Bayesian Analysis (2020) **15**, Number 1, pp. 133–157

Detecting Structural Changes in Longitudinal Network Data

Jong Hee Park[†,*] and Yunkyu Sohn[‡]

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 network change-point model (HNC) 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.

Keywords: network latent space, hidden Markov model, WAIC, military alliance.

1 Introduction

Multi-way array representation of network data is becoming a standard way to model 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 dataset can be represented as a three-way array $\mathcal{Y} = \{\mathbf{Y}_t | t \in \{1, \ldots, T\}\} \in \mathbb{R}^{N \times N \times T}$ in which its t-th slice is an $N \times N$ square matrix $\mathbf{Y}_t = \{y_{i,j,t} | i, j \in \{1, \ldots, N\}\}$. Here $y_{i,j,t}$ informs the dyadic relationship between actors i and j at time t. While dynamic modeling of longitudinal networks has been an increasingly important topic in social, biological, and other fields of science, a fully probabilistic treatment of dynamic network process has been challenging due to simultaneous dependence between dyadic and temporal observations that are often associated with fundamental shifts in data generating processes.[1]

*Jong Hee Park was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2013S1A3A2053683). We thank James Fowler, Kosuke Imai, In Song Kim, John Londregan, Jacob Montgomery, Marc Ratkovic, Michael Ward, and Soichiro Yamauchi for helpful comments on earlier drafts of this paper. An open-source `R` package that implements the proposed method, `NetworkChange`, is available at CRAN.

[†]Department of Political Science and International Relations, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, Republic of Korea 08826, jongheepark@snu.ac.kr

[‡]School of Political Science and Economics, Waseda University, 1-6-1 Nishiwaseda Shinjuku-ku, Tokyo, Japan 169-8050, ysohn@aoni.waseda.jp

[1]For a range of examples for longitudinal network analysis, see Holme and Saramäki (2012) and references therein. For a general survey of network models, see Goldenberg et al. (2010).

 https://doi.org/10.1214/19-BA1147

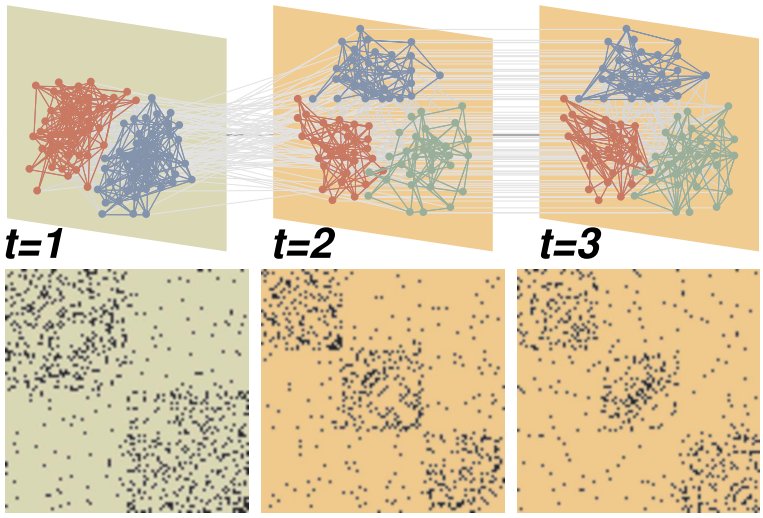


Figure 1: A dynamic network with a change-point. The network has two clusters at $t = 1$ whereas networks at $t = 2$ and $t = 3$ have three clusters. The top panel shows node traits on the latent space of networks and the bottom panel shows it in a matrix format. Gray lines between layers in the top panel 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. Links between clusters are colored in white. Olive and yellow colors shown in the backgrounds of top and bottom panels indicate the hidden regimes of latent space shared among layers.

In this paper, we present a Bayesian method to model change-point process in longitudinal networks by employing a Bayesian framework of multilinear array decomposition by Hoff (2011) and Hoff (2015). 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; Ridder et al., 2016).

Figure 1 shows an example of longitudinal network data with a change-point. The top panel shows node traits on the latent space of networks and the bottom panel shows it in a matrix format. The distance between each pair of nodes on the space 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 in the example).

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 algorithms 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 to a mixture of χ^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). 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. 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.

[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.

For example, let $\{z_{i,j,t}\}$ be a continuous latent variable that captures the link probability between i and j observed at time t and $\mathbf{x}_{i,j,t}$ be a vector of known covariates affecting $\{z_{i,j,t}\}$. Then, based on the notion of multilayer exchangeability (Hoff, 2009b), MTRM models $\{z_{i,j,t}\}$ as a function of covariates, node-specific random effects ($\{\mathbf{u}_1, \ldots, \mathbf{u}_N\}$) and a diagonal matrix of time-specific random effects ($\mathbf{V}_t$):

$$\begin{aligned} \Pr(y_{i,j,t} = 1 | \mathbf{x}_{i,j,t}, \mathbf{u}_i, \mathbf{u}_j, \mathbf{v}_t) &= \Phi(z_{i,j,t}) && (1) \\ z_{i,j,t} &= \mathbf{x}_{i,j,t}\boldsymbol{\beta} + \mathbf{u}_i^T \mathbf{V}_t \mathbf{u}_j + \epsilon_{i,j,t} && (2) \\ \epsilon_{i,j,t} &\sim \mathcal{N}(0,1) && (3) \end{aligned}$$

where $\{\mathbf{u}_1, \ldots, \mathbf{u}_N\}$ represent (time-constant) R-dimensional latent node positions and $\mathbf{V}_t = \begin{bmatrix} v_{1,t} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & v_{R,t} \end{bmatrix}$ is a diagonal matrix of (time-varying) network generation rules. In a 2-dimensional latent space, for example, $\mathbf{u}_i = (u_{i,1}, u_{i,2})$ indicates node i's latent position, and $\mathbf{u}_i^T \mathbf{V}_1 \mathbf{u}_j$ represents the latent connection probability between node i and node j at time 1.[4] 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. 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 ($\{\mathbf{u}_1, \ldots, \mathbf{u}_N\}$), which are constrained to be constant over time in Hoff (2011), are allowed to change depending on the transition of hidden states.

We call the resulting model a hidden Markov network changepoint model (HNC). HNC 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 networks 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

[4] In a 2-dimensional latent space, $\mathbf{u}_i^T \mathbf{v}_t \mathbf{u}_j$ is

$$\begin{bmatrix} u_{i,1} v_{1,t} u_{j,1} \\ u_{i,2} v_{2,t} u_{j,2} \end{bmatrix} = \begin{bmatrix} u_{i,1} & u_{i,2} \end{bmatrix} \times \begin{bmatrix} v_{1,t} & 0 \\ 0 & v_{2,t} \end{bmatrix} \times \begin{bmatrix} u_{i,1} \\ u_{i,2} \end{bmatrix}. \quad (4)$$

in the latent space approach. We 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 Array Data

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

$$\begin{aligned}
\Pr(y_{i,j,t} = 1 | \mathbf{x}_{i,j,t}, \mathbf{u}_i, \mathbf{u}_j, \mathbf{v}_t) &= \boldsymbol{\Phi}(\mathbf{x}_{i,j,t}\boldsymbol{\beta} + \langle \mathbf{u}_i, \mathbf{v}_t, \mathbf{u}_j \rangle) && (5)\\
\mathbf{U} &\sim \text{matrix normal}(\mathbf{M} = \mathbf{1}\boldsymbol{\mu}_U^T, \mathbf{I}_N, \Psi_U) && (6)\\
\mathbf{V} &\sim \text{matrix normal}(\mathbf{M} = \mathbf{1}\boldsymbol{\mu}_V^T, \mathbf{I}_T, \Psi_V) && (7)\\
\epsilon_{i,j,t} &\sim \mathcal{N}(0, \sigma^2) && (8)
\end{aligned}$$

where $\langle \mathbf{u}_i, \mathbf{v}_t, \mathbf{u}_j \rangle = \sum_{r=1}^{R} u_{i,r} v_{r,t} u_{j,r}$ and matrix normal $(\mathbf{M}, \mathbf{U}, \mathbf{V})$ is an $N \times R$ matrix-variate normal distribution with mean matrix $\mathbf{M}$, row variance $\mathbf{U}$, and column variance $\mathbf{V}$.

The resulting estimates of node-specific latent variables recover a specific type of similarity between nodes that is easily interpretable (Hoff, 2008). If $\mathbf{u}_i$ and $\mathbf{u}_j$ exhibit similar values, they will have similar inner product outcomes with node k's latent position vector $\mathbf{u}_k$. 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 network generation rule parameter $\mathbf{v}_t$ contains the information on what the distance relationships on each dimension of the $\mathbf{U}$ space reveal about their connection probability. For example, $v_{1,t} > 0$ corresponds to the case when a network generation rule for the first dimension at time t is homophilous (or assortative). In words, this indicates that two nodes on the first 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, $v_{1,t} < 0$ corresponds to the case when a network generation rule for the first dimension at time t is heterophilous (or 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

The formulation of MTRM in Equation (5) is designed to estimate consistent regression parameters ($\boldsymbol{\beta}$), considering network effects as a nuisance parameter. Hence it entails a critical weakness in uncovering latent node traits. One of the most prominent latent node traits for changepoint analysis is *meso-scopic network properties*. Meso-scopic network properties indicate characteristics of a network at the group level. Examples of meso-scopic network properties are homophily (the tendency to link with similar others), heterophily (the tendency to link with different others), a community structure, a stochastic block, and a core-periphery structure. The emergence (and changes) of group structures is commonly observed in real-world network data.

Except for exogenous covariates ($\mathbf{x}_{i,j,t}$), Equation (5) has no treatment to account for degree heterogeneity that has been known to confound the recovery of meso-scopic network properties (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 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 similar. 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).

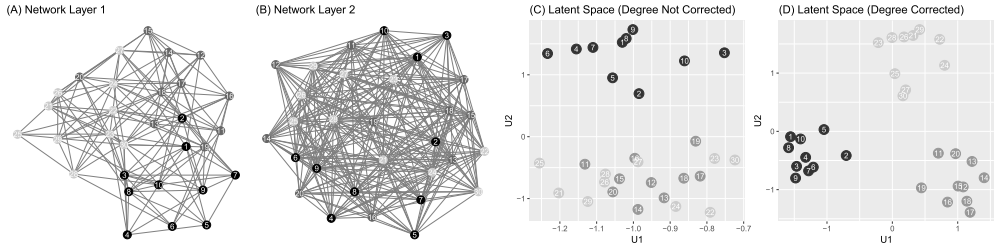


Figure 2: Degree Correction for Group Structure Discovery. Two undirected networks with 30 nodes and 3 groups are generated. While they share exactly same group structure, Layer 1 is a homophilous network and Layer 2 is a heterophilous network. Specifically, we generated the Layer 1 network by setting the within-group link probability (the probability of being connected with in-group nodes) at 0.7 and the between-group link probability (the probability of being connected with out-group nodes) at 0.2. In Layer 2, the within-group link probability is set at 0.2 and the between-group link probability is set at 0.7.

Figure 2 illustrates the problem in a simple setting using a 2 layer network. We generate two undirected networks consisting three groups 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 is much larger than the between group link probability 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 a probit MTRM shown in Equation (5) 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 light gray group from the dark gray 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 ($\mathbf{\Omega}_t$), consisting of the principal eigenvalue ($\lambda_t^{princ} = \max(|\lambda(\mathbf{Y}_t)|)$) and its associated eigenvector (Peixoto, 2013):

$$\Omega_{i,j,t} = \lambda_t^{princ} \tilde{\mathbf{u}}_{i,t} \tilde{\mathbf{u}}_{j,t}^T \tag{9}$$

where $\tilde{\mathbf{u}}_{it}$ is the ith row of the associated eigenvector.[5] In matrix form, we denote the principal eigenmatrix as $\mathbf{\Omega}_t$ for the t-th slice.

Figure 2(D) shows the results from the linear MTRM using the transformed data $\mathcal{B}$ whose t-th slice $\mathbf{B}_t = \mathbf{Y}_t - \mathbf{\Omega}_t$. In accordance with clustering of the positions, k-means clustering result in Figure 2(D) perfectly agrees with true group labels. The use of a null model allows us to recover three distinct groups in the data.[6]

3 The Proposed Method

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 global network properties (*e.g.* average shortest path length or network density) as a structural change. On the other hand, a change in the population statistics of local network properties (*e.g.* transitivity or node degree) can be considered as a structural change. However, global network statistics and local properties cannot fully represent generative processes of dynamic networks as the granularity of the information entailed in such measures is too limited. For instance, it is common to observe networks having exactly same average shortest path length or local network structures whereas their data generational processes are completely different. In contrast, it is hard to imagine a fundamental change in a network's generation process without a change in the structure of subgroups and their connection patterns. For this reason, we define a structural change in networks as a change in meso-scopic features of networks.

[5] Alternatively, one can use a modularity matrix ($M_{i,j,t}$):

$$M_{i,j,t} = y_{i,j,t} - \frac{k_i k_j}{2m}$$

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

[6] One may 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, using a null model is more intuitive and computationally less expensive than including a list of covariates. For more information about uncovering hidden groups from an array of network data, see Sohn and Park (2017).

In the following, we introduce a hidden Markov network change-point model to detect changes in meso-scopic network features. Next, we discuss sampling algorithms of the proposed model. Then, we explain diagnostic methods for the proposed model.

3.1 Hidden Markov Network Change-point Model

As shown by Chib (1998), multiple change-point problems are equivalent to the estimation of a nonergodic (or forward-moving) HMM, which has advantages in avoiding label-switchings of hidden states thanks to the order constraint in hidden states. Let us denote $\mathbf{S}$ as a vector of hidden state variables where S_t is an integer-valued hidden state variable at t

$$\mathbf{S} = \{(S_1, \ldots, S_T) : S_t \in \{1, \ldots, M\}, t = 1, \ldots, T\}, \tag{10}$$

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

Suppose $\boldsymbol{\Theta}$ be a collection of parameters that represents a network generating process of a longitudinal network and $\mathcal{B}$ be a degree-corrected longitudinal network data. Then, we can write the likelihood using the conditional independence assumption of the HMM:

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

From the above notation, the connection between MTRM and HNC becomes clear: HNC is a collection of multiple MTRMs following the Markov dependence of hidden states.

In a regression form, we substitute the regime-specific subscript m by S_t, which is an integer-valued hidden state variable at t. Then,

$$\mathbf{B}_t \quad = \quad \mathbf{U}_{S_t} \mathbf{V}_t \mathbf{U}_{S_t}^T + \mathbf{E}_t \tag{11}$$

$$\mathbf{E}_t \quad \sim \quad \begin{cases} \mathcal{N}_{N\times N}(\mathbf{0}, \sigma_{S_t}^2 \mathbf{I}_N, \mathbf{I}_N) & \text{for} \quad \text{Normal Error} \\ \mathcal{N}_{N\times N}(\mathbf{0}, \gamma_t^{-1} \sigma_{S_t}^2 \mathbf{I}_N, \mathbf{I}_N) & \text{for} \quad \text{Student-}t \text{ Error.} \end{cases} \tag{12}$$

One may concern that the normal distribution of $\mathbf{E}_t$ 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 γ_t follows a gamma distribution ($\mathcal{G}(\nu_0/2, \nu_1/2)$).

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

The resulting prior distributions of $\mathbf{U}$ and $\mathbf{V}$ are matrix-variate normal distributions in which each column vector ($\mathbf{u}_{i,S_t}$ and $\mathbf{v}_t$) follows a multivariate normal distribution. We first discuss the prior distribution of $\mathbf{U}$:

$$\begin{aligned}
\mathbf{U}_{S_t} &\equiv (\mathbf{u}_{1,S_t}, \ldots, \mathbf{u}_{N,S_t})^\top \in \mathbb{R}^{N\times R} && (13)\\
\mathbf{u}_{i,S_t} &\sim \mathcal{N}_R(\boldsymbol{\mu}_{u,S_t}, \Psi_{u,S_t}) && (14)\\
\boldsymbol{\mu}_{u,S_t}|\Psi_{u,S_t} &\sim \mathcal{N}_R(\boldsymbol{\mu}_{0,u_{S_t}}, \Psi_{u,S_t}) && (15)\\
\Psi_{u,S_t} &\equiv \begin{pmatrix} \psi_{1,u,S_t} & \cdots & 0 \\ 0 & \psi_{r,u,S_t} & 0 \\ 0 & \cdots & \psi_{R,u,S_t} \end{pmatrix} && (16)\\
\psi_{r,u,S_t} &\sim \mathcal{IG}\left(\frac{u_0}{2}, \frac{u_1}{2}\right). && (17)
\end{aligned}$$

The prior distributions of $\mathbf{V}$ are similar to $\mathbf{U}$ but one difference is that only diagonal elements of $\mathbf{V}_t$ are modeled as a multivariate normal distribution:

$$\begin{aligned}
\mathbf{V}_t &\equiv \begin{pmatrix} v_{1,t} & \cdots & 0 \\ 0 & v_{r,t} & 0 \\ 0 & \cdots & v_{R,t} \end{pmatrix} && (18)\\
\mathbf{v}_t &\equiv (v_{1,t}, \ldots, v_{R,t})^\top \in \mathbb{R}^{R\times 1} && (19)\\
\mathbf{v}_t &\sim \mathcal{N}_R(\boldsymbol{\mu}_v, \Psi_v) && (20)\\
\boldsymbol{\mu}_v|\Psi_v &\sim \mathcal{N}_R(\boldsymbol{\mu}_{0,v}, \Psi_v) && (21)\\
\Psi_v &= \begin{pmatrix} \psi_{1,v} & \cdots & 0 \\ 0 & \psi_{r,v} & 0 \\ 0 & \cdots & \psi_{R,v} \end{pmatrix} && (22)\\
\psi_{r,v} &\sim \mathcal{IG}\left(\frac{v_0}{2}, \frac{v_1}{2}\right). && (23)
\end{aligned}$$

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

$$S_t|S_{t-1}, \mathbf{P} \sim \mathcal{M}arkov(\mathbf{P}, \pi_0) \quad (24)$$

$$\underbrace{\mathbf{P}}_{M\times M} = (\mathbf{p}_1,\ldots,\mathbf{p}_M) \tag{25}$$

$$\mathbf{p}_i \sim \text{Dirichlet}(\alpha_{i,1},\ldots,\alpha_{i,M}) \tag{26}$$

where π_0 is the initial probability of a non-ergodic Markov chain ($\pi_0 = (1,0,\ldots,0)$). The duration of hidden state m follows a geometric distribution of $1-p_{m,m}$ where $p_{m,m}$ is the mth diagonal element of transition matrix $\mathbf{P}$. The regime change probability can be computed using the posterior draws of hidden states. For example, if one wishes to compute the probability of a regime change from regime 1 to regime 2 between t and $t+1$ is $\frac{1}{G}\sum_{g=1}^{G}\mathcal{I}(S_{t+1}^{(g)}=2|S_t^{(g)}=1)$ where $\mathcal{I}(\cdot)$ indicates a count function and G is the number of MCMC simulation.

3.2 Sampling Algorithm

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

$$p(\boldsymbol{\Theta},\mathbf{P},\mathbf{S}|\mathcal{B}) \propto \mathcal{N}_{N\times N}(\mathbf{B}_1|\boldsymbol{\Theta}_1)\prod_{t=2}^{T}\Big(\mathcal{N}_{N\times N}(\mathbf{B}_t|\mathcal{B}_{t-1},\boldsymbol{\Theta}_{S_t})p(S_t|S_{t-1},\mathbf{P})\Big) \tag{27}$$

$$\prod_{m=1}^{M}\Big(\mathcal{N}_R(\boldsymbol{\mu}_{u,m}|\boldsymbol{\mu}_{0,u,m},\psi_{.,u,m})\mathcal{N}_R(\boldsymbol{\mu}_v|\boldsymbol{\mu}_{0,v},\psi_{.,v,m})\Big) \tag{28}$$

$$\prod_{m=1}^{M}\prod_{r=1}^{R}\Big(\mathcal{IG}(\psi_{r,u,m}|u_{0,m},u_{1,m})\mathcal{IG}(\psi_{r,v,m}|v_{0,m},v_{1,m})\Big) \tag{29}$$

$$\prod_{m=1}^{M}\Big(\mathcal{IG}(\sigma_m^2|c_0,d_0)\mathcal{B}eta(p_{m,m}|a,b)\Big) \tag{30}$$

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

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

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

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

1. $p(\psi_{r,u,m}|\mathcal{B},\mathbf{P},\mathbf{S},\boldsymbol{\Theta}^{-\Psi_{u,m}}) \propto \mathcal{IG}\left(\frac{u_0+N}{2},\frac{\mathbf{U}_{r,m}^T\mathbf{U}_{r,m}+u_1}{2}\right)$.

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

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

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

Step 2 The sampling of $\mathbf{V}, \boldsymbol{\mu}_v, \Psi_v$ is done for each regime. Let

$$\Psi_v = \begin{pmatrix} \psi_{1,v,m} & \cdots & 0 \\ 0 & \psi_{r,v,m} & 0 \\ 0 & \cdots & \psi_{R,v,m} \end{pmatrix}.$$

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

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

Step 3 The sampling of σ_m^2 from $\mathcal{IG}\left(\frac{c_0+N_m\cdot N_m\cdot T_m}{2}, \frac{d_0+\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T b_{i,j,t}-\mu_{i,j,t}}{2}\right)$.

Step 4 Sample $\mathbf{S}$ recursively using Chib (1998)'s algorithm. The joint conditional distribution of the latent states $p(S_0,\ldots,S_T|\boldsymbol{\Theta},\mathcal{B},\mathbf{P})$ can be written as the product of T numbers of independent conditional distributions:

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

Using Bayes' Theorem, Chib (1998) shows that

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

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

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

Step 5 Sample each row of $\mathbf{P}$ from the following Beta distribution:

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

where $P_{k,k}$ 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 HNC 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 HNC with M numbers of latent states ($\mathcal{M}_M$) can be computed as follows:

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

where $\{\boldsymbol{\mu}_u^*, \psi_{\cdot,u}^*, \boldsymbol{\mu}_v^*, \psi_{\cdot,v}^*, \sigma^{2*}, \mathbf{P}^*\}$ are posterior means of MCMC outputs. The log likelihood is computed by summing log predictive density values evaluated at posterior

means across all states and over all upper triangular array elements as follows:

$$\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=i+1}^{N-1}\sum_{m=1}^{M} p(b_{i,j,t}|\mathcal{B}_{t-1}^{\text{upper}}, \boldsymbol{\mu}_{u,m}^{*}, \psi_{.,u,m}^{*}, \boldsymbol{\mu}_{v,m}^{*}, \psi_{.,v,m}^{*}, \sigma_{m}^{2*}, \mathbf{P}^{*}, S_t = m, \mathcal{M}_M)$$
$$p(S_t = m|\mathcal{B}_{t-1}^{\text{upper}}, \boldsymbol{\mu}_{u,m}^{*}, \psi_{.,u,m}^{*}, \boldsymbol{\mu}_{v,m}^{*}, \psi_{.,v,m}^{*}, \sigma_{m}^{2*}, \mathbf{P}_{m}^{*}, \mathcal{M}_M).$$

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 HNC, the log posterior density evaluated at posterior means is decomposed into seven blocs:

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

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 HNC (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 HNC 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 HNCs given a varying number of breaks.

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

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

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

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

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 block-structured network data. Due to space limitation, we only report the results of two simulation tests (a block-splitting break and a block-merging break followed by a block-splitting break) here. Additional simulation results (no break, a block-splitting break, and a block-merging break followed by a block-splitting break) are available in the web appendix.

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 block 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 block 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 HNCs 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.

[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.

4.2 Block-Structured Networks with a Block-Splitting Break

The first test is dynamic network data with a single group-splitting break. The ground truth is that the number of latent groups changes from 2 to 3 in the middle. Table 1 shows the results of the simulation. Again, we read the model diagnostic results from 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 HNC. 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.[8] Interestingly, the network generation rule in the last column 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 ± 1.4 around the estimated break points on average while simulated break points of the one break model stay constant.

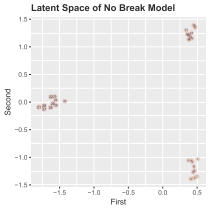
4.3 Block-Structured Networks with Two Breaks

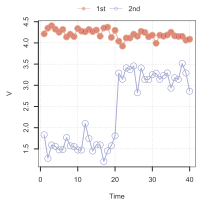
Next, we check whether our method correctly recovers more complicated network changes. The first break is planted at $t = 10$ and it corresponds to a block-merging change. Another break is planted at $t = 30$, which corresponds to a block-splitting change. Thus, the number of latent groups changes from 3, 2 to 3. Table 2 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,

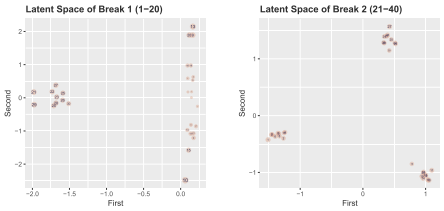
[8] Chib (1995)'s algorithm is based on the summation of log likelihoods evaluated at posterior means and hence sensitive to the presence of singleton states in high dimensional time series data. In contrast, WAIC relies on the log pointwise predictive density as a measure of the goodness of the fit and its variance as a penalty. Since the log pointwise predictive density is averaged over the entire MCMC scan ($\frac{1}{G}\sum_{g=1}^{G} p(\mathcal{B}_t^{\text{upper}}|\boldsymbol{\Theta}^{(g)}, \mathbf{P}^{(g)}, \mathcal{M}_M)$), it is less sensitive to singular components in high dimensional mixture models like HNC. This is why WAIC outperforms in the break number detection in the context of HNC.

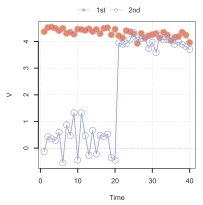
Break number	=	0
WAIC	=	13744
-2*log marginal	=	13648
-2*log likelihood	=	13592



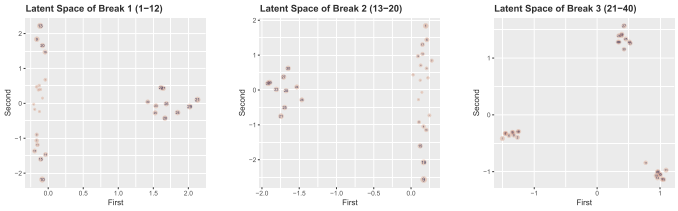


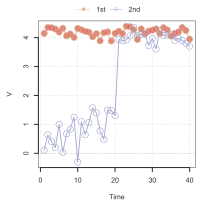
Break number	=	1
WAIC	=	13016
-2*log marginal	=	12922
-2*log likelihood	=	12810
Average Loss	=	0.00



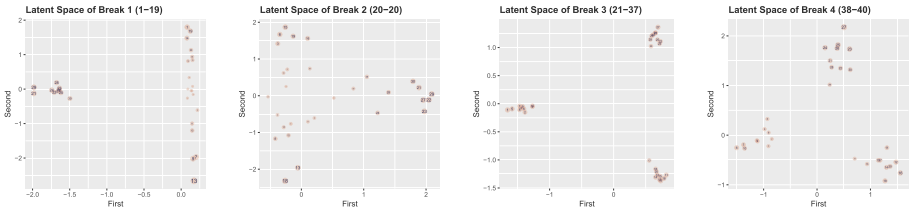


Break number	=	2
WAIC	=	13067
-2*log marginal	=	12929
-2*log likelihood	=	12786
Average Loss	=	0.18





Break number	=	3
WAIC	=	13064
-2*log marginal	=	12877
-2*log likelihood	=	12738
Average Loss	=	0.49



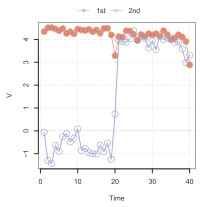
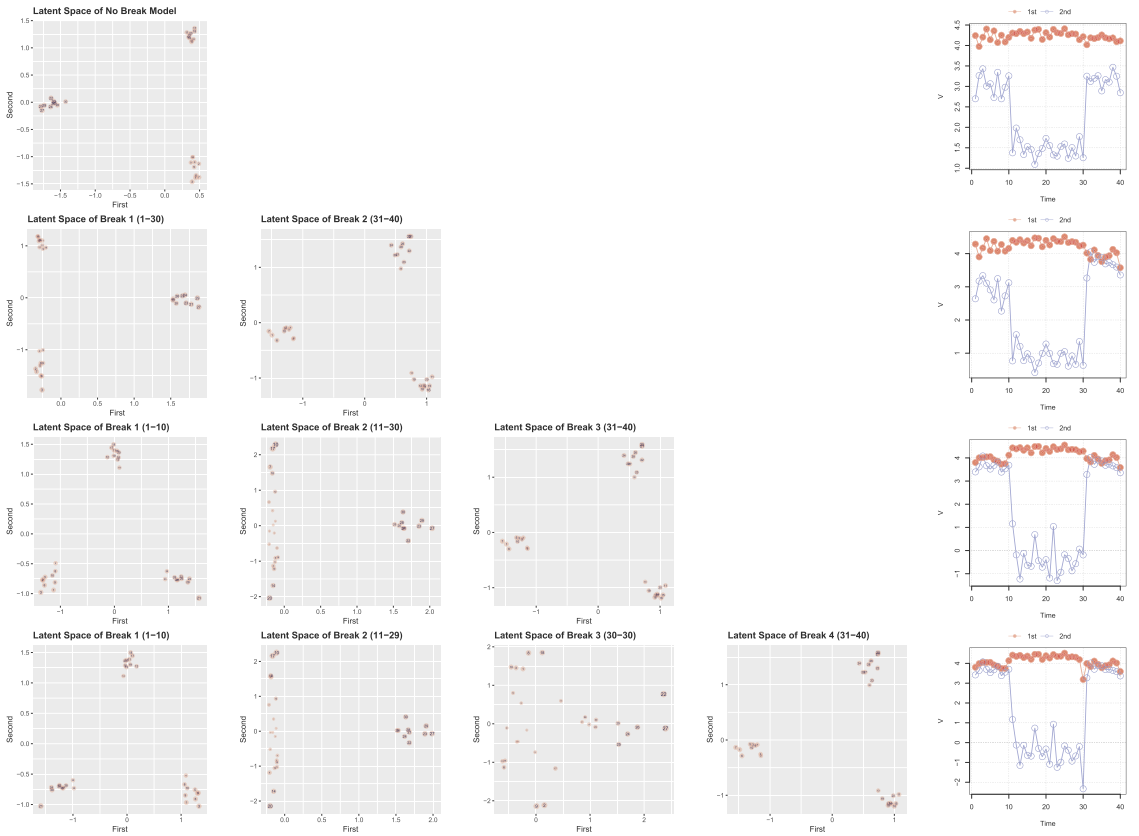


Table 1: Simulation Results of Block-structured Networks with a Block-splitting Break. The ground truth is one break and the underlying group structure changes from a two group structure to a three group structure in the middle.



Break number	=	0
WAIC	=	13727
-2*log marginal	=	13630
-2*log likelihood	=	13574

Break number	=	1
WAIC	=	13583
-2*log marginal	=	13431
-2*log likelihood	=	13324
Average Loss	=	0.27

Break number	=	2
WAIC	=	13202
-2*log marginal	=	13043
-2*log likelihood	=	12892
Average Loss	=	0.00

Break number	=	3
WAIC	=	13203
-2*log marginal	=	13028
-2*log likelihood	=	12856
Average Loss	=	1.59

Table 2: Simulation Results of Block-structured Networks with Two Breaks. The ground truth is two breaks and the underlying group structure changes from a two group structure to a three group structure right after $t = 10$ and from a three group structure to a two group structure right after $t = 30$. The ground truth is two breaks.

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 at the recovered latent states, the two break HNC correctly recovers the two underlying changes between $t = 10$ and $t = 11$ (block merging) and between $t = 30$ 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 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 the 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 when there is no singleton state while WAIC performs consistently well 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 the proposed method to analyze structural changes in international 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 using a principled statistical approach. A main reason is the lack of statistical methods that model unknown structural changes in longitudinal network data.

To highlight how HNC works in a simple setting, we use a small dataset consisting of seven "major powers" (Austria-Hungary, France, Germany, Italy, Japan, Russia, and the United Kingdom) from 1816 to 1938.[9] We aggregated every 2 year network from the original annual binary networks 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 the early 20th century. Also, the period from 1816 to 1938 corresponds to the era of shifting alliances among major powers. Thus, changes in the structure of alliance indicate changes in the international system.

Figure 3 shows the model diagnostic results for the major power alliance data. 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 regimes. 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.

[9]The source of the data is Gibler (2008). Additional results using a larger dataset are available in the web appendix.

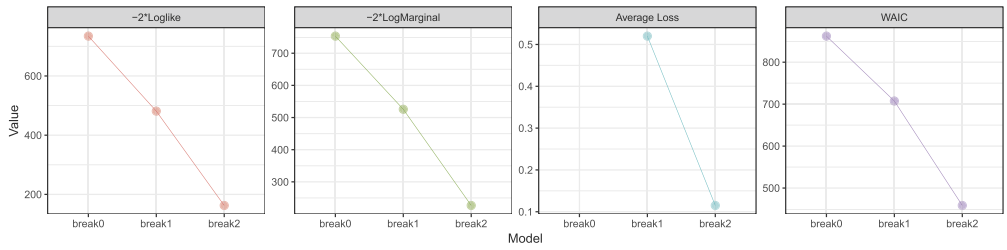


Figure 3: Model Diagnostics of Major Power Alliance Network Changes, 1816 - 1938.

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.[10] 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 observation is the centrality of Austria-Hungary which connects the 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 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, 2009).

The second observation 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 observation is the timing of the second break between 1890 and 1892. Scholars of international relations and historians have focused on the formation of the Dual Alliance between Germany and Austria-Hungary in 1879 and a sequence of alliances that followed, led by a structural change in the balance of the power system (Snyder, 1997; Vermeiren, 2016).[11] These series of events transformed a web of shift-

[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.

[11] 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.

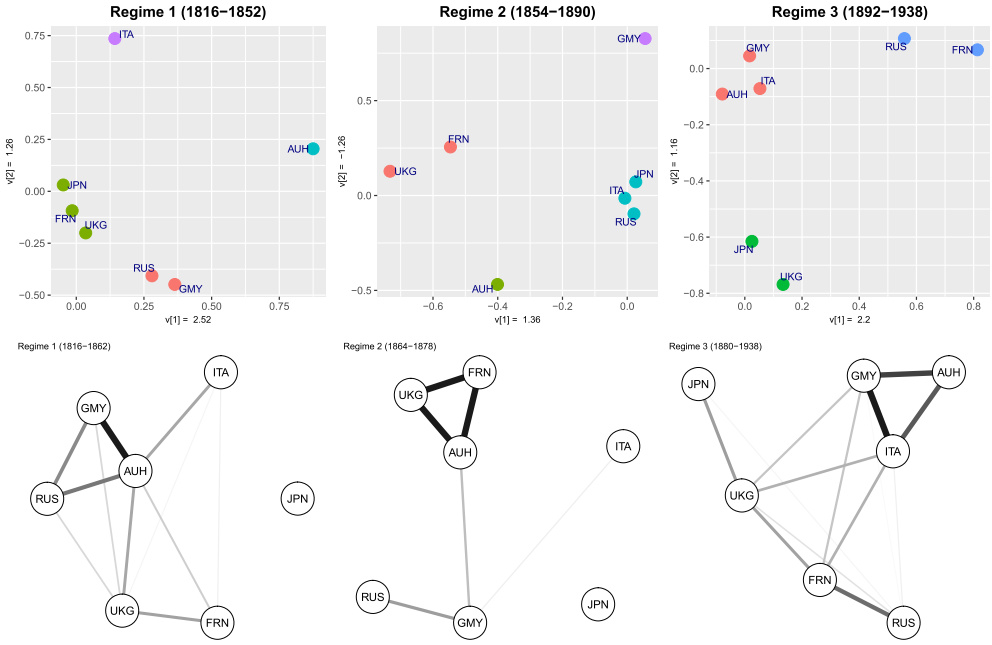


Figure 4: Changing Node Positions and Network Topology of Military Alliance Networks Among Major Powers, 1816 - 1938: Node colors (online) indicate clusters. Regime averages of v_t values for each dimension are reported on each axis (top panel). Line widths are proportional to the duration of alliance links (bottom panel). Included states are Austria-Hungary (AUT), France (FRA), Germany (GMY), Italy (ITA), Japan (JAP), Russia (RUS), and the United Kingdom (GBR). Each node is colored by its cluster label obtained through k-means clustering.

ing alliances into a clearly diverged group structure consisting of two clusters: Austria-Hungary, Germany, and Italy on 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.

6 Concluding Remarks

In this article, we presented HNC 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 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. The 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 HNC 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, HNC provides an important tool to investigate changes in meso-scale properties of longitudinal network data. Changes in meso-scale network properties are important quantities that correspond to fundamental changes in the network generating process.

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).

References

Bartolucci, F., Marino, M. F., and Pandolfi, S. (2018). "Dealing with Reciprocity in Dynamic Stochastic Block Models." *Computational Statistics & Data Analysis*, 123: 86–100. MR3777087. doi: https://doi.org/10.1016/j.csda.2018.01.010. 135

Benson, A. R., Gleich, D. F., and Leskovec, J. (2016). "Higher-order organization of complex networks." *Science*, 353(6295): 163–166. 153

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer. MR2247587. doi: https://doi.org/10.1007/978-0-387-45528-0. 147

Björck, A. (1996). *Numerical Methods for Least Squares Problems*. SIAM. MR1386889. doi: https://doi.org/10.1137/1.9781611971484. 140

Burt, R. S. (2009). *Structural Holes: The Social Structure of Competition*. Harvard University Press. 151

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. MR1166846. doi: https://doi.org/10.2307/3315430. 140

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, 35–1. 137, 138

Chib, S. (1995). "Marginal Likelihood From the Gibbs Output." *Journal of the American Statistical Association*, 90(432): 1313–1321. MR1379473. 137, 144, 145, 147

Chib, S. (1998). "Estimation and Comparison of Multiple Change-point Models." *Journal of Econometrics*, 86(2): 221–241. MR1649222. doi: https://doi.org/10.1016/S0304-4076(97)00115-2. 140, 141, 143, 144

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. 135

Cribben, I. and Yu, Y. (2016). "Estimating Whole-Brain Dynamics by Using Spectral Clustering." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. MR3632344. doi: https://doi.org/10.1111/rssc.12169. 134, 135

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. MR1780272. doi: https://doi.org/10.1137/S0895479896305696. 153

Desmarais, B. A. and Cranmer, S. J. (2012). "Statistical Mechanics of Networks: Estimation and Uncertainty." *Physica A*, 391(4): 1865–1876. 135

Drton, M. (2009). "Likelihood Ratio Tests and Singularities." *Ann. Statist.*, 37(2): 979–1012. MR2502658. doi: https://doi.org/10.1214/07-AOS571. 135

Gelman, A., Hwang, J., and Vehtari, A. (2014). "Understanding Predictive Information Criteria for Bayesian Models." *Statistics and Computing*, 24(6): 997–1016. MR3253850. doi: https://doi.org/10.1007/s11222-013-9416-2. 145, 146

Gibler, D. (2008). *International Military Alliances, 1648-2008*. CQ Press. 150

Goldenberg, A., Zheng, A. X., Fienberg, S. E., and Airoldi, E. M. (2010). "A Survey of Statistical Network Models." *Found. Trends Mach. Learn.*, 2(2): 129–233. 133

Guhaniyogi, R. and Dunson, D. B. (2015). "Bayesian Compressed Regression." *Journal of the American Statistical Association*, 110(512): 1500–1514. MR3449050. doi: https://doi.org/10.1080/01621459.2014.969425. 140

Guo, F., Hanneke, S., Fu, W., and Xing, E. P. (2007). "Recovering Temporally Rewiring Networks: A Model-based Approach." *Proceedings of the 24 th International Conference on Machine Learning*, 321–328. MR2930645. 134, 135

Han, S. and Dunson, D. B. (2018). "Multiresolution Tensor Decomposition for Multiple Spatial Passing Networks." *CoRR*, abs/1803.01203. 133

Hanneke, S., Fu, W., and Xing, E. P. (2010). "Discrete Temporal Models of Social Networks." *Electronic Journal of Statistics*, 4: 585–605. MR2660534. doi: https://doi.org/10.1214/09-EJS548. 135

Hartigan, J. A. (1985). "A Failure of Likelihood Asymptotics for Normal Mixtures." In LeCam, L. and Olshen, R. A. (eds.), *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, volume 2, 807–810. Belmont, California: Wadsworth Statistics/Probability Series. MR0822066. 147

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. MR2758643. doi: https://doi.org/10.1214/10-AOAS329. 134

Hoff, P. (2007). "Model Averaging and Dimension Selection for the Singular Value

Decomposition." *Journal of the American Statistical Association*, 102(478): 674–685. MR2325118. doi: https://doi.org/10.1198/016214506000001310. 153

Hoff, P. D. (2008). "Modeling Homophily and Stochastic Equivalence in Symmetric Relational Data." In Platt, J., Koller, D., Singer, Y., and Roweis, S. (eds.), *Advances in Neural Information Processing Systems 20*, 657–664. Cambridge University Press. 137

Hoff, P. D. (2009a). "Multiplicative Latent Factor Models for Description and Prediction of Social Networks." *Computational & Mathematical Organization Theory*, 15(4): 261–272. 133

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. MR2749840. doi: https://doi.org/10.1198/jcgs.2009.07177. 136

Hoff, P. D. (2011). "Hierarchical Multilinear Models for Multiway Data." *Computational Statistics & Data Analysis*, 55: 530–543. MR2736574. doi: https://doi.org/10.1016/j.csda.2010.05.020. 133, 134, 135, 136, 140, 153

Hoff, P. D. (2015). "Multilinear Tensor Regression for Longitudinal Relational Data." *The Annals of Applied Statistics*, 9(3): 1169–1193. MR3418719. doi: https://doi.org/10.1214/15-AOAS839. 133, 134, 135

Holme, P. and Saramäki, J. (2012). "Temporal Networks." *Physics Reports*, 519(3): 97–125. 133

Johndrow, J. E., Bhattacharya, A., and Dunson, D. B. (2017). "Tensor Decompositions and Sparse Log-linear Models." *Annals of Statistics*, 45(1): 1–38. MR3611485. doi: https://doi.org/10.1214/15-AOS1414. 133

Karrer, B. and Newman, M. E. (2011). "Stochastic Blockmodels and Community Structure in Networks." *Physical Review E*, 83(1): 016107. MR2788206. doi: https://doi.org/10.1103/PhysRevE.83.016107. 137, 138

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. MR1279653. doi: https://doi.org/10.1093/biomet/81.1.27. 142

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. 133

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT press. 140

Newman, M. E. (2006). "Modularity and Community Structure in Networks." *Proceedings of the National Academy of Sciences*, 103(23): 8577–8582. 138

Newman, M. E. (2010). *Networks: An Introduction*. Oxford University Press. MR2676073. doi: https://doi.org/10.1093/acprof:oso/9780199206650.001.0001. 138

Newman, M. E. and Girvan, M. (2004). "Finding and Evaluating Community Structure in Networks." *Physical Review E*, 69(2): 026113. MR2282139. doi: https://doi.org/10.1103/PhysRevE.74.036104. 139

Peixoto, T. P. (2013). "Eigenvalue Spectra of Modular Networks." *Physical Review Letters*, 111(9): 098701–5. 139

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*, 2942–2948. 133

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. 134, 135

Robins, G. L. and Pattison, P. E. (2001). "Random Graph Models for Temporal Processes in Social Networks." *Journal of Mathematical Sociology*, 25(5–41). 135

Rohe, K., Chatterjee, S., and Yu, B. (2011). "Spectral Clustering and the High-Dimensional Stochastic Blockmodel." *The Annals of Statistics*, 39(4): 1878–1915. MR2893856. doi: https://doi.org/10.1214/11-AOS887. 152

Rothenberg, G. E. (1968). "The Austrian Army in the Age of Metternich." *Journal of Modern History*, 40(2): 156–165. 151

Snijders, T. A. B., Steglich, C. E. G., and Schweinberger, M. (2006). *Longitudinal Models in the Behavioral and Related Sciences*, 41–71. Routledge. 135

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. MR3074604. doi: https://doi.org/10.1007/978-1-4614-1800-9_129. 135

Snyder, G. H. (1997). *Alliance Politics*. Cornell University Press. 151

Sohn, Y. and Park, J. H. (2017). "Bayesian Approach to Multilayer Stochastic Block Model and Network Changepoint Detection." *Network Science*, 5(2): 164–186. 139

van Dyk, D. A. and Park, T. (2008). "Partially Collapsed Gibbs Samplers." *Journal of the American Statistical Association*, 103(482): 790–796. MR2524010. doi: https://doi.org/10.1198/016214508000000409. 142

Vermeiren, J. (2016). *The First World War and German National Identity: The Dual Alliance at War*. Cambridge University Press. 151

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. MR3191998. doi: https://doi.org/10.1214/13-AOAS696. 134, 135

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. 135

Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*, Cambridge University Press. 137

Watanabe, S. (2010). "Asymptotic Equivalence of Bayes Cross Calidation and Widely applicable information criterion in singular learning theory." *Journal of Machine Learning Research*, 11: 3571–3594. MR2756194. 137, 145

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. MR2840178. doi: https://doi.org/10.1214/10-AOAS403. 135

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. MR3059083. doi: https://doi.org/10.1214/12-AOS1036. 138