DETECTING STRUCTURAL CHANGES IN LONGITUDINAL NETWORK DATA

BY JONG HEE PARK∗,†, AND YUNKYU SOHN‡

Seoul National University† and Princeton University‡

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

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

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

∗Corresponding Author

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

†For a range of examples for longitudinal network analysis, see Holme and Saramäki (2012) and references therein.
structure between dyadic observations over time. Importantly, it allows us to jointly estimate time-specific network generation rules with time-constant latent node positions. Let \( U = (u_1, \ldots, u_N)^T \in \mathbb{R}^{N \times R} \) 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 network generation rules \( v_t \) at time \( t \) or \( t \)th layer as follows:

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

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

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

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

The resulting estimates of node-specific latent variables recover a specific type of similarity between nodes, so that nodes with similar connections profiles are likely to have similar values (Hoff, 2008). If \( u_i \) and \( u_k \) exhibit similar values, they will have similar inner product outcomes with node \( j \)'s latent position vector \( u_j \), corresponding to the notion of stochastic equivalence in network theory (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 same sign.

In this paper, we develop a hidden Markov multilinear tensor model (HMTM) that extends Hoff (2011)'s MTRM. HMTM combines MTRM with a hidden Markov model (HMM) using Bayesian inference. Conventional approaches to dynamic network modeling usually extend a framework of static networks by assuming smooth topological changes over time (Robins and Pattison, 2001; Hanneke et al., 2010; Desmarais and Cranmer, 2012; Snijders et al., 2006, 2010; Westveld and Hoff, 2011; Ward et al., 2013). However, longitudinal network datasets typically show irregular dynamics, implying multiple changes in their data generating parameters. For example, a human brain network study shows that functional connectivity networks between
brain regions are highly non-stationary, consisting of many latent states (Cribben et al., 2013). In the proposed method, HMTM, latent node positions follow a discrete Markov process in which parameters are allowed to change multiple times whenever there is a symptom of fundamental change in the latent traits. As shown in other applications (Baum et al., 1970; Chib, 1998; Robert et al., 2000; Cappe et al., 2005; Scott et al., 2005; Frühwirth-Schnatter, 2006; Teh et al., 2006), the conditional independence assumption in HMM allows us to model each local regime using a well-known parametric model. Suppose $\Theta$ be a collection of parameters that represent a network generating process of a longitudinal network $Y$. Then, the model is decomposed as

$$p(Y|\Theta) = \int p(S_1|\Theta)p(Y_1|S_1, \Theta) \prod_{t=2}^{T} \sum_{m=1}^{M} p(Y_t|\Theta_m) Pr(S_t = m|S_{t-1}, \Theta) dS,$$

where $p(Y_t|\Theta_m)$ a generative network model at regime $m$ and $S = (S_1, \ldots, S_T)$ are discrete hidden states. In our model, the duration of hidden state $m$ follows a geometric distribution of $1 - p_{mm}$ where $p_{mm}$ is the $m$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. $\frac{1}{G} \sum_{g=1}^{G} \mathcal{I}(S^{(g)}_t \neq S^{(g)}_{t-1})$).

We introduce a degree correction formulation (Karrer and Newman, 2011; Chaudhuri et al., 2012) to account for degree heterogeneity that may hinder stochastic equivalence trait recovery. We will show that the degree correction method makes a crucial difference in the recovery of group-structures in longitudinal network data, controlling for factors coming from disturbance other than the data generating parameters. We also discuss model diagnostics of HMTMs using the approximate log marginal likelihood (Chib, 1995) and the Watanabe-Akaike Information Criterion (WAIC) (Watanabe, 2010). Although the proposed method can be used to analyze longitudinal network datasets from various fields, we would like to highlight that HMTM has a strong substantive appeal to the fields where abrupt changes in the structural properties of networks have important substantive implications. For example, many historical questions in the social sciences are deeply related to the identification of distinct periods or critical events (Abbott, 2001; Mahoney and Rueschemeyer, 2003; Pierson, 2004). For this goal, social scientists increasingly resort to HMMs and related models to identify hidden regimes from historical time series data (Hamilton, 1989; Kim and Nelson, 1998; Western and Kleykamp, 2004; Spirling, 2007; Park, 2011; Pang et al.,
While network is one of the most prevalent forms of dataset in social sciences, the lack of methodological tools for the identification of critical regimes prevents scholars from using longitudinal network dataset to answer such important questions in their fields.

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

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

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

Guo et al. (2007) and Wang et al. (2014) model network change-points by allowing shifts of parameter values in ERGMs. However, both models exhibit computational inefficiency, in which the maximum size of network analyzed was 11 nodes (Guo et al., 2007) and 6 nodes respectively (Wang et al., 2014). Ridder et al. (2016) use stochastic block models (SBM) as a probabilistic model for network generation of each time-specific network and identify the
existence of a single break by comparing the bootstrapped distribution of the log-likelihood ratio between a null model (a SBM without a break) and an alternative (a SBM with a break). However, the asymptotic distribution of a SBM with a break approaches to a mixture of $\chi^2$-distributions and hence a log-likelihood ratio test statistic used by Ridder et al. (2016) does not meet the regularity condition (Drton, 2009).

2. The Proposed Method. Our task of developing a dynamic network model for structural changes must start from the question of “What constitutes structural changes in networks?” On the one hand, one can think of a change in summary statistics of 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 microscopic features of networks (e.g. community structures, stochastic blocks, core-periphery structures) (Nowicki and Snijders, 2001; Newman, 2006; Fortunato, 2010; Tibély et al., 2011; Sporns, 2014). Technically, various approaches for microscopic 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 microscopic features of networks. We support this claim by using synthetic examples and show that this perspective is effective enough to recover fundamental aspects of changes in network generation.

2.1. Latent Space Approach with Degree-correction. We take the latent space approach in which the proximity of nodes in the latent space implies similar connecting profiles (Hoff et al., 2002; Hoff, 2008). The latent space becomes a stochastic equivalence space in terminology of social network analysis (Wasserman and Faust, 1994). In particular, we use a MTRM as a baseline for dynamic network modeling. However, the formulation of MTRM in Equation (1) entails a critical weakness in uncovering microscopic network features because there is no treatment to account for degree heterogeneity except exogenous covariates ($x_{i,j,t}$). The implicit assumption in this latent model is that the expected degree of nodes having a similar role (i.e. proximal in the latent space) is identical while the distribution of degree in empirical networks is highly heterogeneous and skewed following power law or expo-
ential distributions (Clauset et al., 2009). This problem is well known in the network science literature and various degree-correction methods have been proposed in the literature of network block discovery (Newman, 2006, 2010; Karrer and Newman, 2011; Chaudhuri et al., 2012; Zhao et al., 2012).

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

\[
\omega_{ijt} = \lambda_{t}^{\text{princ}} \mathbf{u}_{it} \tilde{\mathbf{u}}_{jt}^T,
\]

where \( \mathbf{u}_{it} \) is the \( i \)th row of the associated eigenvector.\(^2\) The planted three group structure is well recovered in panel (B).

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

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

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

Fig 1: Degree Correction for Group Structure Discovery: A 90 node undirected network with 3 communities is generated using the within-group link probability of 0.5 and the between-group link probability of 0.05.
2.2. Hidden Markov Multilinear Tensor Model. In a HMTM with \( k \) breaks, the probability distribution of a degree-corrected symmetric temporal network data (\( B_t = Y_t - \Omega_t \)) is modeled as a Markov mixture of \( k \) MTRMs in which hidden states move forward (Chib, 1998). As shown by Chib (1998) and Park (2012), multiple change-point problems are equivalent to the estimation of a nonergodic (or forward-moving) HMM, which has advantages in latent state identification and parameter estimation, thanks to the order constraint in latent states. Let us denote \( S_t \) as a hidden state variable and \( P \) as a \( (k+1) \times (k+1) \) transition matrix where \( p_{i,i} \) is the \( i \)th diagonal element of \( P \). Then, a HMTM is

\[
B_t = Y_t - \Omega_t \tag{7}
\]

\[
B_t = \beta J_N + U_{S_t} V_t U_{S_t}^T + E_t \tag{8}
\]

\[
S_t | S_{t-1}, P \sim \text{Markov}(P, \pi_0) \tag{9}
\]

\[
p_{i,i} \sim \text{Beta}(a_0, b_0) \tag{10}
\]

\[
E_t \sim \mathcal{N}_{N \times N}(0, \sigma_{S_t}^2 I_N, I_N),
\]

where \( \pi_0 \) is the initial probability (\( \pi_0 = (1, 0, \ldots, 0) \)), \( J_N \) is \( N \times N \) all-ones matrix. \( \beta J_N \) 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 \( U_m \) using the Gram-Schmidt process (Björck, 1996; Guhaniyogi and Dunson, 2015) in each simulation step. Hoff (2011)’s hierarchical scheme centers rows of \( U_m \) around its global mean (\( \mu_{u,m} \)) using a multivariate normal distribution. This does not guarantee the orthogonality of each latent factor in \( U_m \). The lack of orthogonality makes the model unidentified, causing numerical instability in parameter estimation and model diagnostics (Murphy, 2012; Guhaniyogi and Dunson, 2015).

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

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

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

The use of inverse-Wishart distribution for the prior distribution of a variance parameter (\( \Psi_{u,m}, \Psi_{v,m} \)) comes at a great cost because choosing informative inverse-Wishart prior distributions for \( \Psi_{u,m} \) and \( \Psi_{v,m} \) is not easy.
(Chung et al., 2015) and a poorly specified inverse-Wishart prior distribution has serious impacts on the marginal likelihood estimation. In our trials, the log posterior inverse-Wishart density of \( \Psi_{u,m} \) and \( \Psi_{v,m} \) often goes to a negative infinity, failing to impose proper penalties. In HMTM, the off-diagonal covariance of \( U_m \) 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.

Prior distributions for each row vector of regime specific latent positions (\( U_m \)) are multivariate normal distributions conditioned on a variance parameter (\( \Psi_{u,m} \)): \[ \{u_{1,m}, \ldots, u_{N,m}\} \sim N_R(\mu_{u,m}, \Psi_{u,m}) \] and \( \mu_{u,m} \sim N(\mu_{0,u,m}, \Psi_{u,m}) \). Similarly, prior distributions for each row vector of \( V \) are also multivariate normal distributions on a variance parameter (\( \Psi_{v,m} \)): \[ \{v_{1,m}, \ldots, v_{T,m}\} \sim N_R(\mu_{v,m}, \Psi_{v,m}) \] and \( \mu_{v,m} \sim N(\mu_{0,v,m}, \Psi_{v,m}) \). The prior distribution for regime specific error variances (\( \sigma^2_m \)) is an inverse gamma distribution (\( IG(c_0/2, d_0/2) \)).

The prior distribution for transition probabilities is the Beta distribution (\( Beta(a, b) \)) and the prior distribution for \( \beta \) is a normal distribution (\( N(\beta_0, \Sigma_0) \)).

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

\[
p(\Theta, P, S | B) \propto \prod_{m=1}^{M} \left( N_{N \times N}(B_1 | B_{t-1}, \Theta_{S_t})p(S_t | S_{t-1}, P) \right)
\]

\[
\times \prod_{m=1}^{M} \left( N_{R}(\mu_{u,m}, \Psi_{u,m})N_{R}(\mu_{v,m}, \Psi_{v,m}) \right)
\]

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

\[
\times \prod_{m=1}^{M} \left( IG(\sigma^2_m | c_0, d_0)Beta(p_{mm} | a, b) \right)N(\beta | b_0, B_0)
\]

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):

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

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

$$\log \hat{p}(B_{\text{upper}} | M_M) = \log p(B_{\text{upper}} | \mu^*_u, \psi^*_u, \mu^*_v, \psi^*_v, \beta^*, \sigma^{2*}, P^*, M_M)$$

the log likelihood

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

the log prior density of posterior means

$$- \sum_{m=1}^{M} \log p(\mu^*_{u,m}, \psi^*_{u,m}, \mu^*_{v,m}, \psi^*_{v,m}, \beta^*, \sigma^{2*}_m, p^*_m, | B_{\text{upper}}, M_M)$$

the log posterior density of posterior means

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

$$\sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=i+1}^{N-1} \sum_{m=1}^{M} p(b_{i,j,t} | B_{t-1}^{\text{upper}}, \mu^*_{u,m}, \psi^*_{u,m}, \mu^*_{v,m}, \psi^*_{v,m}, \beta^*, \sigma^{2*}_m, P^*, S_t = m, M_M)$$

$$p(S_t = m | B_{t-1}^{\text{upper}}, \mu^*_{u,m}, \psi^*_{u,m}, \mu^*_{v,m}, \psi^*_{v,m}, \beta^*, \sigma^{2*}_m, P^*, M_M).$$

The second measure of model uncertainty is WAIC (Watanabe, 2010). WAIC approximates the expected log pointwise predictive density by subtracting a bias for the effective number of parameters from the sum of log pointwise predictive density. WAIC approximates leave-one-out cross validation (LOO-CV) in singular models and hence can serve as a metric for out-of-sample predictive accuracy of HMTM (Gelman et al., 2014). Predictive accuracy is a good standard for detecting the number of breaks because
overfitting is a major concern in analysis using mixture models and HMMs. Also, the cost of computation is very low as WAIC is computed from MCMC outputs. Note that WAIC of HMTM partitions the data into $T$ pieces of conditional density, and hence the one-step ahead prediction density must be used. Using the formula suggested by Gelman et al. (2014), WAIC for HMTM with $M$ number of latent states ($\mathcal{M}_M$) is

$$\text{WAIC}_{\mathcal{M}_M} = -2 \left( \sum_{t=1}^{T} \log \left[ \frac{1}{G} \sum_{g=1}^{G} p(B^{\text{upper}}_t | \mu^{(g)}_u, \psi^{(g)}_u, \mu^{(g)}_v, \psi^{(g)}_v, \beta^{(g)}, \sigma^{2,(g)}, P^{(g)}, \mathcal{M}_M) \right] \right) - \text{bias for the effective number of parameters}$$

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 $\log \hat{p}(B^{\text{upper}} | \mathcal{M}_M)$ for easy comparison with WAIC following the advice of Gelman et al. (2014): The smaller the deviance, the better the accuracy.

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

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

\(^3\)The other results are reported in the supplementary information.
and $t = 30$ in the case of two breaks. We fit four different HMTMs from no break ($M_0$) to three breaks ($M_3$) and compare their model diagnostics, recovered latent spaces, and time-varying network generation rules. From Table 1 to Table 3, the first column shows the number of imposed breaks and their model-fits. The columns in the middle display estimated regime-specific latent spaces for each model. Note that there are $k + 1$ regimes for $k$ break model: $k = 0, \ldots, 3$. The last column shows estimated time-varying network generation rule parameters for $R = 2$.

Table 1 summarizes the results from the no break case. In the no break case, the longitudinal network was generated from a two block structure, consisting of a 10 node block and a 20 node block respectively, and the underlying block structure remains constant. The reading of the results starts from model diagnostics in the first column. While the approximate log marginal likelihood incorrectly favors the one break model ($M_1$), WAIC correctly shows that the no break model ($M_0$) fits the data best. The no break model ($M_0$) in Table 1 correctly recovers the latent block structure. Note that one dimensional information is sufficient to identify the planted block structure and hence the second dimensional network generation rule ($v_2$) oscillates around zero throughout time.

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

Table 2 summarizes the simulation results from the block-splitting change example. The ground truth is the single break HMTM in the third row. Again, WAIC correctly identifies the single break model as the best-fitting
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>
<thead>
<tr>
<th>Region 1</th>
<th>Region 2</th>
<th>Region 3</th>
<th>Region 4</th>
<th>Region 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Latent Space of Break 1 (1-8)
Latent Space of No Break Model
Latent Space of Break 2 (2-39)
Latent Space of Break 3 (40-40)
Latent Space of Break 4 (37-40)

Example calculation:
- Log likelihood = -7204
- WAIC = 14653
- Break = 2

Example calculation:
- Log likelihood = -7191
- WAIC = 14589
- Break = 1

- Parameter estimation and analysis for different regions and models.
model while the approximate log marginal likelihood favors the three break HMTM. The one break model ($M_1$) in Table 2 correctly recovers the block-splitting latent change and the second dimensional network generation rule ($v_2$) jumps from 0 to a positive number in the middle as the number of blocks increases from 2 to 3.

Next, we check whether HMTM correctly recovers more complicated planted network changes. In this example, we generated a synthetic longitudinal network dataset with two different changes. The first break is defined by a block-merging change at $t = 11$ and the second break is defined by a block-splitting change at $t = 31$. Table 3 compares the simulation results from four competing models ($M_0$ to $M_3$). Again, WAIC correctly detects $M_2$ as the best-fitting model while the approximate log likelihood favors $M_3$. The two break HMTM correctly recovers the two underlying changes at $t = 11$ (block merging) and $t = 31$ (block splitting). Changes in the relative size of network generation rules (the last column) inform us the types of changes underlying network structures go through. For example, when the number of blocks changes from 2 to 3 in the transition to Regime 3, $v_2$ returns to its previous level at Regime 1.

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

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

Break = 0

\[ \log \text{likelihood} = -6405 \]

\[ B | M_p \hat{\log} \times 2^{−1} \]

\[ \text{WAIC} = 13016 \]

Break = 1

\[ \log \text{likelihood} = -6393 \]

\[ B | M_p \hat{\log} \times 2^{−1} \]

\[ \text{WAIC} = 13067 \]

Break = 2

\[ \log \text{likelihood} = -6405 \]

\[ B | M_p \hat{\log} \times 2^{−1} \]

\[ \text{WAIC} = 12929 \]

Break = 3

\[ \log \text{likelihood} = -6393 \]

\[ B | M_p \hat{\log} \times 2^{−1} \]

\[ \text{WAIC} = 13648 \]
Table 2: Simulation Analysis of Temporal Network Data with a Block-Splitting Structure. The ground truth is two breaks from a three block structure to a two block structure right after \( t = 10 \) and from a two block structure to a three block structure starting from \( t = 30 \). The ground truth is two breaks from a one block structure to a three block structure right after \( t = 10 \) and from a three block structure to a two block structure right after \( t = 30 \).
Table 4 shows the model diagnostic results for the MTRM and five HMTMs fitted to the major power alliance data set. We report Geweke’s z-score for $\sigma^2_m$ for convergence diagnostics in addition to WAIC and the approximate log marginal likelihood. HMTM with four breaks is favored both by WAIC and the approximate log marginal likelihood. Models with more than four breaks show strong signs of nonconvergence due to the existence of redundant states. HMTM with three breaks also shows a sign of nonconvergence (Geweke’s z-score = 2.31), which is caused by a fewer number of hidden states than necessary.

Figure 2 visualizes changes in latent node positions of major powers (top) and changing patterns of the major-power network topology (bottom) from the four break HMTM. Regime-specific network generation rule parameters $v_{rt}$ are reported in axis labels. Several substantive findings are noteworthy. First, the findings of the first two regimes clearly demonstrate the centrality of Austria-Hungary, connecting groups of major powers. Historians call 33 years after the end of Napoleonic Wars “the age of Metternich,” highlighting the central role played by Chancellor of Austria-Hungary in the European balance of power system after Napoleonic Wars (Rothenberg, 1968). The second panel of Figure 2 clearly shows how Austria-Hungary, a weaker state than other major powers, was able to hold a diplomatic power over other major powers. The position of Austria-Hungary in the major power alliance network is highly critical in the sense that the removal of Austria-Hungary would have made the major power alliance network completely

---

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

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

Fig 2: Changing Node Positions and Network Topology of Military Alliances 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.

Regime 1 (1816−1830)
Regime 2 (1832−1880)
Regime 3 (1882−1918)
Regime 4 (1920−1954)
Regime 5 (1956−2012)
disconnected. In the language of social network analysis, Austria-Hungary filled a “structural hole” in the major power alliance network at the time, playing the role of broker (Burt, 2005, 2009; Stovel and Shaw, 2012). However, the distinct role of Austria-Hungary as broker appeared most clearly during Regime 2 (1832-1880), which only partially overlaps with “the age of Metternich” (1815-1848).

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

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

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

---

7A shrewd European historian succinctly summarizes it as follows:

It was not until Metternich that a statesman appeared who had not only internalized the concept [the balance of power] but was given the opportunity to create a new international structure that explicitly embodied it. His less perceptive successors allowed it to collapse. Bismarck recreated it, although on a far less stable basis. Again his successors allowed it to collapse. The First World War came about not because of the unstable power balance created by competing alliances . . . but because the German Empire was no longer interested in maintaining a power balance (Howard, 1994, 134).
5. Concluding Remarks. In this article, we presented HMTM as a statistical method to detect and analyze changes in structural properties of longitudinal network data. The proposed method has several advantages over existing dynamic network models.

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

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

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

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

Appendix 1. MCMC Algorithm for Hidden Markov Tensor Model.

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

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

**Part 1**

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

1. $p(ψ_{r,u,m}|B,P,S,Θ^{-ψ_{u,m}}) \propto IG\left(\frac{ν_0+N}{2},\frac{U_{r,m}^T ψ_{r,u,m} + u_1}{2}\right)$.

2. $p(μ_{u,m}|B,P,S,Θ^{-μ_{u,m}}) \propto$ multivariate normal$(U_m^T 1/(N + 1), Ψ_{u,m}/(N + 1))$.

3. $p(U_m|B,P,S,Θ^{-U_m}) \propto$ matrix normal$_{N \times R}(\tilde{M}_{u,m}, I_N, \tilde{Ψ}_{u,m})$ where

$$\tilde{Ψ}_{u,m} = (Q_{u,m}/σ_m^2 + Ψ^{-1}_{u,m})^{-1}$$

$$\tilde{M}_{u,m} = (L_{u,m}/σ_m^2 + 1μ_{u,m}^T Ψ^{-1}_{u,m}) \tilde{Ψ}_{u,m}$$

$$Q_{u,m} = (U_m^T U_m) \odot (U_m^T V_m^T)$$

$$L_{u,m} = \sum_{j,t: t \in S_t = m} b_{i,j,t} \odot (U_{m,i} \odot U_{m,j})$$

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

**Step 2** The sampling of $V_r, μ_r, Ψ_r$ is done for each regime. Let $Ψ_v = \begin{pmatrix} ψ_{1,v,m} & \ldots & 0 \\ 0 & ψ_{r,v,m} & 0 \\ 0 & \ldots & ψ_{R,v,m} \end{pmatrix}$.

1. $p(ψ_{r,v,m}|B,P,S,Θ^{-ψ_{v,m}}) \propto IG\left(\frac{ν_0+T}{2},\frac{V_{r,m}^T ψ_{r,v,m} + v_1}{2}\right)$.

2. $p(μ_{v,m}|B,P,S,Θ^{-μ_{v,m}}) \propto$ multivariate normal$(V_m^T 1/(T + 1), Ψ_{v,m}/(T + 1))$.

3. $p(V_m|B,P,S,Θ^{-V_m}) \propto$ matrix normal$_{T \times R}(\tilde{M}_{v,m}, I_T, \tilde{Ψ}_{v,m})$ where

$$\tilde{Ψ}_{v,m} = (Q_{v,m}/σ_m^2 + Ψ^{-1}_{v,m})^{-1}$$

$$\tilde{M}_{v,m} = (L_{v,m}/σ_m^2 + 1μ_{v,m}^T Ψ^{-1}_{v,m}) \tilde{Ψ}_{v,m}$$

$$Q_{v,m} = (U_m^T U_m) \odot (U_m^T U_m)$$

$$L_{v,m} = \sum_{i,j} b_{i,j} \odot (U_{m,i} \odot U_{m,j})$$

**Step 3** The sampling of $β$ from $N(b_1, B_1)$ where

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

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

$1(S = m)$ is the number of time units allocated to state $m$ and $μ_{i,j,t}$ is an element of $U_{S_t} Λ_t U_{S_t}^T$.

**Step 4** The sampling of $σ_m^2$ from $IG\left(\frac{ν_0+N_m N T_m}{2},\frac{d_0+∑_{i=1}^N (1)∑_{j=1}^N (2) b_{i,j,t} - β - μ_{i,j,t}}{2}\right)$.

Part 2
Step 5  Sample $S$ recursively using Chib (1998)’s algorithm. The joint conditional distribution of the latent states $p(S_0, \ldots, S_T | \Theta, B, P)$ can be written as the product of $T$ numbers of independent conditional distributions:

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

Using Bayes’ Theorem, Chib (1998) shows that

$$p(S_t | S_{t+1}, \Theta, B, P) \propto p(S_t | \Theta, B, P) \cdot p(S_{t+1} | S_t, P).$$

State probabilities given all data Transition probability at $t$

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: $p(P | B, S, \Theta)$

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 $k+1$.

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 blocks:

$$\log p(\mu^*_u, \psi^*_u, \mu^*_v, \psi^*_v, \beta^*, \sigma^2 | P) = \log p(\mu^*_u | B) + \sum_{r=1}^{R} \log p(\psi^*_r | B, \mu^*_u) + \log p(\mu^*_v | B, \mu^*_u, \psi^*_u) + \sum_{r=1}^{R} \log p(\psi^*_r | B, \mu^*_u, \psi^*_u, \mu^*_v) + \log p(\beta^* | B, \mu^*_u, \psi^*_u, \mu^*_v, \psi^*_v) + \log p(\sigma^2 | B, \mu^*_u, \psi^*_u, \mu^*_v, \psi^*_v, \beta^*).$$

Let $\Theta$ indicate a parameter vector beside hidden states ($S$) and a transition matrix ($P$): $\Theta = \{ \mu_u, \psi_{u,u}, \mu_v, \psi_{v,v}, \beta, \sigma^2 \}$. 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 (11) can be computed by Chib (1995)’s candidate formula:

Step 1

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

Step 2

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

Step 3

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

Step 4

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

Step 5

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

Step 6

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

Step 7

\[ p(P^* | B, \mu_{\cdot,u}^*, \psi_{\cdot,u}^*, \mu_{\cdot,v}^*, \psi_{\cdot,v}^*, \beta^*, \sigma^2_{\cdot,u}) \approx \int p(P^* | B, \mu_{\cdot,u}^*, \psi_{\cdot,u}^*, \mu_{\cdot,v}^*, \psi_{\cdot,v}^*, \beta^*, \sigma^2_{\cdot,u}, S) dp(S | B) \]
References.


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


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

JONG HEE PARK
SEOUL NATIONAL UNIVERSITY

YUNKYU SOHN
PRINCETON UNIVERSITY