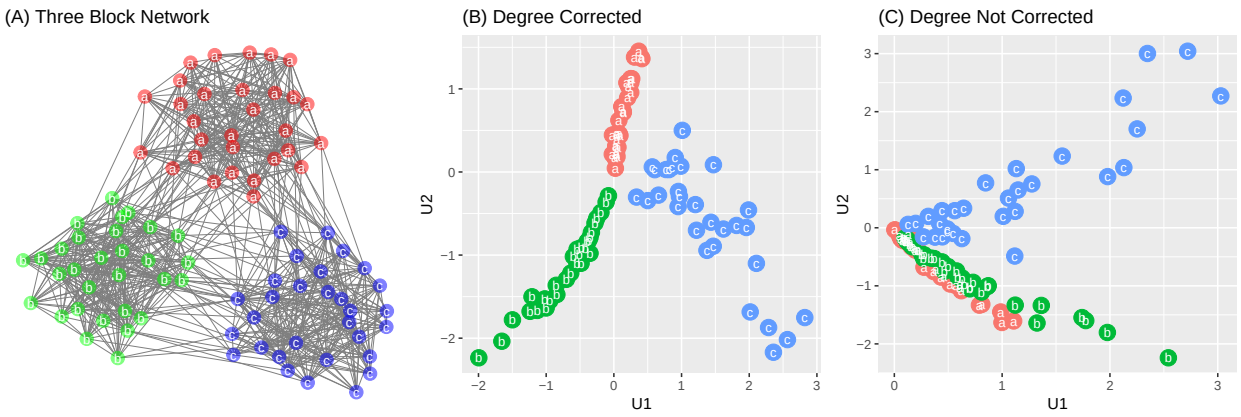


Figure 1: *Degree Correction for Block Discovery: 90 by 90 undirected network data are generated using the within-block link probability of 0.5 and the between-block link probability of 0.05.*

However, there is a critical weakness in uncovering *meso-scopic network features* related to node groups, such as community structures or core-periphery structures because nodes belonging to the same group or cluster have the identical baseline expected degrees. Since degree heterogeneity in empirical networks is common, and often such skewness can be extreme following power law or exponential distributions (Clauset et al., 2009), researchers need a null model that conditions on degree-related properties of nodes. This problem is well known in the network science literature and various degree-correction methods have been proposed (Newman, 2006, 2010; Karrer and Newman, 2011; Chaudhuri et al., 2012; Zhao et al., 2012).

Figure 1 illustrates the problem. We generate an undirected three block network data set with 90 nodes in panel (A). Panel (C) shows recovered latent node positions from a multilinear tensor regression model with a probit link. It is clear that the model fails to recover the planted three block structure. In contrast, the same model applied to a degree corrected data in panel (B) successfully recovers the planted three block structure. In panel (B), we

used an additive null model (ω_{ijt}), consisting of 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:

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

where $\tilde{\mathbf{u}}_{it}$ is the ith row of the associated eigenvector.[2]

2.2 Hidden Markov Multilinear Tensor Model

In a HMTM with k numbers of 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 numbers of 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 S_t denote a hidden state variable, $\mathbf{P}$ is a $k+1$ by $k+1$ transition matrix, and $p_{i,i}$ is the ith diagonal

[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. Both methods are available in `NetworkChange`.

element of $\mathbf{P}$. Then, a HMTM is

$$\begin{aligned}
\mathbf{B}_t &= \mathbf{Y}_t - \mathbf{\Omega}_t && (7)\\
\mathbf{B}_t &= \beta \mathbf{J}_N + \mathbf{U}_{S_t} \mathbf{\Lambda}_t \mathbf{U}_{S_t}^T + \mathbf{E}_t && (8)\\
S_t | S_{t-1}, \mathbf{P} &\sim \mathcal{M}arkov(\mathbf{P}, \pi_0) \\
p_{i,i} &\sim \mathcal{B}eta(a_0, b_0) && (9)\\
\mathbf{E}_t &\sim \mathcal{N}_{N \times N}(\mathbf{0}, \sigma_{S_t}^2 \mathbf{I}_N, \mathbf{I}_N). && (10)
\end{aligned}$$

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, which 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).

Second, we use independent inverse-gamma distributions instead of inverse-Wishart dis-

tribution for the prior distribution of a variance parameter $(\Psi_{u,m}, \Psi_{v,m})$:

$$\begin{aligned} \Psi_u &= \begin{pmatrix} \psi_{1,u,m} & \dots & 0 \\ 0 & \psi_{r,u,m} & 0 \\ 0 & \dots & \psi_{R,u,m} \end{pmatrix} \\ \psi_{r,u,m} &\sim \mathcal{IG}\left(\frac{u_0}{2}, \frac{u_1}{2}\right). \end{aligned}$$

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 covariance of $\mathbf{U}_m$ is constrained to be 0, thanks to the Gram-Schmidt process, and the 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 while keeping the conjugacy intact.

Prior distributions for each row vector of regime specific latent positions $(\mathbf{U}_m)$ are multivariate normal distributions conditioned on a variance parameter $(\Psi_{u,m})$: $\{\mathbf{u}_{1,m}, \dots, \mathbf{u}_{N,m}\} \sim \mathcal{N}_R(\boldsymbol{\mu}_{u,m}, \Psi_{u,m})$ 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): $\{\mathbf{v}_1, \dots, \mathbf{v}_T\} \sim \mathcal{N}_R(\boldsymbol{\mu}_v, \Psi_v)$ 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 dis-

tribution for β is a normal distribution ($\mathcal{N}(b_0, B_0)$).

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

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

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):

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

We discuss the details of the sampling algorithm in the supplementary information.

2.3 Assessing Model Uncertainty using Marginal Likelihood and WAIC

In this article, we consider two fully Bayesian measures of model uncertainty. Note that undirected temporal networks have unique information in the upper triangular array so

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 the 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:

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

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:

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

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 a HMTM (Gelman et al., 2014). Predictive accuracy is a good standard for break number detection because overfitting is a major concern in the 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

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

where G is the MCMC simulation size, $V[\cdot]$ indicates a variance, and $\boldsymbol{\Theta}^{(g)}, \mathbf{P}^{(g)}$ are the gth simulated outputs.

3 Simulation Studies

We test the proposed method for 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.[3] We will check whether the proposed method correctly recovers (1) the number of planted breaks and (2) the type of planted network changes. Planted clusters in network data are generated by an assortative rule in which nodes belonging to the same cluster has a higher link probability ($p_{in} = 0.5$) than nodes belonging to different clusters ($p_{out} = 0.05$). The length of time layers is 40. The planted break occurs at $t = 20$ in case of single break, (2) and (3), and $t = 10$ and $t = 30$ in case of two breaks. We fit four different HMTMs from no break to three breaks and compare model diagnostics, estimated latent spaces, and network generation rules. In Table 1 to Table 3, we report WAIC in the deviance scale following Gelman et al. (2014) and multiply -2 to the approximate log marginal likelihood for easy comparison with WAIC. 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 time-varying network generation rules for $R = 2$.

Table 1 summarizes the results from the no break case. The synthetic network is generated by a two block structure and persists over time. WAIC is smallest in the no break model while the approximate log marginal likelihood favors the one break model. The approximate log marginal likelihood consistently prefers overfit models across different network change types in our simulation tests, the pattern of which is more pronounced in the case with breaks. We will discuss this issue in details shortly. The latent space and the network generation rule of the planted two block network is well recovered by estimates of the con-

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

stant HMTM in the second row. The proposed method and WAIC correctly identifies the case of no break in dynamic networks.

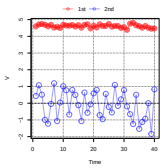
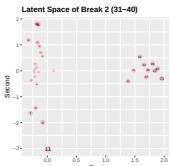
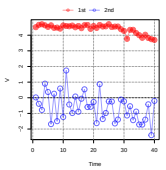
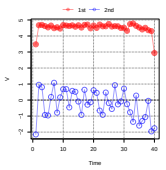
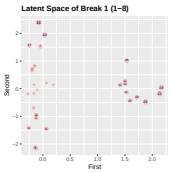
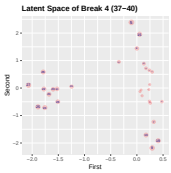
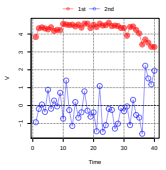
Model Fit	Latent Space ($\mathbf{U}_m$) Changes				Generation Rule ($\mathbf{v}_t$)
	Regime 1	Regime 2	Regime 3	Regime 4	
Break = 0 WAIC = 14573 $-2 \times \log p(\mathcal{B}\|\mathcal{M}_k) = 14435$ log likelihood = -7191					
Break = 1 WAIC = 14589 $-2 \times \log p(\mathcal{B}\|\mathcal{M}_k) = 14404$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

Table 2 summarizes the simulation results from a block-splitting change. The ground truth is the single break HMTM in the third row. WAIC correctly identifies the single break model while the approximate log marginal likelihood favors the three break HMTM. Note that models with redundant states (the two break model and the three break model) have single observation states (latent states consisting of a single observation). Regime 1 for HMTM with two breaks and Regime 1 and Regime 3 for HMTM with three breaks are single observation states. In model diagnostics, HMTMs with single observation states are similar to mixture models with singular components; they produce overestimated log likelihoods (Hartigan, 1985; Bishop, 2006). Since Chib (1995)'s algorithm is based on the

summation of log likelihoods evaluated at posterior means at the end of MCMC sampling and log likelihoods serve as a measure of the goodness of the fit, models with single observation states are favored by the approximate log likelihood estimate. 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.

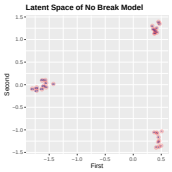
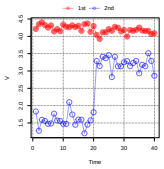
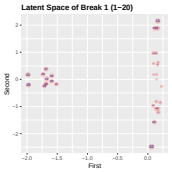
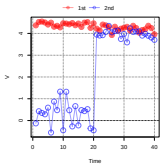
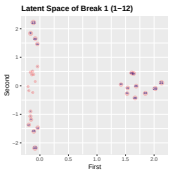
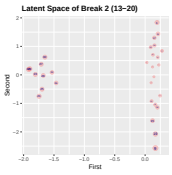
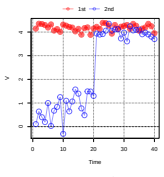
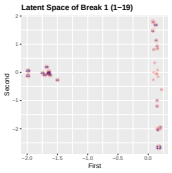
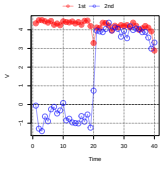
Model Fit	Latent Space ($\mathbf{U}_m$) Changes				Generation Rule ($\mathbf{v}_t$)
	Regime 1	Regime 2	Regime 3	Regime 4	
Break = 0 WAIC = 13744 $-2 \times \log p(\mathcal{B}\|\mathcal{M}_k) = 13648$ log likelihood = -6796					
Break = 1 WAIC = 13016 $-2 \times \log p(\mathcal{B}\|\mathcal{M}_k) = 12922$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

Next, we check whether HMTM correctly recovers the planted network changes in a more complicated setting. In this example, we generated a synthetic longitudinal network data

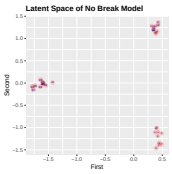
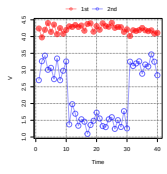
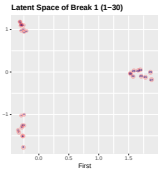
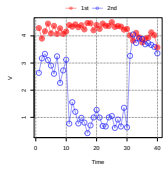
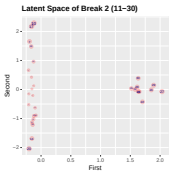
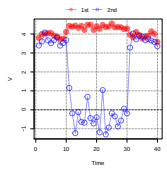
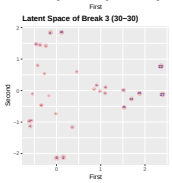
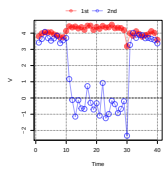
Model Fit	Latent Space ($\mathbf{U}_m$) Changes				Generation Rule ($\mathbf{v}_t$)
	Regime 1	Regime 2	Regime 3	Regime 4	
Break = 0 WAIC = 13727 $-2 \times \log p(\mathcal{B}\|\mathcal{M}_k) = 13630$ log likelihood = -6787					
Break = 1 WAIC = 13583 $-2 \times \log p(\mathcal{B}\|\mathcal{M}_k) = 13431$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

set with two structural changes. The first break is a block-merging change right after $t = 10$ and the second change is a block-splitting change right after $t = 30$. Table 2 summarizes the simulation results from the block-merging-splitting changes. Again, WAIC correctly finds the two break model while the approximate log likelihood favors the three break model. The latent space The two break HMTM correctly recovers the two underlying changes at $t = 11$ (block merging) and $t = 31$ (block splitting). Note that when block a and block b are merged into one block in Regime 2, the second element of regime-averaged network generation rule (v_2) becomes small while v_1 remains large. Changes in the relative size of network generation rules indicate us corresponding changes in the underlying network structure. In Regime 3, the number of blocks increases back to 3 and v_2 returns to its previous level at Regime 1.

More test results are available in the supplementary information.

Overall, our simulation results show that the network break number detection using WAIC of HMTM is highly effective in correctly identifying the planted break numbers. Also, the simulation shows that HMTM successfully recovers various types of block structure changes through regime-changing latent node positions and time-varying network generation rules.

4 Applications

4.1 Military Alliance Network Data

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. One reason is the lack of statistical methods that directly model structural changes in longitudinal network data, which we address in this paper. The other 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 divide our analysis into two parts. The first part uses the data from "major powers" that play a dominant role in shaping the structure of the international system and tend to last for a long time in the international system. 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 analysis. In the second part, we focus on the postwar period in which the number of states and alliance links among them explode. After removal of isolated nodes during the postwar period, our sample contains 104 states.[4]

4.2 Analysis of Alliance Networks Among Major Powers, 1816 - 2012

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.*

	$\mathcal{M}_0$	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_5$
$-2 * \log p(\mathcal{B}\|\mathcal{M}_k)$	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 results of the model diagnostics for the major power data set. We report Geweke's z-score for σ^2_m for convergence diagnostics with WAIC and the approximate log marginal likelihood. HMTM with four breaks are 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. 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.[5] Regime-specific network generation rule parameters v_{rt} are reported in axis labels.

[4]We sum every 2 year network to increase the density of each layer. Changing the granularity of aggregation does not change the substantive findings.

[5]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.

Except Regime 1, v_{rt} continues to be positive, indicating a strong homophily in major power alliance networks after 1830. It is interesting to find the central role of Austria-Hungary (AUH) during the second regime (1832-1880), connecting groups of major powers. The removal of Austria-Hungary would make the major power alliance network 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 (Gould and Fernandez, 1989; Burt, 2005, 2009; Stovel and Shaw, 2012). Although historians commonly assume that the age of Metternich lasted 33 years after the end of Napoleonic Wars (Rothenberg, 1968), HMTM uncovers (1) the period of Austria-Hungary's importance in the European balance of power system and (2) the specific role Austria-Hungary played in the military alliance network. Austria-Hungary's importance waned in Regime 2 largely thanks to the rise of Germany and the major power alliance network has gone through a structural change since Regime 3. To put it simply, no "*courtier honnête* (honest broker)" filling structural holes has existed in the major power alliance network since the 1880s.[6] Likewise, the findings of HMTM shed new light on historians' reading of the 19th international system, an example of which is well summarized in the following quote:

> It was not until Metternich that a statesman appeared who had not only internalized the concept but was given the opportunity to create a new international structure that explicitly embodied it. His less perceptive successors allowed it to

[6]The origin of the structural change lies in the Dual Alliance between Germany and Austria-Hungary in 1879 and a sequence of alliances that followed it (**??**). 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.

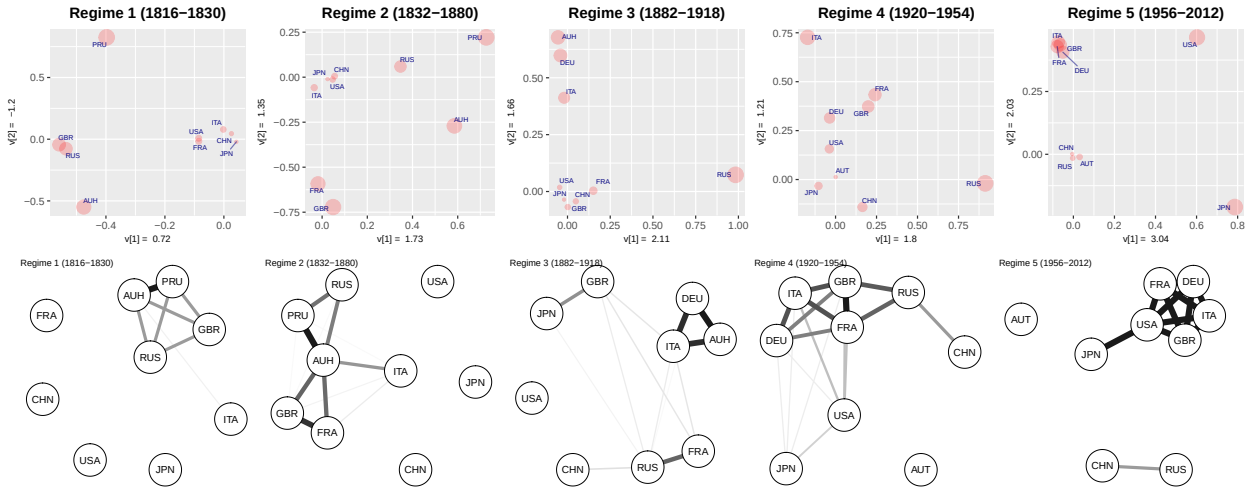


Figure 2: *Changing Node Positions and Network Topology of Military Alliance Networks Among Major Powers, 1816 - 2012: Regime averages of v_t values for each dimension are reported in the axis (top panel). Line widths (bottom panel) are proportional to the duration of alliance links.*

> 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).

Table 5: *Bayesian Model Diagnostics of Postwar Military Alliance Networks Among 104 Countries, 1946 - 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.*

	$\mathcal{M}_0$	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$
$-2 * \log p(\mathcal{B}\|\mathcal{M}_k)$	133661	124522	128234	125109	125300
WAIC	140107	145419	142328	137480	138717
Geweke's z-score (σ^2_m)	1.44	1.27	1.18	1.14	1.88

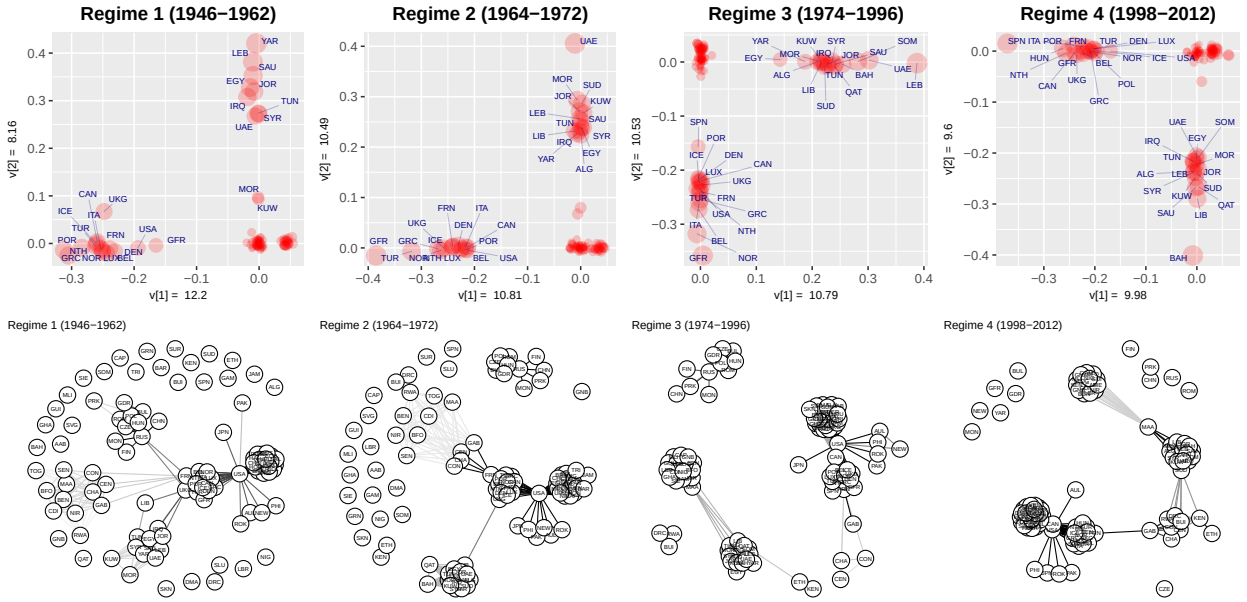


Figure 3: *Changing Node Positions and Changing Clustering Patterns in Postwar Military Alliance Networks Among 104 Countries, 1946 - 2012: In the top panel, regime averages of v_t values for each dimension are reported in the axis.*

4.3 Analysis of Postwar Military Alliance Networks, 1946 - 2012

Table 5 shows the results of the model diagnostics for the postwar data set. The three break model is preferred by both WAIC and the approximate log marginal likelihood. Models of more than four breaks show a sign of nonconvergence caused by redundant hidden states. The estimated three breaks and regime-specific latent node positions are displayed in Figure 3. The bottom panel shows changing network topology in the postwar military alliance networks. First, network generation rule parameters v_{rt} are symmetric (except the the first regime) and positive in both dimensions, which indicates that the coalition structure of postwar military alliance networks is multidimensional and shows a sign of a strong homophily. In the beginning of the Cold War, the first dimension of military alliance networks divided countries into two groups, consisting of countries close to the North Atlantic Treaty Organi-

zation (NATO) members and countries close to the Warsaw Treaty Organization members. However, this division has become less important since the second regime. The rise of the Non-Aligned Movement, decolonization, and the isolation of the Warsaw Treaty Organization members in the global alliance network have shifted the structure of the postwar military alliance networks into a multidimensional one.

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 propose method combines a highly flexible generative model of multilayer networks (MTRM) with a HMM, which have 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). Our simulation studies showed that our generative approach to network changes is a powerful tool to detect and analyze various 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 overfit models. Last, 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

compared to local network properties or global summary statistics of networks.

While we only consider undirected networks, our model can be extended to other types of network data such as directed network data or bipartite network data using a singular value decomposition-based approach to temporal network data (De Lathauwer et al., 2000; Hoff, 2007). Also, a hierarchical Dirichlet process prior can be used to endogenously detect break numbers (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. Then, the degree correction using the principal eigenmatrix would be not necessary as the goal of covariate-based analysis is to find regime-dependent associations between covariates and observed networks. In these applications, the latent space formulations serve as a statistical control for latent network effects on tie formations.

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. $p(\psi_{r,u,m}|\mathcal{B}, \mathbf{P}, \mathbf{S}, \mathbf{\Theta}^{-\Psi_{u,m}}) \propto \mathcal{IG}\left(\frac{u_0+N}{2}, \frac{\mathbf{U}_{r,m}^T \mathbf{U}_{r,m} + u_1}{2}\right)$.
2. $p(\boldsymbol{\mu}_{u,m}|\mathcal{B}, \mathbf{P}, \mathbf{S}, \mathbf{\Theta}^{-\boldsymbol{\mu}_{u,m}}) \propto$ multivariate normal$(\mathbf{U}_m^T \mathbf{1}/(N+1), \Psi_{u,m}/(N+1))$.
3. $p(\mathbf{U}_m|\mathcal{B}, \mathbf{P}, \mathbf{S}, \mathbf{\Theta}^{-\mathbf{U}_m}) \propto$ 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. $p(\psi_{r,v,m}|\mathcal{B}, \mathbf{P}, \mathbf{S}, \mathbf{\Theta}^{-\Psi_{v,m}}) \propto \mathcal{IG}\left(\frac{v_0+T}{2}, \frac{\mathbf{V}_{r,m}^T \mathbf{V}_{r,m} + v_1}{2}\right)$.
2. $p(\boldsymbol{\mu}_{v,m}|\mathcal{B}, \mathbf{P}, \mathbf{S}, \mathbf{\Theta}^{-\boldsymbol{\mu}_{v,m}}) \propto$ multivariate normal$(\mathbf{V}_m^T \mathbf{1}/(T_m+1), \Psi_{v,m}/(T_m+1))$.
3. $p(\mathbf{V}_m|\mathcal{B}, \mathbf{P}, \mathbf{S}, \mathbf{\Theta}^{-\mathbf{V}_m}) \propto$ 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

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

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

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:

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

Let $\boldsymbol{\Theta}$ indicate a parameter vector beside hidden states ($\mathbf{S}$) and a transition matrix ($\mathbf{P}$): $\boldsymbol{\Theta} = \{\boldsymbol{\mu}_u, \psi_{.,u}, \boldsymbol{\mu}_v, \psi_{.,v}, \beta, \sigma^2\}$. Let $(\boldsymbol{\Theta}^*, \mathbf{P}^*)$ be posterior means of $(\boldsymbol{\Theta}, \mathbf{P})$. 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

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

Step 2

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

Step 3

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

Step 4

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

Step 5

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

Step 6

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

Step 7

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

References

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

Akoglu, L., Tong, H., and Koutra, D. (2014), "Graph-based Anomaly Detection and Description: A Survey," *arXiv preprint arXiv:1404.4679*, .

Araujo, M., Papadimitriou, S., Günnemann, S., Faloutsos, C., Basu, P., Swami, A., Papalexakis, E. E., and Koutra, D. (2014), "Com2: Fast Automatic Discovery of Temporal ('Comet') Communities," in *Advances in Knowledge Discovery and Data Mining* Springer, pp. 271–283.

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.

Gould, R. V., and Fernandez, R. M. (1989), "Structures of Mediation: A Formal Approach to Brokerage in Transaction Networks," *Sociological Methodology*, 19, 89–126.

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. (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.

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., Song, L., Ahmed, A., Xing, E. P. et al. (2010), "Estimating time-varying networks," *The Annals of Applied Statistics*, 4(1), 94–123.

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

Koutra, D., Papalexakis, E. E., and Faloutsos, C. (2012), Tensorsplat: Spotting latent anomalies in time,, in *Informatics (PCI), 2012 16th Panhellenic Conference on*, IEEE, pp. 144–149.

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.

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.

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.

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 neuroscience," *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.

Volfovsky, A., and Hoff, P. (2015), "Testing for Nodal Dependence in Relational Data Matrices," *Journal of the American Statistical Association*, 110(511), 1037–1046.

Walt, S. M. (1987), *The Origins of Alliances* Cornell University Press.

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.