Detecting Structural Changes in Longitudinal Network Data

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

Yunkyu Sohn
Waseda University
ysohn@aoni.waseda.jp

November 6, 2018

Abstract

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 to find structural changes in the coalition structure among nations.

1 Introduction

Tensor decomposition is becoming a standard means to analyze longitudinal network datasets (Hoff, 2009a, 2011; Rai et al., 2015; Hoff, 2015; Minhas et al., 2016; Johndrow et al., 2017; Han and Dunson, 2018). Compared to network models for static snapshots or matrix-valued datasets, this approach significantly advances our modeling possibility. A longitudinal network data
set can be represented as a tensor $\mathcal{Y} = \{\mathbf{Y}_t|t \in \{1,\ldots,T\}\} \in \mathbb{R}^{N \times N \times T}$, which is an array of $N \times N$ square matrices $\mathbf{Y}_t = \{y_{ijt}|i,j \in \{1,\ldots,N\}\}$. Here $y_{ijt}$ informs the dyadic relationship between actors $i$ and $j$ at time $t$.

While dynamic modeling of tensor-valued longitudinal networks, mainly in a form of reduced-rank decomposition, has been an increasingly important topic in social, biological, and other fields of science, a fully probabilistic treatment of dynamic network process has been a challenging problem due to simultaneous dependence between dyadic and temporal observations that are often associated with fundamental shifts in data generating processes.

By employing a Bayesian framework of tensor decomposition by Hoff (2011) and Hoff (2015), we present a Bayesian method to model change-point process in longitudinal networks. The motivation of our method is firmly based on a common observation in network analysis that longitudinal network datasets frequently exhibit irregular dynamics, implying multiple changes in their data generating processes (e.g. Guo et al., 2007; Heard et al., 2010; Wang et al., 2014; Cribben and Yu, 2016; Barnett and Onnela, 2016; Ridder et al., 2016).

Figure 1 shows an example of longitudinal network data with 3 layers of time series and 90 nodes. The data generating parameters of nodes in each network are depicted by 2 dimensional latent traits at each time point. The distance between each pair of nodes represents their probability of connection, so that proximal nodes are more likely to have connections. As clearly shown by the blocks of matrices at the bottom panel, the clustered patterns of connections are well depicted by the node positions, that can be recovered using our method, on the top panel. Colors in the backgrounds of the layers represent latent regimes inferred using our method. In addition to the latent traits that are specific to each network, one can easily notice that the two cluster networks at $t = 1$ turned into a three cluster network at $t = 2$. Contrary to the dramatic change of the overall network structure at $t = 2$, the network at $t = 3$ exhibits identical node positions to the $t = 2$ network. The same color indicates layers sharing latent node positions. The goal of our method is to uncover 1) latent traits representing data generating processes at each regime sharing those traits (colored node positions/groupings in the top panel) and 2) the timing of unspecified number of changes ($t = \{1\}$ for regime 1 and $t = \{2, 3\}$ for regime 2).

---

For a range of examples for longitudinal network analysis, see Holme and Saramäki (2012) and references therein.
Figure 1: A dynamic network with 3 temporal snapshots embedded in their latent (i.e. unobserved) node positions and latent (colored) regimes of layers that can be recovered using our method (top panel). The same network is represented in a tensor format (bottom panel). Gray lines between layers indicate node identity. Nodes in the same color exhibit dense connections whereas nodes in different colors exhibit sparse connections. Same cluster connections are represented by their node group color and inter-cluster connections are colored in white. The patterns of connections of the three snapshots can be represented as a tensor as shown in the bottom panel. Each network snapshot is shown as a matrix with black dots representing the presence of connections between corresponding node pairs. Olive and yellow colors shown in the backgrounds of top and bottom panels indicate the hidden regimes of latent coordinates shared among layers. $t = 1$ network has distinct characteristics, consisting of two clusters, whereas $t = 2$ and $t = 3$ networks share latent coordinates of three clusters.

Conventional approaches to dynamic network modeling typically extend a static network analysis framework by assuming smooth topological changes over time or applying completely separate models for each time period (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). These methods rely largely on heuristic approaches to detect structural changes in data generating parameters. Recently, several methods for the “network
change-point detection” problem have been proposed, noting the importance of irregular changes in network structures. For example, Cribben and Yu (2016) introduce a two-step approach to network change-point detection in which the cosine distances for the principal eigenvectors of time-specific graph Laplacian matrix are used to find change-points given pre-specified significance thresholds.\(^2\) Another group of studies (Guo et al., 2007; Wang et al., 2014) allow parameter values of exponential random graph models (ERGMs) to change over time. However, both models exhibit computational inefficiency. For instance, the maximum size of network analyzed was 11 nodes in Guo et al. (2007) and 6 nodes in Wang et al. (2014).\(^3\) By incorporating the stochastic blockmodel (SBM) framework, which presumes the existence of discrete node groups, Ridder et al. (2016) propose a method to identify a single parametric break. Ridder et al. (2016)’s method compares the bootstrapped distribution of the log-likelihood ratio between a null model and an alternative. However, the asymptotic distribution of a SBM with a break approaches a mixture of $\chi^2$-distributions. Hence it does not meet the regularity condition of the log-likelihood ratio test statistic (Drton, 2009). A recent approach by Bartolucci et al. (2018) is also restricted to model changes in group membership in the SBM setting when the number of group is fixed. Likewise, existing methods for the “network change-point detection” problem lack the capacity of a fully probabilistic modeling and fail to incorporate uncertainty in the model structure and parameter estimation.

Our approach diverges from previous methods in two significant ways. First, we build a dynamic model using Hoff (2011, 2015)’s multilinear tensor regression model (MTRM), which is a multilayer (i.e. tensor) extension of the latent space approach to network data. MTRM allows us to decompose longitudinal network data into node-specific (or row and column) random effects and time-specific (or layer-specific) random effects.\(^4\) For example, let

\(^2\)Graph Laplacian is one of the most well-known linear operators for adjacency matrix that is designed to minimize the summed quadratic distances between latent positions of connected (unconnected) node pairs for an assortative (dissortative) network.

\(^3\)In the framework of the temporal exponential random graph models (TERGM), Cranmer et al. (2014) pre-tested the existence of parametric breaks in global network statistics. 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 principled tools to select the number of parametric breaks.

\(^4\)These two effects correspond to the node positions at the top panel of Figure 1 and data generating parameters associated with the global patterns of connections respectively.
\{z_{i,j,t}\}$ be latent propensities to form a link between $i$ and $j$ observed at time $t$ and $x_{i,j,t}$ be a vector of known covariates affecting $\{z_{i,j,t}\}$. Then, based on the notion of multilayer exchangeability (Hoff, 2009b), MTRM models the latent propensity of link formation ($\{z_{i,j,t}\}$) as a function of covariates, node-specific random effects ($\{u_1, \ldots, u_N\}$) and time-specific random effects ($V_t$):

\[
\Pr(y_{i,j,t} = 1|x_{i,j,t}, u_i, u_j, V) = \Phi(z_{i,j,t}) \tag{1}
\]

\[
z_{i,j,t} = x_{i,j,t}\beta + u_i^T V_t u_j + \epsilon_{i,j,t} \tag{2}
\]

\[
\epsilon_{i,j,t} \sim \mathcal{N}(0, 1) \tag{3}
\]

where $\{u_1, \ldots, u_N\}$ represent (time-constant) $R$-dimensional latent node positions and $V_t$ is a diagonal matrix of (time-varying) node-connection rules. As we will explain in details, this multiplicative decomposition is highly useful for the joint estimation of time-varying network generation rules in conjunction with latent node positions that are constant for the duration of a hidden regime. Different from SBM formulation, the continuous multidimensional node position formulation let us to model any underlying latent structure including both group-structured networks, treated by SBM, and networks without group substructures that are unable to be modeled by SBM.

The second departure of our approach from existing methods is the use of hidden Markov model (HMM) to characterize the change-point process. 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 turns out to be highly useful to model unknown changes in the latent network traits. More specifically, latent node positions ($\{u_1, \ldots, u_N\}$), which are constrained to be constant over time in Hoff (2011), are allowed to change over time depending on the transition of hidden states.

The resulting model is a hidden Markov multilinear tensor model (HMTM) as it combines Hoff (2011)’s MTRM with HMM. HMTM assumes that a dynamic network process can be modeled as discrete changes in the latent space representation of network layers at each time point. These changes reflect fundamental shifts in structural properties of network under consideration. For example, structural changes in military alliance networks reflect the transformation of the international system such as the balance of power system during the Concert of Europe, the bifurcated system in the run-up to the World War I, and the bipolar structure during the Cold War, as we
will see shortly.

The proposed method has several notable contributions to longitudinal network analysis. First, we show that degree heterogeneity hinders the recovery of meaningful traits in the latent space approach and demonstrate that degree correction formulations (Karrer and Newman, 2011; Chaudhuri et al., 2012) make a crucial difference in the recovery of ground-truth group structures underlying our example data generation. Second, our method provides an important tool to understand dynamic network processes by allowing researchers to model fundamental changes in factors underlying the evolution of longitudinal networks. Changes in longitudinal network data can take a variety of forms and our method does not restrict the types of network generating models. Finally, we provide an open-source R package, NetworkChange, that implements all the methods introduced in the paper including Bayesian model diagnostic tools: the approximate log marginal likelihood (Chib, 1995), the Watanabe-Akaike Information Criterion (WAIC) (Watanabe, 2010). We report the performance test results of these diagnostics.

2 Understanding Multilinear Tensor Regression Model

2.1 Latent Space Model for Tensor

Let \( U = (u_1, \ldots, u_N) \top \in \mathbb{R}^{N \times R} \) be the \( R \)-dimensional latent node positions of \( N \) nodes and \( v_t = (v_{1t}, \ldots, v_{Rt}) \in \mathbb{R}^{R} \) 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 (\( u_i \) for node \( i \) and \( u_j \) for node \( j \)) and layer-specific node-connection rules (\( v_t \) at time \( t \) or \( t \)th layer) as follows:

\[
\Pr(y_{i,j,t} = 1|x_{i,j,t}, u_i, u_j, v_t) = \Phi(x_{i,j,t}\beta + \langle u_i, v_t, u_j \rangle) \quad (4)
\]

\[
U \sim \text{matrix normal}(M = 1\mu_U, I_N, \Psi_U) \quad (5)
\]

\[
V \sim \text{matrix normal}(M = 1\mu_V, I_T, \Psi_V) \quad (6)
\]

\[
\epsilon_{i,j,t} \sim \mathcal{N}(0, \sigma^2) \quad (7)
\]

where \( \langle u_i, v_t, u_j \rangle = \sum_{r=1}^{R} u_{ir} v_{rt} u_{jr} \) and matrix normal\((M, U, V)\) is a \( N \times R \) matrix-variate normal distribution with mean matrix \( M \), row variance \( U \), and column variance \( V \).
The resulting estimates of node-specific latent variables recover a specific type of similarity between nodes that is easily interpretable (Hoff, 2008). If \( u_i \) and \( u_j \) exhibit similar values, they will have similar inner product outcomes with node \( k \)'s latent position vector \( u_k \). This means that the probability of connection with \( k \) is analogous for \( i \) and \( j \). This corresponds to the notion in network theory that nodes \( i \) and \( j \) are structurally equivalent (Wasserman and Faust, 1994). In addition, the generation rule parameter \( v_t \) contains the information on what the distance relationships on each dimension of the \( U \) space reveal about their connection probability. For example, \( v_{rt} > 0 \) corresponds to the case when a network generation rule for the \( r \)th dimension at time \( t \) is homophilous (assortative). In words, \( v_{rt} > 0 \) indicates that two nodes on \( r \)th 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 \( r \)th 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 the same sign.

2.2 Degree Correction

One of most important features of the proposed method is detecting changes in meso-scopic network properties, grouping of nodes, such as homophilous or heterophilous groups and core-periphery substructures in a network. The emergence (and changes) of group structures is commonly observed in real-world network data (Borgatti and Everett, 1999; Nowicki and Snijders, 2001; Newman, 2006; Fortunato, 2010; Xu, 2015; Ridder et al., 2016). The formulation of MTRM in Equation (4), however, is designed to recover consistent regression parameters (\( \beta \)) considering network effects as a nuisance parameter. Hence it entails a critical weakness in uncovering latent meso-scopic network features.

Except for exogenous covariates (\( x_{i,j,t} \)), Equation (4) has no treatment to account for degree heterogeneity that has been known to confound the group structure recovery (e.g. Newman, 2006, 2010; Karrer and Newman, 2011). The intuition is that the distribution of degrees in empirical networks is highly heterogeneous and skewed following power law or exponential distributions (Clauset et al., 2009) while the implicit assumption in the group structure recovery is that the expected degree of nodes having a similar role (i.e. proximal in the latent space or belonging to the same group) is simi-
lar. 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 2: Degree Correction for Group Structure Discovery: Two undirected networks with 30 nodes and 3 groups are generated. In Layer 1, the within-group link probability of 0.5 and the between-group link probability of 0.2 (homophily) and in Layer 2 the within-group link probability of 0.2 and the between-group link probability of 0.5 (heterophily).

Figure 2 illustrates the problem in a simple setting using a 2 layer network where we assume no change in node positions. We generate two undirected networks consisting 3 groups, each group composed of 10 nodes, in Figure 2(A) and Figure 2(B). The number of groups and node membership are identical but the connection rules are opposite. In Figure 2(A), the within-group link probability (0.5) is much larger than the between-group link probability (0.2) and the probabilities are flipped in Figure 2(B). The node colors indicate group memberships and the lines indicate links. Without assuming a change between 2 layers, our task is to identify 3 hidden groups from the data using their recovered node positions.

We first fit an probit MTRM shown in Equation (4) with an unconditional mean parameter without external covariates. Then, we applied the k-means clustering algorithm to the estimated latent node positions to identify group membership. The results are reported in Figure 2(C). The MTRM fails to distinguish the green group from the blue group and researchers will conclude erroneously that the data are generated by two groups.

A simple fix to this problem is to use a null model to control for the expected level of associations among pairs of nodes. One example is an additive null model \( \omega_{ijt} \), consisting of the principal eigenvalue \( \lambda_{t}^{\text{princ}} = -1 \).
max(|λ(Y_t)|)) and its associated eigenvector (Peixoto, 2013):

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

(8)

where \( \tilde{u}_{it} \) is the \( i \)th row of the associated eigenvector.\(^5\) In matrix form, we denote the principal eigenmatrix as \( \Omega_t \).

Figure 2(D) shows the results from the linear MTRM on the transformed data (\( B_t = Y_t - \Omega_t \)). Colors are allocated by the \( k \)-means clustering analysis. The use of a null model allows us to recover three distinct blocs in the data. Of course, one can think of an alternative way of controlling for the expected level of associations by including a list of external covariates. However, when the goal is to identify hidden groups, not coefficients of covariates, using a null model is more intuitive and computationally less expensive than including a list of covariates.

3 The Proposed Method

In order to develop a dynamic network model for structural changes, we must start from the question, “What constitutes structural changes in networks?” On the one hand, one can think of a change in summary statistics of macroscopic 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 microscopic 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;

\(^5\)Alternatively, one can use a modularity matrix (\( M_{ijt} \)):

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

where \( m = \sum_{i=1}^{N} k_i \) and \( k_i \) is the sum of weights for \( i \) (Newman and Girvan, 2004). Both methods are available in NetworkChange.
Tibély et al., 2011; Sporns, 2014). Technically, various approaches for mesoscopic 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 mesoscopic 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.

3.1 Hidden Markov Multilinear Tensor Model

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 $S$ as a vector of hidden state variables where $S_t$ is an integer-valued hidden state variable at $t$

$$S = \{(S_1, \ldots, S_T) : S_t \in \{1, \ldots, M\}, t = 1, \ldots, T\},$$

and $P$ as a $M \times M$ transition matrix where $p_{i,i}$ is the $i$th diagonal element of $P$ and $M$ is the number of hidden states.

Then, the probability distribution of a degree-corrected longitudinal network data with $M - 1$ breaks can be modeled as a Markov mixture of $M$-component MTRMs using the conditional independence assumption of the HMM. Suppose $\Theta$ be a collection of parameters that represent a network generating process of a longitudinal network. Then,

$$p(B|\Theta) = \int p(S_1|\Theta)p(B_1|S_1, \Theta) \prod_{t=2}^T \sum_{m=1}^M p(B_t|\Theta_m) \Pr(S_t = m|S_{t-1}, \Theta) dS,$$

where $p(B_t|\Theta_m)$ a generative network model at regime $m$. Here, the duration of hidden state $m$ follows a geometric distribution of $1 - p_{mm}$ where $p_{mm}$ is the $m$th 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. $\sum_{g=1}^G I(S_t^{(g)} \neq S_{t-1}^{(g)})$).

Following the MTRM, the HMTM decomposes the degree corrected network data at $t$ as a bilinear product of latent node positions and dimension weights subject to hidden state changes:

$$B_t = U_{S_t}V_tU_{S_t}^T + E_t$$

$$E_t \sim \begin{cases} \mathcal{N}_{N \times N}(0, \sigma^2_{S_t}I_N, I_N) & \text{for Normal Error} \\ \mathcal{N}_{N \times N}(0, \gamma_t^{-1}\sigma^2_{S_t}I_N, I_N) & \text{for Student-t Error} \end{cases}$$
One may concern that the normal distribution of $E_t$ does not fit the data very well. In that case, the above model can be modified to include a Student-$t$ distributed error (Carlin and Polson, 1991) where the prior distribution of $\gamma_t$ follows a gamma distribution ($\mathcal{G}(\nu_0/2, \nu_1/2)$).

For prior distributions of $U$ and $V$, we follow Hoff (2011)’s hierarchical scheme with two major modifications. First, we orthogonalize each column of $U_{st}$ using the Gram-Schmidt process (Björck, 1996; Guhaniyogi and Dunson, 2015) in each simulation step. Hoff (2011)’s hierarchical scheme centers rows of $U_{st}$ around its global mean ($\mu_{u,st}$) using a multivariate normal distribution. This does not guarantee the orthogonality of each latent factor in $U_{st}$. 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 distribution for the prior distribution of a variance parameter ($\Psi_{u,s,t}$, $\Psi_v$). The use of inverse-Wishart distribution for the prior distribution of a variance parameter ($\Psi_{u,s,t}$, $\Psi_v$) comes at a great cost because choosing informative inverse-Wishart prior distributions for $\Psi_{u,m}$ and $\Psi_v$ 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 $\Psi_{u,s,t}$ and $\Psi_v$ often goes to a negative infinity, failing to impose proper penalties. In HMTM, the off-diagonal cov-variance of $U_m$ is constrained to be 0, thanks to the Gram-Schmidt process, and the off-diagonal covariance of $V$ is close to 0 as $v_t$ measures time-varying weights of independent $U_m$. Thus, inverse-gamma distributions resolve a computational issue without a loss of information.

The resulting prior distributions of $U$ and $V$ are matrix-variate normal distributions in which each column vector ($u_{i,s,t}$ and $v_t$) follows a multivariate normal distribution. We first discuss the prior distribution of $U$:

\[
\begin{align*}
U_{st} &\equiv (u_{1,s,t}, \ldots, u_{N,s,t})^\top \in \mathbb{R}^{N \times R} \\
u_{i,s,t} &\sim \mathcal{N}_R(\mu_{u,s,t}, \Psi_{u,s,t}) \\
\mu_{u,s,t} | \Psi_{u,s,t} &\sim \mathcal{N}_R(\mu_{0,u,s,t}, \Psi_{u,s,t}) \\
\Psi_{u,s,t} &\equiv \begin{pmatrix} \psi_{1,u,s_t} & \cdots & 0 \\ 0 & \psi_{r,u,s_t} & 0 \\ 0 & \cdots & \psi_{R,u,s_t} \end{pmatrix} \\
\psi_{r,u,s_t} &\sim \mathcal{IG}(\nu_0/2, \nu_1/2).
\end{align*}
\]
The prior distributions of $V$ are similar to $U$ but one difference is that only diagonal elements of $V_t$ are modeled as a multivariate normal distribution:

$$V_t \equiv \begin{pmatrix} v_{1,t} & \ldots & 0 \\ 0 & v_{r,t} & 0 \\ \vdots & \ddots & \ddots \end{pmatrix}$$ (17)

$$v_t \equiv (v_{1,t}, \ldots, v_{R,t})^\top \in \mathbb{R}^{R \times 1}$$ (18)

$$v_t \sim N_R(\mu_v, \Psi_v)$$ (19)

$$\mu_v | \Psi_v \sim N_R(\mu_{0,v}, \Psi_v)$$ (20)

$$\Psi_v = \begin{pmatrix} \psi_{1,v} & \ldots & 0 \\ 0 & \psi_{r,v} & 0 \\ 0 & \ldots & \psi_{R,v} \end{pmatrix}$$ (21)

$$\psi_{r,v} \sim IG \left( \frac{v_0}{2}, \frac{v_1}{2} \right).$$ (22)

Then, we complete the model building by introducing HMM-related prior specifications following Chib (1998):

$$S_t | S_{t-1}, P \sim Markov(P, \pi_0)$$ (23)

$$P_{M \times M} = (p_1, \ldots, p_M)$$ (24)

$$p_i \sim Dirichlet(\alpha_{i,1}, \ldots, \alpha_{i,M})$$ (25)

where $\pi_0$ is the initial probability of a non-ergodic Markov chain ($\pi_0 = (1, 0, \ldots, 0)$).

### 3.2 Sampling Algorithm

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

$$p(\Theta, P, S|B) \propto N_{N \times N}(B_1|\Theta_1) \prod_{t=2}^{T} \left( N_{N \times N}(B_t|B_{t-1}, \Theta_{S_t}) p(S_t|S_{t-1}, P) \right)$$

$$\prod_{m=1}^{M} \left( N_R(\mu_{u,m}|\mu_{0,u,m}, \psi_{u,m}) N_R(\mu_{v}|\mu_{0,v}, \psi_{v,m}) \right)$$

$$\prod_{m=1}^{M} \prod_{r=1}^{R} \left( IG(\psi_{r,u,m}|u_{0,m}, u_{1,m}) IG(\psi_{r,v,m}|v_{0,m}, v_{1,m}) \right)$$

$$\prod_{m=1}^{M} \left( IG(\sigma^2_m|c_0, d_0) Beta(p_{mm}|a, b) \right)$$

where $$B_{t-1} = (B_1, \ldots, B_{t-1})$$. 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):

$$p(\Theta, P, S|B) = p(\Theta|B, P, S) p(P|B, S) p(S|B) .$$

The sampling algorithm of the HMTM can be summarized as follows:

**Step 1** The sampling of regime specific $$U, \mu, \Psi_u$$ consists of the following three steps for each regime $$m$$. Let $$\Psi_u = \begin{pmatrix} \psi_{1,u,m} & \ldots & 0 \\ 0 & \psi_{u,u,m} & 0 \\ 0 & \ldots & \psi_{R,u,m} \end{pmatrix}$$.

1. $$p(\psi_{r,u,m}|B, P, S, \Theta^{-\psi_{u,m}}) \propto IG \left( \frac{u_0 + N}{2}, \frac{U_{r,m}U_{r,m}^{T} + u_1}{2} \right)$$.
2. $$p(\mu_{u,m}|B, P, S, \Theta^{-\mu_{u,m}}) \propto \text{multivariate normal}(U_{m}^{T}1/(N+1), \Psi_{u,m}/(N+1))$$.
3. $$p(U_{m}|B, P, S, \Theta^{-U_{u,m}}) \propto \text{matrix normal}_{N \times R}(M_{u,m}, I_N, \tilde{\Psi}_{u,m})$$ where

$$\tilde{\Psi}_{u,m} = (Q_{u,m}/\sigma^2_{u,m} + \psi^{-1}_{u,m})^{-1}$$

$$M_{u,m} = (L_{u,m}/\sigma^2_{u,m} + 1\mu_{u,m}^{T}\psi^{-1}_{u,m})\tilde{\Psi}_{u,m}$$

$$Q_{u,m} = (U_{m}^{T}U_{m}) \circ (V_{m}^{T}V_{m})$$

$$L_{u,m} = \sum_{j,t: t \in S_t = m} b_{j,t} \otimes (U_{m,j} \circ V_{m,t})$$

4. Orthogonalize $$U_{m}$$ using the Gram-Schmidt algorithm.
Step 2  The sampling of $V, \mu_v, \Psi_v$ is done for each regime. Let $\Psi_v = \\
\begin{pmatrix}
\psi_{1,v,m} & \ldots & 0 \\
0 & \psi_{r,v,m} & 0 \\
0 & \ldots & \psi_{R,v,m}
\end{pmatrix}.
\\
1. $p(\psi_{r,v,m}|B, P, S, \Theta^{-\Psi_{v,m}}) \propto IG\left(\frac{m+T}{2}, V_{r,m}V_{r,m}+\psi_{r,v,m}\right)$.
2. $p(\mu_{v,m}|B, P, S, \Theta^{-\mu_{v,m}}) \propto \text{multivariate normal}(V_{T_{m}}^{-1}/(T_{m}+1), \Psi_{v,m}/(T_{m}+1))$.
3. $p(V_{m}|B, P, S, \Theta^{-V_{m}}) \propto \text{matrix normal}_{T_{m} \times R}(\bar{M}_{v,m}, I_{T_{m}}, \bar{\Psi}_{v,m})$ where
\begin{align*}
\bar{\Psi}_{v,m} &= (Q_{v,m}/\sigma_{m}^{2} + \Psi_{v,m}^{-1})^{-1} \\
\bar{M}_{v,m} &= (I_{v,m}/\sigma_{m}^{2} + I_{v,m}^{T}\Psi_{v,m}^{-1}) \bar{\Psi}_{v,m} \\
Q_{v,m} &= (U_{m}^{T}U_{m}) \circ (U_{m}^{T}U_{m}) \\
L_{v,m} &= \sum_{i,j} b_{i,j} \otimes (U_{m,i} \circ U_{m,j}).
\end{align*}

Step 3  The sampling of $\sigma_{m}^{2}$ from $IG\left(\frac{c_{0}+N_{m}+d_{0}+\sum_{i,j} b_{i,j} \otimes (U_{m,i} \circ U_{m,j})}{2}, \frac{d_{0}+\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T_{m}} b_{i,j,t} - \mu_{i,j,t}}{2}\right)$.

Step 4  Sample $S$ recursively using Chib (1998)'s algorithm. The joint conditional distribution of the latent states $p(S_0, \ldots, S_T|\Theta, B, P)$ can be written as the product of $T$ numbers of independent conditional distributions:
\begin{align*}
p(S_0, \ldots, S_T|\Theta, B, P) &= p(S_T|\Theta, B, P) \cdots p(S_t|S_{t+1}, \Theta, B, P) \cdots p(S_0|S^1, \Theta, B, P).
\end{align*}

Using Bayes' Theorem, Chib (1998) shows that
\begin{align*}
p(S_t|S_{t+1}, \Theta, B, P) &\propto p(S_{t}|\Theta, B_{1:t}, P) \cdot p(S_{t+1}|S_{t}, P) \cdot \text{State probabilities given all data Transition probability at } t
\end{align*}

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 ($P$). 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).

During the burn-in iterations, if sampled $S$ has a state with single observation, randomly sample $S$ with replacement using a pre-chosen perturbation weight ($w_{\text{perturb}} = (w_1, \ldots, w_M)$).
Step 5 Sample each row of $P$ from the following Beta distribution:

$$p_{kk} \sim \text{Beta}(a_0 + j_{k,k} - 1, b_0 + j_{k,k+1})$$

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 $k + 1$.

We provide the sampling details of the HMTM with a Student-$t$ distributed error the supplementary material.

### 3.3 Assessing Model Uncertainty

We provide three metrics for model diagnostics and break number detection: the approximate log marginal likelihood method, WAIC, and the average loss of break points. 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:

$$\log \hat{p}(\mathcal{B}^{\text{upper}}|\mathcal{M}_M) = \log p(\mathcal{B}^{\text{upper}}|\mu^*_u, \psi^*_u, \mu^*_v, \psi^*_v, \sigma^2, P^*, \mathcal{M}_M)$$

$$+ \sum_{m=1}^{M} \log p(\mu^*_{u,m}, \psi^*_{u,m}, \mu^*_{v,m}, \psi^*_{v,m}, \sigma^2, p^*_{m,m}|\mathcal{M}_M)$$

$$- \sum_{m=1}^{M} \log p(\mu^*_{u,m}, \psi^*_{u,m}, \mu^*_{v,m}, \psi^*_{v,m}, \sigma^2, p^*_{m,m}|\mathcal{B}^{\text{upper}}, \mathcal{M}_M)$$

where $\{\mu^*_u, \psi^*_u, \mu^*_v, \psi^*_v, \sigma^2, P^*\}$ 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:

\[
\sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=i+1}^{N-1} \sum_{m=1}^{M} p(b_{i,j,t}|B_{t-1}^{upper}, \mu_{u,m}^{*}, \psi_{i,u,m}^{*}, \mu_{v,m}^{*}, \psi_{i,v,m}^{*}, \sigma_{m}^{2*}, P^{*}, S_{t} = m, \mathcal{M}_{M}) p(S_{t} = m|B_{t-1}^{upper}, \mu_{u,m}^{*}, \psi_{i,u,m}^{*}, \mu_{v,m}^{*}, \psi_{i,v,m}^{*}, \sigma_{m}^{2*}, P^{*}, \mathcal{M}_{M}).
\]

The computation of the log posterior density of posterior means requires a careful blocking in a highly parameterized model as discussed in Chib (1995). In our HMTM, the log posterior density evaluated at posterior means is decomposed into seven blocs:

\[
\log p(\mu_{u}^{*}, \psi_{i,u}^{*}, \mu_{v}^{*}, \psi_{i,v}^{*}, \sigma^{2*}, P^{*}|B) = \log p(\mu_{u}^{*}|B) + \sum_{r=1}^{R} \log p(\psi_{r,u}^{*}|B, \mu_{u}^{*}) + \log p(\mu_{v}^{*}|B, \mu_{u}^{*}, \psi_{i,u}^{*}) + \sum_{r=1}^{R} \log p(\psi_{r,v}^{*}|B, \mu_{u}^{*}, \psi_{i,u}^{*}, \mu_{v}^{*}) + \log p(\sigma^{2*}|B, \mu_{u}^{*}, \psi_{i,u}^{*}, \mu_{v}^{*}, \psi_{i,v}^{*}) + \log p(P^{*}|B, \mu_{u}^{*}, \psi_{i,u}^{*}, \mu_{v}^{*}, \psi_{i,v}^{*}, \sigma^{2*}).
\]

The second measure of model diagnostics 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 the estimated WAIC scores are dependent upon latent state estimates. The dependence on estimated latent states indicate that our measure of WAIC cannot be used to predict future networks given the sequence of network observation. Instead, we aim to use WAIC to compare predictive accuracies of HMTMs given a varying number of breaks.

Using the formula suggested by Gelman et al. (2014), WAIC for HMTM
with $M$ number of latent states ($\mathcal{M}_M$) is

$$WAIC_{\mathcal{M}_M} = -2 \left( \sum_{t=1}^{T} \log \left[ \frac{1}{G} \sum_{g=1}^{G} p(\mathcal{B}_t^\text{upper} | \mu_u^{(g)}, \psi_{u,u}, \mu_v^{(g)}, \psi_{v,v}, \sigma_2^{(g)}, P^{(g)}, S^{(g)}, \mathcal{M}_M) \right] \right) - \sum_{t=1}^{T} V_g \left[ \log p(\mathcal{B}_t^\text{upper} | \mu_u^{(g)}, \psi_{u,u}, \mu_v^{(g)}, \psi_{v,v}, \sigma_2^{(g)}, P^{(g)}, S^{(g)}, \mathcal{M}_M) \right]$$

the expected log pointwise predictive density

bias for the effective number of parameters

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

The last measure of model diagnostics is the average loss of break points. The inclusion of redundant break points (e.g. imposing two breaks on a single break process) produces an instability in draws of hidden state variables. An easy way to check the existence of redundant breaks is to estimate average variances of simulated break points. This measure is equivalent to the average loss of break points assuming the simulation mean of break points ($\bar{\tau}_m$) as true break points:

$$\text{Average Loss} = \frac{1}{M} \sum_{m=1}^{M} \left( \frac{1}{G} \sum_{g=1}^{G} (\bar{\tau}_m - \tau_m^{(g)})^2 \right)$$

where $G$ is the MCMC simulation size and $M$ is the total number of breaks. The average loss is close to 0 if simulated break points are highly stable. Average Loss becomes larger if at least one of break points swings widely in each simulation.

4 Simulation Studies

In this section, we check the performance of the proposed method using simulated group-structured network data. We consider five different cases of group-structured network changes:

1. group-structured networks with no break (Table 1)
2. group-structured networks with a group-splitting break (Table 2)

3. group-structured networks with a group-merging break (Table 3 in Supplementary Material)\(^6\)

4. group-structured networks with a group-merging break followed by a group-splitting break (Table 3)

5. group-structured networks with a group-splitting break followed by a group-merging break (Table 4 in Supplementary Material)

4.1 Simulation Setup

Blocks in simulated data were generated by an assortative rule in which nodes belonging to the same group had a higher connection probability \((p_{in} = 0.5)\) than nodes belonging to different groups \((p_{out} = 0.05)\).\(^7\) In the group merging examples, two groups were merged so that the tie formation probability between the members of the two groups changed from \(p_{out}\) to \(p_{in}\). In the group splitting examples, an existing group split into two equal size groups so that the connection probability between the members of the two different groups 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 \(t = 10\) 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.

4.2 Block-structured Networks with No Break

We first check how sensitive our proposed method is to the false positive bias (i.e. detecting breaks when there is no break). We generate block-structured networks with no break.\(^8\) Table 1 summarizes the results from the no break

---

\(^6\)To save space, we report the simulation results of a group-merging break and of a group-splitting break followed by a group-merging break in Supplementary Material.

\(^7\)i.e. \(\forall b_i = b_j, p_{ij} = p_{in}\) and \(\forall b_i \neq b_j, p_{ij} = p_{out}\) where \(b_i\) and \(b_j\) are group labels for nodes \(i\) and \(j\) respectively. This simple formulation, with two difference values for the block diagonal connection probability and the off-diagonal connection probability, is called the planted partition model.

\(^8\)More specifically, the longitudinal network was generated from a two group structure, consisting of a 10 node group and a 20 node group respectively, and the underlying group structure remains constant.
and the underlying group structure is a two-group network.

Table 1: Simulation Results of Block-structured Networks with No Break. The ground truth is no break.

| Break number | Average Loss | WAIC |
|--------------|--------------|------|
| 1            | 0.05         | 14310 |
| 2            | 0.08         | 14589 |
| 3            | 0.05         | 14310 |
| 4            | 0.08         | 14589 |

Average Loss = 0.08

-2 log likelihood = 14310
-2 log marginal = 14435
WAIC = 14589

Break number = 1

Break number = 2

Break number = 3

Break number = 4

Latent Space of Break 1 (1−1)
Latent Space of Break 1 (1−8)
Latent Space of Break 2 (2−39)
Latent Space of Break 2 (9−31)
Latent Space of Break 2 (31−40)
Latent Space of Break 3 (38−38)
Latent Space of Break 3 (40−40)
Latent Space of Break 4 (37−40)
Latent Space of No Break Model

The ground truth is no break.
Table 2: Simulation Results of Block-structured Networks with a Block-splitting Break. The ground truth is

\[
\text{Average Loss} = 0.49
\]

\[
-2\log \text{likelihood} = 12810
\]

\[
-2\log \text{marginal} = 12922
\]

\[
\text{WAIC} = 13016
\]

| Break number | Break number | Break number | Break number |
|--------------|--------------|--------------|--------------|
| 1            | 2            | 3            | 0            |

| Latent Space of Break 1 (1−19) | Latent Space of Break 2 (13−20) | Latent Space of Break 3 (21−37) | Latent Space of Block-structured Networks |
|-------------------------------|---------------------------------|---------------------------------|-------------------------------------------|
| Time                         | Time                            | Time                            | Time                                      |
| V                             | V                               | V                               | V                                         |
| 0 10 20 30 40                | 0 1 2 3 4                       | 0 1 2 3 4                       | 0 1 2 3 4                                  |
| 1st 2nd                      | 1st 2nd                         | 1st 2nd                         | 1st 2nd                                   |

In the middle, we break and the underlying group structure changes from a two group structure to a three group structure.
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 ($M_1$), WAIC correctly shows that the no break model ($M_0$) fits the data best. Also, the average loss of break points shows somewhat unstable movements. On average, simulated break points of the one break model swing ±0.57 around the estimated break point and simulated break points of the two break model swing ±1.48 around the estimated break points.

If we look at the estimated latent space of the no break model, it correctly recovers the latent two-group structure while the one break model identifies another latent state almost identical to the first one. The approximate log marginal likelihood fails to penalize the recovery of the redundant latent state. However, as we add more breaks, the approximate log marginal likelihood successfully penalizes the model for the existence of redundant states in the two break model and in the three break model.

4.3 Block-structured Networks with a Block-splitting Break

Now, we move the case of dynamic network data with a single group-splitting break. That is, the ground truth is that the number of latent groups changes from 2 to 3 in the middle. Table 2 shows the results of the simulation. Again, we start to read from the model diagnostic results in the first column.

WAIC correctly identifies the single break model as the best-fitting model while the approximate log marginal likelihood favors the three break HMTM. As we have seen in the previous example, the approximate log marginal likelihood fails to penalize the model with redundant breaks. The source of the problem is the singleton state (a latent state consisting of a single observation). A similar problem has been noticed in finite mixture models with singular components (Hartigan, 1985; Bishop, 2006). In the three break model, the second state has only one observation, which increases the log likelihood dramatically. Since a singleton state is highly unlikely in reality, researchers can ignore false diagnostic results simply by checking the existence of singleton states.9 Interestingly, the network generation rule in the last column

---

9Chib (1995)’s algorithm is based on the summation of log likelihoods evaluated at posterior means and hence sensitive to the presence of singleton states in high dimensional time
shows almost identical patterns, regardless of the number of imposed breaks. Note that the second dimensional network generation rule \(v_2\) jumps to a large positive number in the middle as the number of groups increases from 2 to 3.

The average loss of break points clearly favors the one break model. Adding redundant breaks increases the average loss of break points significantly. For example, simulated break points of the three break model swing \(\pm 1.4\) around the estimated break points on average while simulated break points of the one break model stay constant.

### 4.4 Block-structured Networks with Two Breaks

Last, we check whether our proposed method correctly recovers more complicated network changes. The first break is planted at \(t = 10\) and it corresponds to a group-merging change. Another break is planted at \(t = 30\), which corresponds to a group-splitting change. Thus, the number of latent groups changes 3, 2, and 3. Table 3 reports the results of the two break test.

WAIC correctly detects \(\mathcal{M}_2\) as the best-fitting model while the approximate log likelihood favors \(\mathcal{M}_3\), the pattern of which is constant in our simulation. Again, the presence of a singleton state in the three break model is the source of the problem for the approximate log likelihood. The average loss of break points correctly favors the two break model. Fitting the one break model increases the average loss of break points because the latent state sampler falls into either of the breaks. In contrast, adding more than two breaks increases the average loss of break points significantly because the existence of a redundant break makes it difficult to pin down break points in simulations.

If we look the recovered latent states, the two break HMTM correctly recovers the two underlying changes between \(t = 10\) and \(t = 11\) (group merging) and between \(t = 30\) and \(t = 31\) (group 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, 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 \((\frac{1}{G} \sum_{g=1}^{G} p(B_{upper}^{(g)}, \Theta^{(g)}, P^{(g)}, \mathcal{M}_M))\), 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.
Table 3: Simulation Results of Block-structured Networks with Two Breaks. The ground truth is two breaks.

| Model Fit Latent Space | Break number | Average Loss | -2*log likelihood | -2*log marginal | WAIC |
|------------------------|--------------|--------------|--------------------|-----------------|------|
| Latent Space of Break 1 (1−10) | 1 | 0.00 | 12892 | 13043 | 13583 |
| Latent Space of Break 2 (11−30) | 0 | 0.27 | 13574 | 13630 | 13727 |
| Latent Space of Break 3 (31−40) | 2 | 1.59 | 13028 | 13054 | 13431 |
| Latent Space of Break 4 (31−40) | 3 | ∗ | 13030 | 13057 | 13758 |

The ground truth is two breaks.
when the number of groups 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 clearly show that our proposed model and multiple metrics for model diagnostics work very well in (1) correctly identifying the number of breaks, (2) recovering latent group structure changes, and (3) identifying state-specific latent node positions. The approximate log marginal likelihood performs well where there is no singleton state while WAIC performs steadily regardless of the existence of singleton states. The average loss also shows steady performance, signaling the existence of redundant states in a tested model.

5 Applications

In this section, we apply our proposed method to the analysis of structural changes in military alliance networks. 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 unknown structural changes in longitudinal network data.

To illustrate our method, we start from a simple example using a small data set, consisting of seven “major powers” (Austria-Hungary, France, Germany, Italy, Japan, Russia, and the United Kingdom) from 1816 to 1938 (Gibler, 2008). We aggregated every 2 year network to increase the density of each layer. Changing the granularity of aggregation does not change the substantive findings. These seven major powers are main players of the balance of power system in Europe during the 19th century and two world wars in the 20th century. Also, the period from 1816 to 1938 corresponds to the era of shifting alliances among major powers. Thus, the structure of alliance among major powers will clearly display how the distribution of military power in the international system changes over time. Then, we apply our method to a larger data set of 104 postwar states after removing isolated nodes.
5.1 Major Power Alliance Network Changes, 1816 - 1938

Figure 3 shows the model diagnostic results for the major power alliance data set. We dropped the results for the models with more than 3 breaks as they show strong signs of non-convergence, which indicate the existence of redundant states. All metrics of model diagnostics point to the two break model as the best-fitting model. In particular, the average loss of break points significantly drops in the two break model.

Figure 4 visualizes changes in latent node positions of major powers (top) and changing patterns of the major-power network topology (bottom) from the two break model. Node colors (online) indicate clusters of each node using the k-means clustering method. Regime-specific network generation rule parameters $v_{rt}$ are reported in axis labels. Several substantive findings are noteworthy.

The first notable finding is the centrality of Austria-Hungary, connecting groups of major powers, between 1816 and 1890. This period includes what historians call “the age of Metternich” (1815-1848) (Rothenberg, 1968). After the end of Napoleonic Wars, Chancellor of Austria-Hungary played an important role in maintaining the European balance of power system. The first dimension of Regime 1 clearly distinguishes Austria-Hungary from the other major powers. In Regime 2 (1854-1890), Germany challenged the position of Austria-Hungary. However, throughout Regime 1 and Regime 2, the network position of Austria-Hungary remained highly critical in the sense that

---

10 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.
Figure 4: Changing Node Positions and Network Topology of Military Alliance Networks Among Major Powers, 1816 - 1938: Node colors (online) indicate clusters. 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. Included states are Austria-Hungary (AUT), France (FRA), Germany (GMY), Italy (ITA), Japan (JAP), Russia (RUS), and the United Kingdom (GBR).
the removal of Austria-Hungary would have made the major power alliance network completely 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).

The second notable finding is the timing of the first break between 1852 and 1854. This break point coincides with the outbreak of the Crimean War. In this war, Russia was defeated by the united powers of Britain, France, Austria-Hungary, and Prussia (Germany). The rise of Germany led by Otto von Bismarck and the defeat of Russia marked the first break in the balance of power system.

The third notable finding is the timing of the second break between 1890 and 1892. Scholars of international relations and historians consider the formation of the Dual Alliance between Germany and Austria-Hungary in 1879 and a sequence of alliances that followed it as a structural change in the balance of power system (Snyder, 1997; Vermeiren, 2016). These series of events transformed a web of shifting alliances into a clearly diverged group structure, consisting of two clusters: Austria-Hungary, Germany, and Italy on the one hand and France, Russia, and the United Kingdom on the other. The network diagram of the third regime (bottom-right) shows members of the two clusters, which formed each side of belligerents in World War I.

5.2 Postwar Alliance Network Changes, 1946 - 2012

Now, we focus on the postwar period in which the number of states and alliance links among them exploded. After removing isolated nodes, our sample contained 104 states. Figure 5 shows the results of the model diagnostics. Although the two break model is preferred by the approximate log marginal likelihood, the average loss of break points and WAIC favor the one break model. The estimated break point is between 1978 and 1980.

Figure 6 shows changes in latent node positions (top) and changing patterns of the network topology (bottom). What we can see is that the number of latent groups does not change during the postwar period. What changed

---

11First, 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.
is the strength of the connection among groups of nodes, denoted by $v_{rt}$ in each axis. Both numbers in the $x$ and $y$ axes are large and positive, implying that the coalition structure of postwar military alliance networks is multidimensional with a sign of a strong homophily. However, after 1980, the strength of the within-group connection rule became weaker as more countries are connected with outside group members. This reflects the fact that the coalition structure of postwar military alliance networks was highly dense in the beginning of the Cold War. The Cold War division became less important over time due to the détente and the rise of the Non-Aligned Movement. The results suggest that the structural change in the postwar military alliance network came earlier than the collapse of the Soviet Union.

6 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 (Rohé 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.
Figure 6: Changing Node Positions and Network Topology of Postwar Military Alliance Networks, 1946 - 2012: Node colors (online) indicate clusters. Regime averages of $v_1$ values for each dimension are reported in the axis (top panel). Line widths (bottom panel) are proportional to the duration of alliance links.
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. One difficulty in adding covariates in hidden Markov models is to avoid the endogeneity between state transition and effects of covariates (Kim et al., 2008).
MCMC Algorithm for Hidden Markov Tensor Model

For each $t$ layer, generate $B_t = Y_t - \Omega_t$ by choosing a null model ($\Omega_t$).

Set the total number of changepoints $M$ and initialize $(U, \mu_u, \Psi_u, V, \mu_v, \Psi_v, \beta, \sigma^2, S, P)$.

Part 1

Step 1 The sampling of regime specific $U, \mu, \Psi_u$ 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. $p(\psi_{r,u,m} | B, P, S, \Theta^{-\Psi_{u,m}}) \propto IG \left( \frac{\nu_u + N}{2}, \frac{U_{r,m}^T U_{r,m} + u}{2} \right)$.
2. $p(\mu_{u,m} | B, P, S, \Theta^{-\mu_{u,m}}) \propto \text{multivariate normal} \left( \frac{U_{m}^T 1}{(N + 1)}, \frac{\Psi_{u,m}}{(N + 1)} \right)$.
3. $p(U_m | B, P, S, \Theta^{-U_m}) \propto \text{matrix normal} \left( \tilde{M}_{u,m}, I, \tilde{\Psi}_{u,m} \right)$ where 
   \[
   \tilde{\Psi}_{u,m} = \frac{Q_{u,m}/\sigma^2_{m} + \Psi_{u,m}^{-1}}{2},
   \tilde{M}_{u,m} = (L_{u,m}/\sigma^2_{m} + 1\mu_{u,m}^T \Psi_{u,m}^{-1}) \tilde{\Psi}_{u,m},
   Q_{u,m} = (U_{m}^T U_{m} \circ (V_{m}^T V_{m})),
   L_{u,m} = \sum_{j,t: t \in S_{i} = m} b_{i,j,t} \otimes (U_{m,i} \circ V_{m,j,t})
   \]

4. Orthogonalize $U_m$ using the Gram-Schmidt algorithm.

Step 2 The sampling of $V, \mu_v, \Psi_v$ 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. $p(\psi_{r,v,m} | B, P, S, \Theta^{-\Psi_{v,m}}) \propto IG \left( \frac{\nu_v + T}{2}, \frac{V_{r,m}^T V_{r,m} + v}{2} \right)$.
2. $p(\mu_{v,m} | B, P, S, \Theta^{-\mu_{v,m}}) \propto \text{multivariate normal} \left( \frac{V_{m}^T 1}{(T + 1)}, \frac{\Psi_{v,m}}{(T + 1)} \right)$.
3. $p(V_m | B, P, S, \Theta^{-V_m}) \propto \text{matrix normal} \left( \tilde{M}_{v,m}, I_{T_m}, \tilde{\Psi}_{v,m} \right)$ where 
   \[
   \tilde{\Psi}_{v,m} = \frac{Q_{v,m}/\sigma^2_{m} + \Psi_{v,m}^{-1}}{2},
   \tilde{M}_{v,m} = (L_{v,m}/\sigma^2_{m} + 1\mu_{v,m}^T \Psi_{v,m}^{-1}) \tilde{\Psi}_{v,m},
   Q_{v,m} = (U_{m}^T U_{m} \circ (U_{m}^T U_{m})),
   L_{v,m} = \sum_{i,j} b_{i,j} \otimes (U_{m,i} \circ U_{m,j})
   \]

31
Step 3  The sampling of $\beta$ from $\mathcal{N}(b_1, B_1)$ where

$$B_1 = (B_0^{-1} + \sum_{m=1}^{M} \sigma_m^{-2} N^2 \mathbf{1}(S = m))^{-1}$$

$$b_1 = B_1 \times \left( B_0^{-1} b_0 + \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} b_{i,j,t} - \mu_{i,j,t} \right).$$

$\mathbf{1}(S = m)$ is the number of time units allocated to state $m$ and $\mu_{i,j,t}$ is an element of $U_S \Lambda_t U_T^T$.

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

Part 2

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

$$p(S_0, \ldots, S_T | \Theta, \mathcal{E}, P) = p(S_T | \Theta, \mathcal{E}, P) \cdots p(S_t | S_{t+1}, \Theta, \mathcal{E}, P) \cdots p(S_0 | S_1, \Theta, \mathcal{E}, P).$$

Using Bayes’ Theorem, Chib (1998) shows that

$$p(S_t | S_{t+1}, \Theta, \mathcal{E}, P) \propto \frac{p(S_t | \Theta, B_{1:t}, P)}{p(S_{t+1} | S_t, P)}.$$

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 ($P$). 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 $S$ has a state with single observation, randomly sample $S$ with replacement using a pre-chosen perturbation weight ($w_{\text{perturb}} = (w_1, \ldots, w_M)$).

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

Step 6  Sample each row of $P$ from the following Beta distribution:

$$p_{kk} \sim \text{Beta}(a_0 + j_{k,k} - 1, b_0 + j_{k,k+1})$$

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 $k+1$.
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:

\[
\log p(\mu^*_u, \psi^*_{...u}, \mu^*_v, \psi^*_{...v}, \beta^*, \sigma^{2*}, P^*|B) = \log p(\mu^*_u|B) + \sum_{r=1}^{R} \log p(\psi^*_{r,u}|B, \mu^*_u) + \log p(\mu^*_v|B, \mu^*_u, \psi^*_{...u}, \mu^*_v) + \log p(\beta^*|B, \mu^*_u, \psi^*_{...u}, \mu^*_v, \psi^*_{...v}) + \log p(\sigma^{2*}|B, \mu^*_u, \psi^*_{...u}, \mu^*_v, \psi^*_{...v}, \beta^*) + \log p(P^*|B, \mu^*_u, \psi^*_{...u}, \mu^*_v, \psi^*_{...v}, \beta^*, \sigma^{2*}).
\]

Let \( \Theta \) indicate a parameter vector beside hidden states (S) and a transition matrix (P): \( \Theta = \{ \mu_u, \psi_{...u}, \mu_v, \psi_{...v}, \beta, \sigma^2 \} \). Let \((\Theta^*, P^*)\) be posterior means of \((\Theta, P)\). Using Chib (1995)'s formula to compute the approximate log marginal likelihood,

\[
p(\Theta^*, P^*|B) = \frac{p(B|\Theta^*, P^*)p(\Theta^*, P^*)}{m(B)}
\]

\[
m(B) = \frac{p(B|\Theta^*, P^*)p(\Theta^*, P^*)}{p(\Theta^*, P^*|B)}
\]

\[
\log m(B) = \log p(B|\Theta^*, P^*) + \log p(\Theta^*, P^*) - \log p(\Theta^*, P^*|B).
\]

The quantities in the right hand side of Equation (30) can be computed by Chib (1995)'s candidate formula:

Step 1

\[
p(\mu^*_u|B) \approx \int p(\mu^*_u|B, \psi_{...u}, \mu_v, \psi_{...v}, \beta, \sigma^2, P, S)dp(\psi_{...u}, \mu_v, \psi_{...v}, \beta, \sigma^2, P, S|B)
\]

Step 2

\[
p(\psi^*_{r,u}|B, \mu^*_u) \approx \int p(\psi^*_{r,u}|B, \mu^*_u, \mu_v, \psi_{...v}, \beta, \sigma^2, P, S)dp(\mu_v, \psi_{...v}, \beta, \sigma^2, P, S|B)
\]

33
Step 3
\[ p(\mu_u^*, \mu_v^*, \psi_{\cdot,u}^*) \approx \int p(\mu_u^*, \mu_v^*, \psi_{\cdot,u}^*, \psi_{\cdot,v}^*, B, \mu_u^*, \mu_v^*, \beta, \sigma^2, P, S)dp(\psi_{\cdot,u}^*, \beta, \sigma^2, P, S|B) \]

Step 4
\[ p(\psi_{\cdot,u}^*, B, \mu_u^*, \mu_v^*, \sigma^2) \approx \int p(\psi_{\cdot,u}^*, B, \mu_u^*, \mu_v^*, \beta, \sigma^2, P, S)dp(\beta, \sigma^2, P, S|B) \]

Step 5
\[ p(\beta^* | B, \mu_u^*, \psi_{\cdot,u}^*, \mu_v^*, \psi_{\cdot,v}^*) \approx \int p(\beta^* | B, \mu_u^*, \psi_{\cdot,u}^*, \mu_v^*, \psi_{\cdot,v}^*, \sigma^2, P, S)dp(\beta^*, \sigma^2, P, S|B) \]

Step 6
\[ p(\sigma^2 | B, \mu_u^*, \psi_{\cdot,u}^*, \mu_v^*, \psi_{\cdot,v}^*, \beta^*) \approx \int p(\sigma^2 | B, \mu_u^*, \psi_{\cdot,u}^*, \mu_v^*, \psi_{\cdot,v}^*, \beta^*, P, S)dp(\sigma^2, P, S|B) \]

Step 7
\[ p(P^* | B, \mu_u^*, \psi_{\cdot,u}^*, \mu_v^*, \psi_{\cdot,v}^*, \beta^*, \sigma^2) \approx \int p(P^* | B, \mu_u^*, \psi_{\cdot,u}^*, \mu_v^*, \psi_{\cdot,v}^*, \beta^*, \sigma^2, S)dp(S|B) \]
Sampling of HMTM with a Student-\(t\) Distributed Error

We modify the sampling of \(U\) based on a Student-\(t\) distributed error as follows:

1. \[ p(U_m|B, P, S, \Theta^{-U_m}) \propto \text{matrix normal}_{N \times R}(M_{u,m}, I_N, \tilde{\Psi}_{u,m}) \]

where

\[
\begin{align*}
\tilde{\Psi}_{u,m} & = (Q_{u,m}/\sigma_m^2 + \Psi_{u,m}^{-1})^{-1} \\
M_{u,m} & = (L_{u,m}/\sigma_m^2 + 1\mu_{u,m}^T\Psi_{u,m}^{-1})\tilde{\Psi}_{u,m} \\
Q_{u,m} & = (U_m^TU_m) \odot (V_m^T\Gamma_m V_m) \\
L_{u,m} & = \sum_{j,t: t \in S_t=m} b_{i,j,t} \otimes (U_{m,j} \odot (V_{m,t} \gamma_t))
\end{align*}
\]

where \(\Gamma_m\) is a \(T_m\) by \(T_m\) diagonal matrix with \(\gamma_t\) corresponding to regime \(m\).

Next, the sampling of \(\sigma_m^2\) can be done using \(IG\left(\frac{c_0+N_m-N_m \cdot T_m}{2}, \frac{d_0+\sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \gamma_t(b_{i,j,t} - \beta - \mu_{i,j,t})}{2}\right)\).

Then, the draws of \(\gamma_t\) are obtained from \(G(\nu_2/2, \nu_3,t/2)\) where

\[
\begin{align*}
\nu_2 & = \nu_0 + 1 \\
\nu_3,t & = \nu_1 + \sigma_s^{-2} \left( \sum_{i=1}^N \sum_{j=1}^N b_{i,j,t} - \beta - \mu_{i,j,t} \right)
\end{align*}
\]
Figure 7: Block Label Recovery from Block-Merging-Splitting Changes: $v[1]$ indicates the first element of regime-averaged network generation rule parameter and $v[2]$ is the second element of the same parameter. Node color and labels indicate the planted group membership.

Figure 7 demonstrates the performance of discrete block label recovery using a block-merging-splitting example with $T = 40$. Panel (a) shows the planted block structures at the beginning of each regime and panel (b) shows recovered latent node positions and their block memberships that are identified by Hartigan and Wong (1979)'s $k$-means clustering algorithm. The two break HMTM correctly recovers the two planted changes at $T = 11$ (block merging) and $T = 31$ (block splitting). When block $a$ and block $b$ are merged into one block in Regime 2, the second dimension becomes redundant and the second element of the regime-averaged network generation rule ($v_2$) becomes small (0.98). In Regime 3, the number of blocks increases back to 3 and $v_2$ returns to its previous level at Regime 1.

Block Structure Recovery from Latent Space Estimates

Figure 7 demonstrates the performance of discrete block label recovery using a block-merging-splitting example with $T = 40$. Panel (a) shows the planted block structures at the beginning of each regime and panel (b) shows recovered latent node positions and their block memberships that are identified by Hartigan and Wong (1979)'s $k$-means clustering algorithm. The two break HMTM correctly recovers the two planted changes at $T = 11$ (block merging) and $T = 31$ (block splitting). When block $a$ and block $b$ are merged into one block in Regime 2, the second dimension becomes redundant and the second element of the regime-averaged network generation rule ($v_2$) becomes small (0.98). In Regime 3, the number of blocks increases back to 3 and $v_2$ returns to its previous level at Regime 1.
Further Simulation Results

In this section, we report additional simulation results that are not reported in the manuscript due to space limitation. First, we report simulation results from a block-merging network change. Then, we report simulation results from block-splitting-merging network changes. The structure of the reported tables is identical the ones discussed in the manuscript.

Results of a Block-Merging Network Change

Table 4 summarizes the simulation results from the block-merging change example. The ground truth is the single break HMTM in the third row. WAIC correctly identifies the single break model as the best-fitting model while the approximate log marginal likelihood favors the three break HMTM. The one break model ($M_1$) in Table 4 correctly recovers the block-merging latent change and the second dimensional network generation rule ($v_2$) drops from a positive number to 0 in the middle as the number of blocks decreases from 3 to 2.
Table 4: Simulation Analysis of Block Merging MTRM. The ground truth is one break. All setup is adapted to the simulation examples in the main text. Two 10 node blocks were merged into a single block.
Results of Block-Splitting-Merging Network Changes

Table 5 summarizes the simulation results from the block-splitting-merging change example. The ground truth is the two break HMTM (\( M_2 \)). WAIC correctly identifies \( M_2 \) as the best-fitting model while the approximate log marginal likelihood favors \( M_3 \). Note that Regime 3 of \( M_3 \) has only two observations, which inflates the approximate log marginal likelihood. The results of \( M_2 \) correctly recovers the block-splitting-merging change in the latent node positions and network generation rules.
Table 5: Simulation Analysis of Block Split-Merging MTRM. The ground truth is two breaks. All parameters are analogous to the example depicted in Table 3 of the main text except the order of the changes.

| Regime 1 | Regime 2 | Regime 3 | Regime 4 | Latent Space (Un) Changes | Break number | \(-2\)log likelihood | WAIC |
|----------|----------|----------|----------|--------------------------|--------------|----------------------|-------|
|          |          |          |          |                           | 0            | 12714                |       |
|          |          |          |          |                           | 1            | 12806                |       |
|          |          |          |          |                           | 2            | 13366                |       |

Average Loss = 0.00
Average loss = 0.00
Average loss = 0.00
Average loss = 0.00
Average loss = 0.00

\(-2\)log likelihood = 12714
\(-2\)log marginal = 12946
\(-2\)log likelihood = 12806
\(-2\)log marginal = 12946
\(-2\)log likelihood = 13366
\(-2\)log marginal = 13648

WAIC = 13034
WAIC = 13031
WAIC = 13608
WAIC = 13744
WAIC = 13608
WAIC = 13744

Break number = 2
Break number = 3
Break number = 2
Break number = 0
Break number = 2
Break number = 0
Software Implementation

```r
require(NetworkChange)
set.seed(1973)

## Generate an array (30 by 30 by 40) with block transitions from 2 blocks to 3 blocks
Y <- MakeBlockNetworkChange(n=10, T=40, type ="split")
G <- 100 ## only 100 mcmc scans to save time

## Fit models
out0 <- NetworkStatic(Y, R=2, mcmc=G, burnin=G, verbose=G, Waic=TRUE)
out1 <- NetworkChange(Y, R=2, m=1, mcmc=G, burnin=G, verbose=G, Waic=TRUE)
out2 <- NetworkChange(Y, R=2, m=2, mcmc=G, burnin=G, verbose=G, Waic=TRUE)
out3 <- NetworkChange(Y, R=2, m=3, mcmc=G, burnin=G, verbose=G, Waic=TRUE)
outlist <- list(out0, out1, out2, out3)

## The true model is out1
WaicCompare(outlist)
MarginalCompare(out)
BreakPointLoss(list(out1, out2, out3))[[1]]

## plot latent node positions
plotU(out1)

## plot layer-specific network generation rules
plotV(out1)
```

Figure 8 shows the output of `plotU(out1)`, which is the regime-specific latent node positions and Figure 9 shows the output of `plotV(out1)`, which is time-varying network generation rules.
Figure 8: Software Outputs of \texttt{plotU(out1)}

Figure 9: Software Outputs of \texttt{plotV(out1)}
References

Barnett, I. and Onnela, J.-P. (2016). Change point detection in correlation networks. Scientific Reports, 6:18893 EP –.

Bartolucci, F., Marino, M. F., and Pandolfi, S. (2018). Dealing with reciprocity in dynamic stochastic block models. Computational Statistics & Data Analysis, 123:86 – 100.

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.

Borgatti, S. P. and Everett, M. G. (1999). Models of core-periphery structures. Social Networks, 21(4):375–395.

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.

Carlin, B. P. and Polson, N. G. (1991). Inference for non-conjugate bayesian models using the gibbs sampler. Canadian Journal of Statistics, 19:399–405.

Chaudhuri, K., Graham, F. C., and Tsiatas, A. (2012). Spectral clustering of graphs with general degrees in the extended planted partition model. In COLT, volume 23, pages 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.

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., 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. 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*. Springer Verlag, Heidelberg.
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 24th International Conference on Machine Learning*, pages 321–328.

Han, S. and Dunson, D. B. (2018). Multiresolution tensor decomposition for multiple spatial passing networks. *CoRR*, abs/1803.01203.

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 LeCam, L. and Olshen, R. A., editors, *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, volume 2, pages 807–810, Belmont, California. Wadsworth Statistics/Probability Series.

Hartigan, J. A. and Wong, M. A. (1979). Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108.

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 Platt, J., Koller, D., Singer, Y., and Roweis, S., editors, *Advances in Neural Information Processing Systems 20*, pages 657–664. Cambridge University Press.
Hoff, P. D. (2009a). Multiplicative latent factor models for description and prediction of social networks. *Computational & Mathematical Organization Theory, 15*(4):261–272.

Hoff, P. D. (2009b). Simulation of the matrix bingham-von mises-fisher distribution, with applications to multivariate and relational data. *Journal of Computational and Graphical Statistics, 18*(2):438 – 456.

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.

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

Johndrow, J. E., Bhattacharya, A., and Dunson, D. B. (2017). Tensor decompositions and sparse log-linear models. *Annals of Statistics, 45*(1):1–38.

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

Kim, C.-J., Piger, J., and Startz, R. (2008). Estimation of markov regime-switching regression models with endogenous switching. *Journal of Econometrics, 143*(2):263–273.

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.
Minhas, S., Hoff, P. D., and Ward, M. D. (2016). A new approach to analyzing coevolving longitudinal networks in international relations. *Journal of Peace Research, 53*(3):491–505.

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.

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.

Rai, P., Wang, Y., and Carin, L. (2015). Leveraging features and networks for probabilistic tensor decomposition. *AAAI’15 Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, pages 2942–2948.

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*, pages 41–71. Routledge.

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.

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.

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.

Wasserman, S. and Faust, K. (1994). *Social network analysis: Methods and applications*, volume 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.

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.

Xu, K. S. (2015). Stochastic block transition models for dynamic networks. *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics*, pages 1079–1087.

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.