Network Science **5** (2): 164–186, 2017.
doi:10.1017/nws.2017.13

Bayesian approach to multilayer stochastic blockmodel and network changepoint detection

YUNKYU SOHN

Department of Politics, Princeton University, Princeton, NJ 08540, USA
Department of Political Science, University of California San Diego, La Jolla, CA 92093, USA
(*e-mail:* ysohn@princeton.edu)

JONG HEE PARK

Department of Political Science and International Relations, Seoul National University, Seoul 151-742, Republic of Korea
(*e-mail:* jongheepark@snu.ac.kr)

Abstract

Network scholars commonly encounter multiple networks, each of which is possibly governed by distinct generation rules while sharing a node group structure. Although the stochastic blockmodeling—detecting such latent group structures with group-specific connection profiles—has been a major topic of recent research, the focus has been given to the assortative group discovery of a single network. Despite its universality, concepts, and techniques for simultaneous characterization of node traits of multilayer networks, constructed by stacking multiple networks into layers, have been limited. Here, we propose a Bayesian multilayer stochastic blockmodeling framework that uncovers layer-common node traits and factors associated with layer-specific network generating functions. Without assuming *a priori* layer-specific generation rules, our fully Bayesian treatment allows probabilistic inference of latent traits. We extend the approach to detect changes in block structures embedded in temporal layers of network time series. We demonstrate the method using synthetic multilayer networks with assortative, disassortative, core-periphery, and overlapping community structures. Finally, we apply the method to empirical social network datasets, and find that it detects significant latent traits and structural changepoints. In particular, we uncover endogenous historical regimes associated with distinct constellations of states in United States Senate roll call vote similarity patterns.

Keywords: *multilayer network, stochastic blockmodel, network changepoint detection, tensor decomposition, network time series, hidden Markov model*

1 Introduction

An observed network can be understood as a specific realization of a network generating function coupled with latent node traits. More often than not, network scholars encounter multiple realizations of networks, each of which is possibly governed by distinct generation rules, while sharing key node traits. For example, a romantic network and a friendship network exhibit strikingly different topological characteristics. Romantic networks are mainly disassortative (i.e. heterophilic, bipartite), whereas friendship networks tend to be assortative (i.e. homophilic). Suppose the sex label is unknown, and we want to infer the information from

the connection profiles of these two networks defined over an identical node set. Simultaneously, estimating common node traits of these two networks obviously provides us with a more statistically rigorous estimation of the unknown group structure than studying each separately does. Prompted by this idea, we develop a general statistical framework that allows a probabilistic inference of common latent node traits (e.g. sex) and layer-specific network generation functions (e.g. assortative or disassortative) from multilayer network data (e.g. romantic and friendship networks).

Statistical inference of latent traits of networks has been a major topic of methodological advance in network science. The majority of methods aim at partitioning nodes into subsets with similar connection profiles. Various techniques have been developed by imposing particular constraints, including assortative grouping methods (community detection) (Bickel & Chen, 2009; Girvan & Newman, 2002; Hoff et al., 2002; Newman, 2006; Guimera & Amaral, 2005; Palla et al., 2007; Ahn et al., 2010; Hoff, 2005; Mucha et al., 2010; Lancichinetti et al., 2011; Moody & White, 2003) and assortativity constraint-free methods (stochastic blockmodel) (Holland et al., 1983; Karrer & Newman, 2011; Nowicki & Snijders, 2001; Newman & Leicht, 2007; Airoldi et al., 2009), and by incorporating specific optimization schemes, including single metric maximization (Girvan & Newman, 2002; Guimera & Amaral, 2005; Ahn et al., 2010; Mucha et al., 2010), spectral (Newman, 2006; Rohe et al., 2011), random walk (Rosvall & Bergstrom, 2008), and probabilistic inference (Hoff et al., 2002; Holland et al., 1983; Newman & Leicht, 2007; Handcock et al., 2007; Hoff, 2005; Hofman & Wiggins, 2008; Airoldi et al., 2009; Lancichinetti et al., 2011; Gopalan & Blei, 2013); for a comprehensive review, see Fortunato (2010).

Among available frameworks for recovering latent traits of networks, we focus on the stochastic blockmodel (Holland et al., 1983; Karrer & Newman, 2011; Nowicki & Snijders, 2001). The stochastic blockmodel is a generalized group-based network formulation, which accommodates an arbitrary group-to-group connection profile among node groups, including assortative, disassortative, core-periphery, and overlapping community structures. While numerous techniques are available for the single-layer network modeling, it remains as a fairly challenging task to recover latent traits from multiple networks with the exception of recent studies.

Generally speaking, multilayer network[1] data can be decomposed into two key factors. The first factor is *layer-common node traits* indicating node group division or role diversification common to multiple layers of networks. The second factor is *layer-specific generation rules*, which indicate layer-specific functions linking node traits with layer-specific network realizations. A generalized statistical framework for multilayer network modeling requires the joint estimation of layer-common node traits and layer-specific generation rules in a way that allows a probabilistic inference of these estimates in order to account for the stochastic nature of the data generating processes.

Considering these requirements, limitations of existing methods seem obvious. Existing methods for latent trait recovery suffer from the lack of a capacity to analyze non-assortative generation rules (Ahn et al., 2010; Mucha et al., 2010; Hoff

[1] For general reviews of multilayer networks, see Boccaletti et al. (2014) and Kivelä et al. (2014).

https://doi.org/10.1017/nws.2017.13 Published online by Cambridge University Press

et al., 2002; Hoff, 2005; Hofman & Wiggins, 2008). Also, the most well-known multilayer community detection method (Mucha et al., 2010) collapses layer-specific factors by pulling all layer-specific connections into a combined set and conducting the community detection process using a single objective function. As a result, this method is incapable of detecting common group structures of nodes belonging to multiple layers when each layer does not exhibit a homogeneous generation rule. Moreover, the majority of existing tensor-based multilayer clustering algorithms (e.g. Mucha et al. (2010) and Kolda & Bader (2009)) employ deterministic methods or single-run stochastic algorithms, which do not allow probabilistic interpretation of their estimates, thereby being vulnerable to over-fitting (Lancichinetti et al., 2011). While not proposed in a Bayesian manner (Han et al., 2015) or not allowing the non-assortative structure recovery (Paul & Chen, 2016), very recent studies incorporate similar formulations to ours and provide consistency proofs for the multilayer stochastic blockmodel inference.

In this article, we present a fully Bayesian approach to the multilayer extension of the stochastic blockmodel by incorporating recent advances in Bayesian parallel factor analysis (PARAFAC) for tensor (or array) data (Kolda & Bader, 2009; Hoff, 2011). Our method is capable of examining multiple forms of large-scale organizations (e.g. assortative, disassortative, core-periphery structures) by employing a minimum level of null modeling and optimization constraints. The model produces layer-specific estimates representing layer-specific network generating processes as well as latent node traits common to a set of layers. Also, the Bayesian framework allows us to incorporate additional levels of model complexity that account for interesting real-world phenomena in network dynamics. An extension introduced in this article is the multiple changepoint detection for latent structures in networks. Our approach provides probabilistic estimates for all quantities of interests, including regime-specific latent node traits, layer-specific generation rules, and timings of structural changes.

2 Problem statement

Figure 1 graphically displays our goal. Panel A shows a two-layer multilayer network. Note that the two layers have identical grouping of nodes despite the fact that the first layer contains an assortative network and the second layer contains a disassortative network. Connections among same group members are colored in red and blue, and inter-group links are colored in gray. Dashed gray lines between the two layers indicate node identity. The first two plots in the bottom row show layer-specific adjacency matrices. The top figure of Panel B shows the objective of the multilayer network analysis where identical groups and distinct generation rules are properly identified. In contrast, the bottom figure of Panel B shows a simple aggregation representation of the two networks where the aggregation collapses the group structure of the original multilayer network.

We aim to extract *latent traits* of a *node-aligned* multilayer network $\mathscr{A} \equiv [\mathbf{A}_t | t \in T] = [A_{ijt} | i, j \in V, t \in T]$. We denote $\mathscr{A}$ as a tensor with tth layer consisting of an adjacency matrix $\mathbf{A}_t$. $\mathscr{A}$ can be represented by an adjacency tensor, which is a three-order array data: $\mathscr{A} \in \mathbb{R}^{N \times N \times |T|}$ where T is a set consisting of the array index and the number of nodes $N \equiv |V|$ in the node set V. In the article, we focus on symmetric network layers where a network in each layer contains undirected

https://doi.org/10.1017/nws.2017.13 Published online by Cambridge University Press

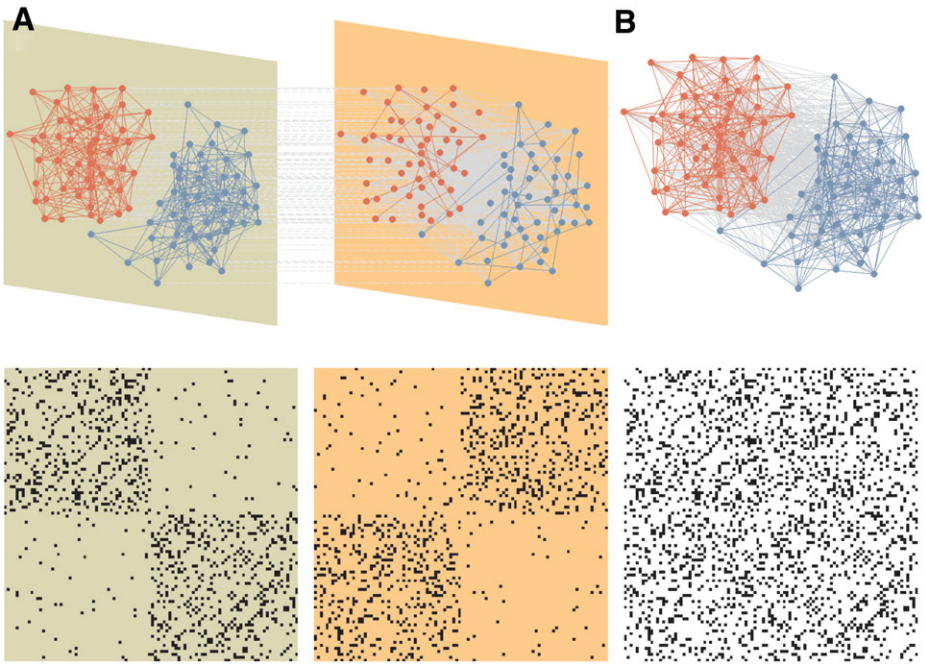

Fig. 1. Motivation for the multilayer representation. (A) A node-aligned multilayer network consisting of two layers, where the first layer is an assortative network whereas the second layer is a disassortative network. Networks in the two layers have identical grouping of nodes. Connections among same group members are colored in red and blue and inter-group links are colored in gray. Dashed gray lines between the two layers indicate node identity. Matrices in the bottom represent layer-specific-adjacency-matrices. (B) Simple aggregation representation of the two layer networks. The adjacency matrix in the bottom demonstrates that the aggregation collapses the group structure of the original multilayer network. (Color online)

links: $A_{ijt} = A_{jit}\ \forall i, j, t$. By *node-aligned*, we indicate that each layer of a multilayer network has an identical set of nodes V to each other.

Latent traits include key underlying variables that serve as layer-common node-specific traits as well as layer-specific functions of network generation (Faust, 1988; Handcock et al., 2007; Hoff et al., 2002; Hoff, 2005). For the case of a single-layer-network, when projected on a one-dimensional trait space, a network $\mathbf{A}$ can be approximated using a generating function $g(\cdot, \lambda)$ and a latent node trait vector $\mathbf{u}$. A statistical modeling procedure reverse-engineers unobserved $\mathbf{u}$ and $g(\cdot, \lambda)$ from $\mathbf{A}$ where $\mathbf{A} \approx g(\mathbf{u}, \lambda)$. It is well known that existing latent space modeling schemes can be categorized into two groups following the scale of $\mathbf{u}$ (Hoff, 2008). On the one hand, the latent position models assume a continuous-valued latent space and recover latent node positions in a continuous scale. Models belonging to this group can be classified into different classes depending on the shape of its kernel function (i.e. similarity metric). For example, the latent distance model assumes A_{ij} as a function of a metric distance between u_i and u_j (Hoff et al., 2002), and the random dot product model assumes A_{ij} as a function of $u_i u_j$ (Hoff, 2005). The second group of the latent space modeling scheme is the latent class model, in which each element

https://doi.org/10.1017/nws.2017.13 Published online by Cambridge University Press

of $\mathbf{u}$ is a categorical variable denoting the membership of each node (Holland et al., 1983; Karrer & Newman, 2011; Nowicki & Snijders, 2001; Newman & Leicht, 2007).

In practice, most well-known latent class models uncover latent node positions first and, by using the estimated low-dimensional positional coordinates, examine the possibility of a group structure (Krzakala et al., 2013; Newman, 2006; Ng et al., 2002; Riolo & Newman, 2014; Rohe et al., 2011; Sussman et al., 2012; Lei & Rinaldo, 2014). In this step-by-step estimation procedure, the goal of the group label inference is to map $h(\mathbf{u_i}) \rightarrow c_i$ where c_i is the group label of node i and h is a conversion function for clustering. Taken together with $\mathbf{u}$, layer-specific generating function $g(\cdot, \cdot, \lambda_t)$ recovers the latent block diagram of network at tth layer. We adopt this step-by-step procedure.

By definition, the stochastic blockmodel assigns an identical connection probability to node pairs (i,k) and (j,k) $\forall k$, if nodes i and j belong to the same group (i.e. block). As a result, nodes from an identical group take an equivalent role in generating a network, and they have identical expected degrees. In order to account for the degree heterogeneity of nodes belonging to a common group, as found in most empirical networks, we adopt a degree correction procedure which will be introduced in the next section.

If the traits are almost identical across the entire set of layers, the estimation of common latent traits is straightforward. However, layer-specific networks may not always have similar node traits across all layers (Holme & Saramäki, 2012). One special case is changes in the block structures of a temporal network. In this case, one needs to divide layers into proper subsets to identify state-specific (i.e. regime-specific) node traits that are assumed to be constant in each subset of layers. We call the layer subsets, each corresponding to a unique state, as *hidden network regimes*, and model their transitions to follow a non-ergodic discrete-time discrete-space Markov process (Chib, 1998). The method is applicable to an arbitrary multilayer network with sequentially ordered layers (i.e. $T_t < T_{t+1}$ for $1 \leqslant t \leqslant |T| - 1$).

3 Multilayer extension of stochastic blockmodel

3.1 Model

Let N be the total number of nodes and G denote the total number of groups. Also, let $\mathbf{C}$ be an $N \times G$ matrix of node group membership underlying the generation of multiple networks. Then, our final goal is to approximate each layer of $\mathscr{A} = [\mathbf{A}_t]$ using a common $\mathbf{C}$ for $t \in T$. Each row vector indicates the corresponding node's membership so that $C_{ik} = 1$ denotes that node i belongs to group k. Since we only allow a single group membership, $C_{ik} \in \{0, 1\}$ and $\sum_k C_{ik}$ is 1.

Before clustering nodes to obtain the group membership matrix $\mathbf{C}$, we first need to infer positional coordinates of nodes in a low-dimensional continuous space. Suppose we have a one-dimensional trait space. Following the previous approaches incorporating null models for the block structure detection (Newman, 2006; Karrer & Newman, 2011), the network in each layer can be understood by using a specific realization of the following model:

$$A_{ijt} = g(u_i, u_j, \lambda_{ijt}) + P_{ijt} \tag{1}$$

https://doi.org/10.1017/nws.2017.13 Published online by Cambridge University Press

where u_i and u_j indicate node-specific variables assigned to i and j respectively, t is layer index, and λ_{ijt} is a pair-specific factor that is not node-specific. P_{ijt} is the expected number of edges calculated using a null model. For a network induced by the blockmodel, since this formulation controls for the degree heterogeneity by introducing the null model $\mathbf{P}$, $g(u_i, u_j, \lambda_{ijt})$ explains the tie formation tendency between nodes i and j, which attributes to their block memberships.

Here, we adopt a random dot product formulation as a specification of Equation (1), so that

$$\begin{aligned} A_{ijt} &= g(u_i, u_j, \lambda_{ijt}) + P_{ijt} \\ &= \lambda_{ijt} u_i u_j + P_{ijt} \end{aligned} \tag{2}$$

(De Lathauwer et al., 2000; Kolda & Bader, 2009; Hoff, 2011). Note that for a single-layer network, this representation is equivalent to the spectral decomposition of $\mathbf{A} - \mathbf{P}$.

We assume that λ_{ijt} is constant for every node pair in network layer t: $\lambda_{ijt} = \lambda_t$ $\forall i, j$.[2] The value of λ_t is estimated in a continuum. Controlling for P_{ijt}, as a node pair becomes close in the $\mathbf{u}$ space, a positive λ_t implies a higher probability of link formation between i and j, whereas a negative λ_t implies a lower probability of link formation. Proximal positions in the $\mathbf{u}$ space imply similar connection profiles. According to the social network terminology, this relationship among proximal nodes refers to the structural equivalence in network formation (Faust, 1988).

After computing $\mathbf{u}$, we aim to find a common $\mathbf{C}$ for $\mathbf{A}_t$ $t \in T$, which satisfies

$$\mathbf{A}_t \approx \mathbf{C}\mathbf{Z}_t\mathbf{C}^T + \mathbf{P}_t \tag{3}$$

where $G \times G$ matrix $\mathbf{Z}_t$ is a symmetric group-to-group connection profile matrix with Z_{lmt} representing the expected edge weight between a node pair belonging to groups l and m, controlling for $\mathbf{P}_t$.

A clear advantage of employing this step-by-step procedure, instead of incorporating a direct discrete classification approach, is that we do not need to specify the number of groups and the structures of block diagrams before conducting the analysis. Instead of having the group label formulation, using the low-dimensional embedding approach reduces a substantial amount of the computational load. This is due to the fact that the low-dimensional coordinates and layer-specific parameters produced by our random dot formulation endogenously determine the shapes of block diagrams. This becomes an enormous advantage especially in the case when there are many free parameters to estimate, including the changepoint extension we will introduce shortly.

3.2 Bayesian computation

Our computational scheme heavily relies on recent advances in Bayesian parallel factor analysis (PARAFAC) for tensor data (Kolda & Bader, 2009; Hoff, 2009, 2011). In Hoff (2009, 2011)'s Bayesian method for factor decomposition of tensor data, recovered node traits are random error components required primarily for

[2] In the R-dimensional case, λ_t becomes an R-dimensional vector: $\langle \lambda_t^1, \cdots, \lambda_t^R \rangle$.

https://doi.org/10.1017/nws.2017.13 Published online by Cambridge University Press

the consistent estimation of regression slopes (Ward et al., 2013). Thus, estimates of recovered node traits are sensitive to the specification of regression covariates and there is no guarantee that estimates of recovered node traits fully represent topological characteristics of multiple networks. We will show that, instead of incorporating exogenous covariates, adopting a degree correction null model in the estimation stage makes it possible to recover a trait space that can be interpreted as a structural equivalence space (Faust, 1988; Hoff et al., 2002; Hoff, 2005).

Let $\mathbf{B}_t \equiv \mathbf{A}_t - \mathbf{P}_t$ and $\mathscr{B} \equiv [\mathbf{B}_t | t \in T]$. Now assume a generalized latent space with R dimensions. To infer $N \times R$ vector array $\mathbf{U}$ and $R \times R$ diagonal matrix $\mathbf{\Lambda}_t = diag(\lambda_t^1, \cdots, \lambda_t^R)$, using a layer specific null model $\mathbf{P}_t$,

$$\begin{aligned}\mathbf{B}_t &= \mathbf{A}_t - \mathbf{P}_t \\ \mathbf{B}_t &= \mathbf{U}\mathbf{\Lambda}_t\mathbf{U}^T + \mathbf{E}_t \\ \mathbf{E}_t &\sim \text{matrix normal}(\mathbf{0}, \sigma^2\mathbf{I}_N, \mathbf{I}_N).\end{aligned}$$

Note that the subsequent Bayesian analysis is based on the weighted matrix $\mathbf{B}_t$, not directly on $\mathbf{A}_t$. Thus, the method is applicable to both discrete (binary, ordinal, or count) and continuous network data.

For Bayesian estimation, we adopt Hoff (2011)'s hierarchical Bayesian scheme for the three mode $(\mathbf{U}, \mathbf{V}, \mathbf{W})$ array data where $\mathbf{W} = \mathbf{U}$, due to the symmetry condition of $\mathbf{A}$, and $\mathbf{V}$ is a $|T|$ by R matrix containing $\boldsymbol{\lambda}_t = (\lambda_t^1, \cdots, \lambda_t^R)$ in each row. The prior distributions of the hierarchical model are as follows:

$$\begin{aligned}\{\mathbf{u}_1^k, , \ldots, \mathbf{u}_N^k\} &\sim \text{multivariate normal}(\boldsymbol{\mu}_k, \Psi_k) \\ \boldsymbol{\mu}_k | \Psi_k &\sim \text{multivariate normal}(\boldsymbol{\mu}_{0,k}, \Psi_k/\kappa_0) \\ \Psi_k &\sim \text{inverse Wishart}(W_{0,k}, v_{0,k}) \\ \mathbf{U}^{(k)} &\sim \text{matrix normal}(\mathbf{M} = \mathbf{1}\boldsymbol{\mu}_k^T, \mathbf{I}_N, \Psi_k) \\ \sigma^2 &\sim \text{inverse gamma}(c_0, d_0).\end{aligned}$$

where $\mathbf{u}_i$ is the ith $(1 \times R)$ row vector, multivariate normal $(\boldsymbol{\mu}_k, \Psi_k)$ is the R-dimensional Gaussian distribution with mean vector $\boldsymbol{\mu}_k$ and variance matrix Ψ_k, inverse Wishart $(W_{0,k}, v_{0,k})$ is the inverse Wishart distribution, which is a conjugate prior for the covariance matrix of a multivariate normal distribution, and inverse gamma(c_0, d_0) is the inverse gamma distribution, which is a conjugate prior for the variance parameter σ^2. Due to their conjugacy, we can construct a Gibbs sampling algorithm for the estimation of the parameters of interest.

The posterior distribution is augmented by hierarchical parameters:

$$p(\mathbf{U}, \mathbf{V}, \sigma^2 | \mathscr{B}) = \int p(\mathbf{U}, \mathbf{V}, \boldsymbol{\mu}_u, \Psi_u, \boldsymbol{\mu}_v, \Psi_v, \sigma^2 | \mathscr{B}) d\boldsymbol{\mu}_u d\Psi_u d\boldsymbol{\mu}_v d\Psi_v \tag{4}$$

The parameters of the integral kernel can be sampled from the following blocked Gibbs sampler (Liu, 1994; Hoff, 2011):

1. Sample $p(\mathbf{U}, \boldsymbol{\mu}_u, \Psi_u | \mathscr{B}, \mathbf{V}, \boldsymbol{\mu}_v, \Psi_v, \sigma^2)$ using the following equality:

$$\begin{aligned}p(\mathbf{U}, \boldsymbol{\mu}_u, \Psi_u | \mathscr{B}, \mathbf{V}, \boldsymbol{\mu}_v, \Psi_v, \sigma^2) = &p(\mathbf{U} | \boldsymbol{\mu}_u, \Psi_u, \mathscr{B}, \mathbf{V}, \boldsymbol{\mu}_v, \Psi_v, \sigma^2) \\ &p(\boldsymbol{\mu}_u | \Psi_u, \mathscr{B}, \mathbf{V}, \boldsymbol{\mu}_v, \Psi_v, \sigma^2) \\ &p(\Psi_u | \mathscr{B}, \mathbf{V}, \boldsymbol{\mu}_v, \Psi_v, \sigma^2)\end{aligned} \tag{5}$$

https://doi.org/10.1017/nws.2017.13 Published online by Cambridge University Press

2. Sample $p(\mathbf{V}, \boldsymbol{\mu}_v, \Psi_v | \mathscr{B}, \mathbf{U}, \boldsymbol{\mu}_u, \Psi_u, \sigma^2)$ using the following equality:

$$\begin{aligned} p(\mathbf{V}, \boldsymbol{\mu}_v, \Psi_v | \mathscr{B}, \mathbf{U}, \boldsymbol{\mu}_u, \Psi_u, \sigma^2) = & p(\mathbf{V} | \boldsymbol{\mu}_v, \Psi_v, \mathscr{B}, \mathbf{U}, \boldsymbol{\mu}_u, \Psi_u, \sigma^2) \\ & p(\boldsymbol{\mu}_v | \Psi_v, \mathscr{B}, \mathbf{U}, \boldsymbol{\mu}_u, \Psi_u, \sigma^2) \\ & p(\Psi_v | \mathscr{B}, \mathbf{U}, \boldsymbol{\mu}_u, \Psi_u, \sigma^2) \end{aligned} \tag{6}$$

3. Sample $p(\sigma^2 | \mathscr{B}, \mathbf{U}, \boldsymbol{\mu}_u, \Psi_u, \mathbf{V}, \boldsymbol{\mu}_v, \Psi_v)$.

The samples of this iterative simulation are simulation equivalent to direct samples from $p(\mathbf{U}, \mathbf{V}, \sigma^2 | \mathscr{B})$ after integrating out augmented parameters (Gelfand & Smith, 1990; Smith & Roberts, 1993). The detailed sampling algorithms are discussed in Supplementary Information.

3.3 Degree correction

A crucial weakness of the naive stochastic blockmodel framework, lacking the degree correction component $\mathbf{P}$, is that nodes belonging to same group have identical expected degree, thereby not allowing degree heterogeneity within a group (Karrer & Newman, 2011; Chaudhuri et al., 2012). One way to resolve this problem is to employ the degree correction approach, which is similar in spirit to extracting covariance structure in the factor analysis (Karrer & Newman, 2011; Fortunato, 2010).

The development of null models for degree correction (i.e. the specification of $\mathbf{P}$ for a single-layer network) has gained a particular attention among scholars studying clustering techniques for networks (Fortunato, 2010). A core component of such null models is to compute the baseline expected number of edges for a pair of nodes.

When inferring a group structure from a latent node trait distribution, node degree can be a reasonable element predicting the baseline edge weight for a node pair. Popular models such as modularity approach (Newman, 2006) and degree corrected stochastic blockmodel (Karrer & Newman, 2011) utilize this idea. For example, an element in the modularity null function is defined as

$$P_{ij}^{modul} = k_i k_j / 2e \tag{7}$$

where k_i is the degree of node i and e is the total number of edges, corresponding to the expected number of links between two nodes given degree values. It is proven that several well-known null models are approximately equivalent to modularity (Newman, 2013). As a result, they are interchangeable in most cases without the loss of generality. For example, when using the spectral elements of $\mathbf{A}$ for the block structure inference, by excluding its principal eigenvalue ($\lambda^{princ} = \lambda(\mathbf{A})_{max}$) and the associated eigenvector ($\mathbf{u}^{princ}$), we can define

$$\mathbf{P}^{princ} = \lambda^{princ} \mathbf{u}^{princ} \mathbf{u}^{princT} \tag{8}$$

While developed under the emphasis on the assortative structure recovery, studies on spectral properties of the adjacency matrix $\mathbf{A}$ and graph Laplacians suggest that eigenspectra of these matrices not only recover assortative structures but also recover arbitrary block structures of networks (Hoff, 2008; Nadakuditi & Newman, 2013; Peixoto, 2013; Rohe et al., 2011). The assortativity constraint in $\mathbf{Z}_t$ corresponds to non-negativity constraint in λ^r (Lee & Seung, 1999). Hence, in

https://doi.org/10.1017/nws.2017.13 Published online by Cambridge University Press

order to accommodate general classes of block diagrams for $\mathbf{Z}_t$, an algorithm needs to allow positional coordinates $\mathbf{u}^r$ corresponding to negative λ^r. In this way, one can succeed in the simultaneous estimation of a common node group structure underlying the generation of a multilayer network consisting of an assortative network layer and a disassortative network layer. Accordingly, in practice, selecting a stack of $\mathbf{u}^r$ corresponding to the largest $|\mathbf{u}^{r\,T}(\mathbf{A}-\mathbf{P})\mathbf{u}^r|$ sequentially yields recovery not limited to assortative structures (Rohe et al., 2011; Peixoto, 2013), in which the corresponding $\{\mathbf{u}^r\}^R$ are the positional coordinates corresponding to the largest $\{|\lambda^r|\}^R$. Due to this setting, our approach does not impose any restriction on the sign of diagonal elements in $\mathbf{\Lambda}_t$ and infers an arbitrary block diagram with symmetric group-to-group interactions.

3.4 Rank selection

An important procedure required for the valid estimation of node traits or a block structure is to choose a proper number of dimensions R for the stack of latent node trait vectors $\mathbf{U}$. This procedure is not required if a researcher knows the rank of group-to-group connection profiles matrix $\mathbf{Z}$ *a priori*. However, this parameter is usually unknown (Fishkind et al., 2013), and one should infer a rank value from the result of an analysis.

One feasible way to resolve this problem is via Bayesian model comparison by selecting a number which minimizes Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002; Hoff, 2011). DIC is an approximated version of the average out-of-sample deviance for the estimates of a parametric model. Accordingly, calculating this value helps one to examine the performance of a model controlling for its over-parameterization. DIC is defined as

$$DIC = \bar{D} + \tilde{p} \tag{9}$$

where $\bar{D}$ is the average of $-2\ln p(y|\theta)$ obtained through Markov chain Monte Carlo (MCMC) for the recovered parameters $\theta = \{\mathscr{B}, \sigma_t^2 | t \in T\}$, $\tilde{p}$ is the effective number of parameters which is equal to $\bar{D} + 2\ln p(y|\theta)$, and $\hat{\theta}$ is the posterior mean of θ. In words, R corresponding to the minimum value of DIC is the optimal number of dimensions derived by simultaneously considering the model fit and the model complexity.

3.5 Ex-post clustering

After obtaining the continuous latent trait values of nodes on the structural equivalence space, one needs to cluster nodes into groups. Clustering techniques for nodes in the low-dimensional trait space have been extensively studied in statistics (Lei & Rinaldo, 2014; Rohe et al., 2011; Sussman et al., 2012). Sussman et al. (2012) and Lei & Rinaldo (2014) show that, when applying the k-means clustering, the spectral components of an adjacency matrix perfectly recover the true node group structure if the size of a network is sufficiently large and its maximum expected degree is as small as $\log N$. Rohe et al. (2011) derive similar results for a graph Laplacian. Peixoto (2013) also obtains a similar result for a general matrix representation of networks using the random matrix approach.

https://doi.org/10.1017/nws.2017.13 Published online by Cambridge University Press

There are multiple options for the specification of $h(\mathbf{u_i}) \rightarrow c_i$, including k-means clustering of $R+1$ groups (Ng et al., 2002) on R-dimensional space. Another idea recently proposed to cluster nodes in a latent space is multiway partitioning which clusters each node on the basis of the proximity of its location to transformed $R+1$ simplex vectors (Riolo & Newman, 2014). A straightforward implementation of this approach for $R=1$ is to bisect node set based on their signs in $\mathbf{u}$ (Newman, 2006). In Supplementary Information, we introduce the clustering procedure.

To summarize, our statistical model to uncover group structures consists of the following two steps:

1. Recover common $\mathbf{U} \in \mathbb{R}^{N \times R}$.
2. Cluster $\mathbf{U}$ and find $\mathbf{C} \in \{0,1\}^{N \times R}$.

In the following section, we will discuss a detailed algorithmic procedure for the changepoint detection which adds the regime inference stage as oppose to the zero-changepoint case discussed so far.

3.6 Detecting changes in block structures

In many occasions, layer labels of multilayer network are sequentially ordered. In this case, researchers often wish to detect dramatic changes in the block structures. For example, the evolution of alliance networks in international politics may experience several breaks in the block structure due to World Wars or the end of the Cold War. However, there have been very few statistical methods for the network changepoint detection. Moreover, a majority of methods for network changepoints identify breaks in global-level network metrics such as the diameter or the average clustering coefficient (Akoglu et al., 2015), missing important information of changes in node-level traits. In our Bayesian framework, it is straightforward to incorporate the estimation of breaks, corresponding to the transitions of node-level traits, and the estimation of time-varying node traits and time-varying network generation rules.

Our task is to infer the locations of M changepoints and estimate regime-specific node traits and layer-specific network generating factors given the M changepoints (i.e. $M+1$ regimes). Let $s(t) : t \rightarrow \textit{regime}$ be a function mapping the layer t to a corresponding regime. In the changepoint model with the R-dimensional latent space, $A_{ijt} = g(u(s(t))_i, u(s(t))_j, \lambda_t^r | 1 \leqslant r \leqslant R)$ consisting of regime-specific $\mathbf{U}(s(t))$ and layer-specific $\langle \lambda_t^1, \ldots, \lambda_t^R \rangle$. We aim to endogenously determine $s(t)$, $\mathbf{U}(s(t))$ and $\mathbf{\Lambda}_t$ for each t.

A changepoint problem is identical to the problem of identifying a transition matrix for hidden regimes. In order to account for multiple changepoints, it is computationally efficient to assume a non-ergodic Markov process for the transition of hidden regimes (Chib, 1998). Due to the non-ergodic Markov process assumption, in our model, we assume that the process converges surely to the final regime as it reaches the last layer. To be specific, the transition matrix $\mathbf{\Gamma}$ is an $(M+1) \times (M+1)$ matrix in which elements other than $\gamma_{m,m}$ and $\gamma_{m,m+1}$ are equal 0 for $1 \leqslant m \leqslant M$ and

https://doi.org/10.1017/nws.2017.13 Published online by Cambridge University Press

$\gamma_{M+1,M+1} = 1$.[3] The resulting model is as follows:

$$\begin{aligned}
\mathbf{B}_t &= \mathbf{U}_{S_t} \mathbf{\Lambda}_t \mathbf{U}_{S_t}^T + \mathbf{E}_t \\
S_t | S_{t-1}, \mathbf{\Gamma} &\sim \text{Markov}(\mathbf{\Gamma}, \pi_0) \\
\gamma_{i,i} &\sim \text{Beta}(a_0, b_0) \\
\mathbf{E}_t &\sim \text{matrix normal}_{N \times N}(\mathbf{0}, \sigma^2 \mathbf{I}_N, \mathbf{I}_N).
\end{aligned}$$

The MCMC algorithm for the above model is similar with that in the previous case except the fact that now we sample S_t and then sample $\mathbf{U}, \mathbf{V}$ from the subset of data partitioned by hidden states. Note that all the $\mathbf{U}$-related parameters $(\mathbf{U}, \boldsymbol{\mu}_u, \mathbf{\Psi}_u)$ are now *state-dependent* unlike the previous model: $(\mathbf{U}_m, \boldsymbol{\mu}_{u,m}, \mathbf{\Psi}_{u,m})$. The simulation of hidden states can be done using Chib (1996)'s recursive method. The computational cost of the proposed algorithm scales linearly with the squared value of the number of nodes and the number of layers. See Supplementary Information for more information.

While the proposed approach is similar in spirit to existing algorithms for anomaly detection in dynamic graphs, in particular to the tensor decomposition-based event detection and clustering-based event detection methods (Akoglu et al., 2015), there are notable differences. The anomaly detection algorithms treat each layer separately and hence do not provide common estimates of node traits within the same regime (Araujo et al., 2014). Also, sliding-window-based hierarchical random graph approach is only capable of modeling assortative structures, lacking layer-specific estimates of generation rules (Peel & Clauset, 2015). In contrast, our approach generates regime-specific estimates of node traits coupled with layer-specific factors. Moreover, existing methods explore parameter changes over time, lacking a statistically valid criterion to identify hidden regime changes (Kolar et al., 2010). Our changepoint model allows a probabilistic assessment of hidden regime transitions in network time series data. Our approach is different to the layer-clustering method of Stanley et al. (2016), as they infer equivalent layers of a *non-ordered* multilayer network, based on latent node traits, without assuming a Markov transition process, which is an essential assumption to have for the time series analysis. Similarly, while Peixoto (2015) introduces an extensive network time series analysis method, it does not accommodate the Markov transition assumption so that the algorithm does not provide parametric estimates of fundamental changes in latent traits.

4 Simulation studies and empirical analysis

4.1 Simulation studies

In order to demonstrate the performance of the proposed approach, we generate multilayer network data from simple block diagrams, in which two nodes belonging to the same cluster have p_{in} chance of connection whereas two nodes belonging to different clusters have p_{out} chance of connection. The number of clusters is set to be 3. The top panels (A and B) of Figure 2 show adjacency matrices of synthetic

[3] $p_{m,n}$ is the probability of moving from regime m to regime n and $\gamma_{m,m+1} = 1 - \gamma_{m,m}$.

https://doi.org/10.1017/nws.2017.13 Published online by Cambridge University Press

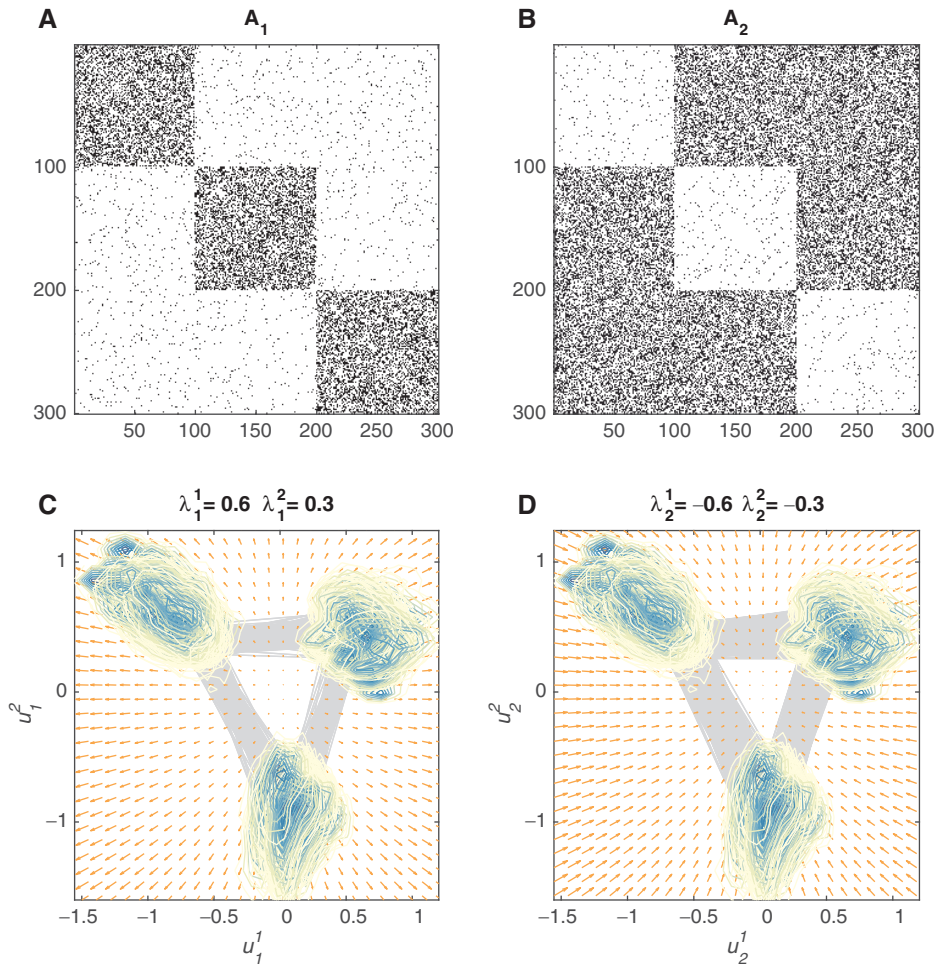

Fig. 2. $R = 2$ Bayesian rank reduction results of the synthetic two-layer example. (A) An assortative network $\mathbf{A}_1$ with three node groups having $p_{in} = 0.37$ and $p_{in} = 0.02$. (B) A disassortative network $\mathbf{A}_2$ with three node groups having $p_{in} = 0.02$ and $p_{in} = 0.37$. After conducting the joint Bayesian analysis of $\mathscr{B} = \{\mathbf{A}_1 - \mathbf{P}_1^{modul}, \mathbf{A}_2 - \mathbf{P}_2^{modul}\}$, we obtain common node positions with confidence regions and layer-specific λ. (C) Recovered estimates of $\mathbf{A}_1$ through 20,000 MCMC simulations with 10,000 burn-in trials. Vector fields indicate the magnitude and the direction of λ_t^r on each axis of the latent space. Arrows pointing outward indicate the assortative tendency of link formation on both axes. Gray lines indicate links between node groups. λ_1^1 and λ_1^2 are mean posterior estimates. (D) Recovered estimates of $\mathbf{A}_2$. Arrows pointing inward indicate the disassortative tendency of link formation on both axises. (Color online)

two-layer networks with planted block structures. Panel A displays an assortative network ($p_{in} \gg p_{out}$) with three node groups and Panel B displays a disassortative network ($p_{in} \ll p_{out}$) with three node groups. Both networks are undirected and unweighted, and planted group labels are identical.

The results of our analysis are presented in the bottom panel of Figure 2. For the synthetic networks with three groups corresponding to rank-3 blocks in Figure 2, the minimum DIC is observed at $R = 2$ (Table 1). This result is consistent with

https://doi.org/10.1017/nws.2017.13 Published online by Cambridge University Press

Table 1. *DIC for the multilayer network depicted in Figure 2.*

	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K=6$
DIC	212666.5	201067.9	201170.9	201151.0	201279.5	201365.3

asymptotic findings in the spectral decomposition literature, which states that a stochastic network induced by a block diagram with rank $R+1$ is sufficient to have R dimensions for its representation in a dot-product space when its degree-related information is excluded (Hoff, 2008; Peixoto, 2013).[4]

By applying the proposed Bayesian method on $\mathscr{B} = \{\mathbf{A}_1 - \mathbf{P}_1^{modul}, \mathbf{A}_2 - \mathbf{P}_2^{modul}\}$, we obtain MCMC samples of common node positions and layer-specific generation rules. In order to visually convey the properties of λ_t, we employ a vector velocity representation. For a layer-specific-network with λ^1 for the first dimension and λ^2 for the second dimension, the vector field function is given by $\lambda^1\hat{\mathbf{e}}_1 + \lambda^2\hat{\mathbf{e}}_2$. In Figure 2(C), vectors point outward in both dimensions, indicating assortative tendencies in both axes. In contrast, vectors in Figure 2(D) point inward, representing disassortative tendencies on both axes. The recovered node positions and their probabilistic distributions are drawn by contour plots in Figures 2(C) and (D). Dark colors denote regions with a higher posterior probability than lighter colors do. The distribution of the contours corresponds to the conditional distribution of node positions given the data: $p(\mathbf{U}|\mathscr{B})$. The gray lines indicate links between nodes.

Next, we test our method to detect changes in the network block structure using synthetic network time series data. We test our method for four generic types of network changes: cluster merging, disassortative to assortative transition, modular to core-periphery transition (Borgatti & Everett, 2000), and gradual cluster overlap (Ahn et al., 2010; Palla et al., 2007). Each type of block structure changes is graphically displayed in the top of each panel in Figure 3.

We generated 40-layer time series network data for each type. In Panel A, 30-node-networks consisting of three 10-node-clusters transform into two-cluster networks. The parameter p^1 is the merging parameter, indicating the probability of inter-group connections between red and yellow nodes. p^1 (dotted black line) changes from 0.05 to 0.5. In Panel B, 30-node-networks consisting of three 10-node-groups change from disassortative to assortative structures. The parameter p^2 (dotted black line) is the assortative transition parameter, which becomes 0 when $p_{out} = 0.5$ and $p_{in} = 0.05$, and 1 when $p_{out} = 0.05$ and $p_{in} = 0.5$. In other words, $p^2 = \frac{p_{in} - p_{out} + 0.45}{0.9}$. In Panel C, 20-node-networks consisting of two 10-node-groups with assortative connections transform into core-periphery networks in which intra-group connection probability between the blue nodes p^3 (dotted black line) drops from 0.5 to the baseline connection probability 0.1. In Panel D, 20-node-networks consisting of two 10-node-groups with assortative connections transform into a densely overlapping community structure. In this example, the parameter p^4 (dotted black line) denotes the size of each cluster to the size of the whole network. As a result, for the 20 nodes case

[4] The degree correction procedure eliminates the degree-related dimension in the recovered latent space. For example, the principal eigenvector of an adjacency matrix for a single-layer network is a type of eigenvector centrality and the first-order approximation of the principal eigenvector corresponds to node degree (Krzakala et al., 2013; Peixoto, 2013).

https://doi.org/10.1017/nws.2017.13 Published online by Cambridge University Press

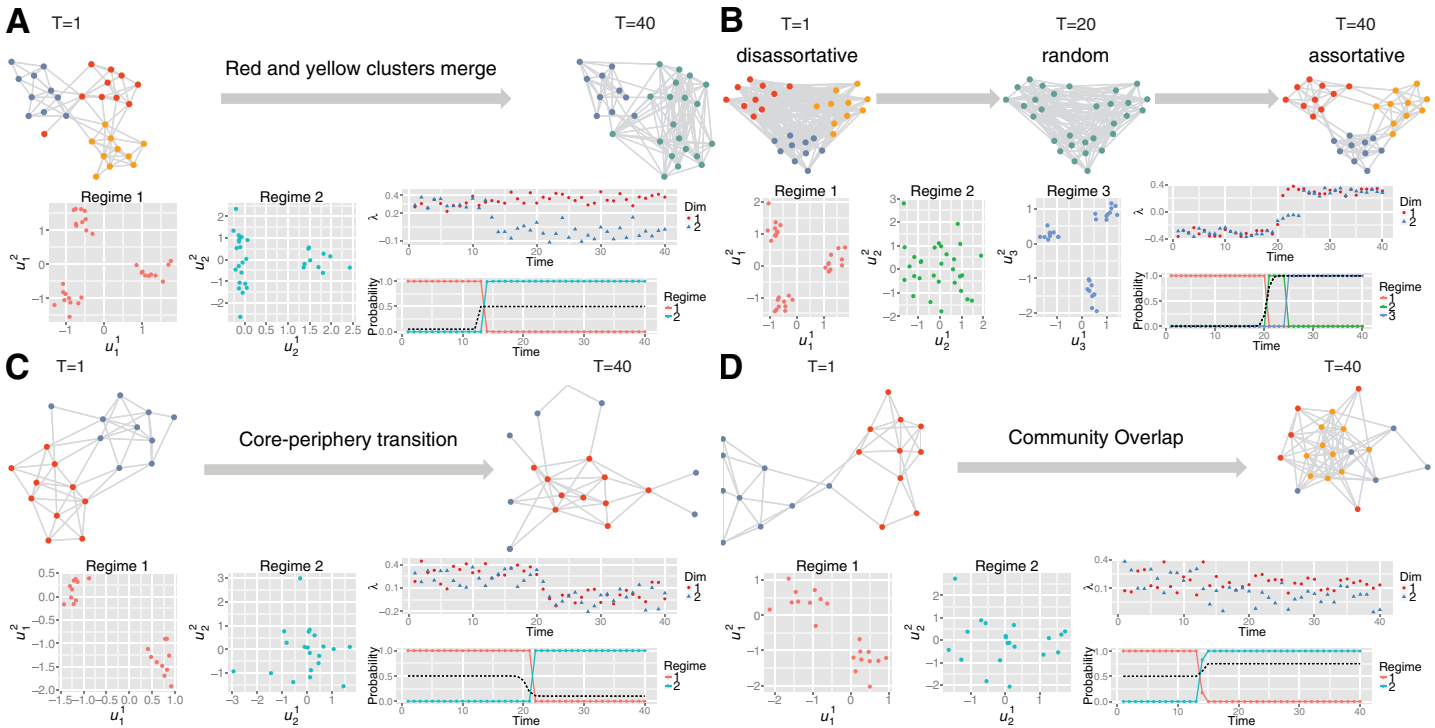

Fig. 3. Changepoint detection results for synthetic planted times series. For each panel, the structures of representative layers of the 40-layer network are depicted on the top. Below, inferred node positions for each regime are plotted. Each right middle subplot indicates the mean λ_t^r estimate on each dimension. Each right bottom subplot shows the regime probability drawn from the changepoint analysis. Black dashed lines indicate the values of the planted single parameters, $\{p^1, p^2, p^3, p^4\}$, for the synthetic time series network data over time. Each time series represents a specific type of network structural changes: (A) cluster merging, (B) disassortative to assortative block structure transition, (C) community structure to core-periphery structure transition, and (D) non-overlapping community structure to overlapping community structure transition, respectively. For detailed explanation of the figure, please see the main text. We conduct 20,000 MCMC simulations and 10,000 burn-in trials for each example. (Color online)

depicted here, $p^4 = 0.5$ corresponds to a non-overlapping structure with 10 nodes for each cluster, whereas $p^4 = 0.75$ refers to a network consisting of 10 nodes with overlapping cluster memberships and 10 nodes with non-overlapping cluster memberships. We employed modularity as a null function in the estimation.

Overall, our method correctly uncovers planted parameter changes in all four cases (Figure 3 bottom right subplots). Two scatter plots in the bottom left of each panel show estimates of latent node traits ($\mathbf{U}_m$) before and after the breaks. The planted block structures are well identified by our method. Note that our method correctly identifies the intermediate transition period, consisting of random networks, in Panel B, uncovering three hidden regimes instead of two. Two plots in the bottom right of each panel display changes in layer-specific λ_t^r (top right) and estimated regime probabilities (bottom right), respectively. In Figure 3(A), for example, we observe a significant drop of λ_t^2 reaching 0 in the second regime, as a result of the transition from three-cluster networks to two-cluster networks. Similarly, in Figure 3(C), the transition of the blue cluster into the peripheral node group, by losing its modular structure, yields a notable drop in both λ_t^1 and λ_t^2. The R-implemented version of the algorithm, used in the study, is reasonably fast, producing 100 samples in 30 seconds on a laptop for the simulated examples analyzed. It is still affordable for larger networks with 1,000 nodes, producing 100 samples within an hour. We expect that C/C++ implementation of the algorithm will yield a substantial improvement in its computational speed.

4.2 Amplifying resolution by joint modeling of multiple networks

Existing studies on spectral graph theory suggest that most graph clustering techniques for a single-layer network suffer from the resolution limit problem. That is, for a sufficiently dense network, it is impossible to identify node groups when $|c_{in} - c_{out}| < q\sqrt{c}$ where c_{in}/N and c_{out}/N denote the probabilities of within-block and without-block link formation, respectively, and $c = \{c_{in} + (q-1)c_{out}\}/q$ denotes the average degree for a network with q equal size groups (Krzakala et al., 2013; Fortunato & Barthelemy, 2007; Nadakuditi & Newman, 2013; Peixoto, 2013). Yet, in practice, given the stochastic nature of the data generating process, well-known detection methods perform worse than the theoretical expectation for finite size networks (Figure 4).

A feasible solution for this problem is to extend the number of observations sharing a node set and group labels. Figure 4 illustrates that our multilayer extension powers up the quality of classification practice. Each network layer in Figure 4 is a realization of a stochastic blockmodel having the same structure (i.e. the three-group homophilic block diagram) with the homophilic network ($N = 300$ and $c = 15$) depicted in Figure 1 (A). As the number of network layers sharing the same group structure increases, the quality of node group labeling improves.

After the identification of the stack of the common node position vectors $\mathbf{U}$, we group the nodes using k-means clustering. Variation of information (Meila, 2003) measured between the recovered labels and the planted labels reaches 0 when these two group label vectors become identical. We observe that, for the single layer case, variation of information is greater than 0 (i.e. the two label vectors are identical) over the entire range of $c_{in} - c_{out}$, despite the fact that the majority of $c_{in} - c_{out}$

https://doi.org/10.1017/nws.2017.13 Published online by Cambridge University Press

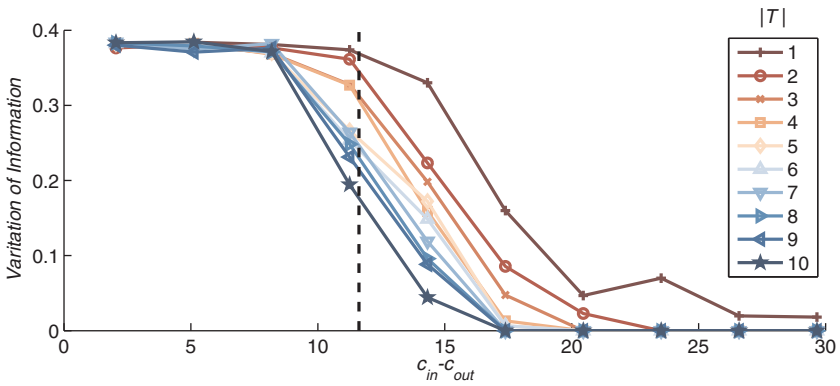

Fig. 4. Impact of the number of layers on the clustering quality. $|T|$ is the number of layers used for the estimation. $\mathbf{P}^{princ}$ is used as the null model Krzakala et al. (2013). Variation of Information is a normalized dissimilarity index which becomes 0 when the group label inferred from the recovered node trait is identical to the planted group labels Meila (2003). Network on each layer is a realization of a three-group homophilic block diagram with the average degree $c \approx 15$ and 100 nodes for each group. Black dashed line indicates the theoretical resolution limit $3\sqrt{15}$ (Krzakala et al., 2013; Fortunato & Barthelemy, 2007; Nadakuditi & Newman, 2013; Peixoto, 2013). We conduct 20,000 MCMC simulations and 10,000 burn-in trials for each case. (Color online)

values exceed the resolution limit. In contrast, increasing the number of layers does diminish variation of information for every $c_{in} - c_{out}$ value, and having five layers guarantees a perfect recovery of the planted node groups above a certain value of $c_{in} - c_{out}$.

4.3 Multilayer social networks online and offline

We analyze real-world multilayer network data among 61 employees (professors, postdoctoral researchers, PhD students, and administration staff) of the Department of Computer Science at Aarhus University (Magnani et al., 2013). The dataset records five types of online and offline social ties covering current working relationships, repeated leisure activities, regularly eating lunch together, co-authorships, and friendships on Facebook, respectively.

Our goal is to identify latent social groups of 61 employees across the five layers, based on the assumption that common latent social groups breed connections on each layer *via* layer-specific generation rules. We conduct $R = 2$ rank reduction on the five-layer network, the results of which are reported in Figure 5. Top panels indicate group-to-group connection matrices, where the number on each block represents the number of links between the corresponding group pair. Bottom left panel shows the recovered node traits with the vector field representing λ_t^r for the network of dyads having lunch together. Bottom right panel shows two-dimensional λ_t^r of each layer. The three groups identified using Riolo & Newman (2014)'s method are denoted as {R,G,B}, respectively in the bottom left. All networks exhibit clear homophilous structures.

The distribution of node positions on the orthogonal axes helps us find differential factors affecting the generation of multiple types of relationships. Given the common

https://doi.org/10.1017/nws.2017.13 Published online by Cambridge University Press

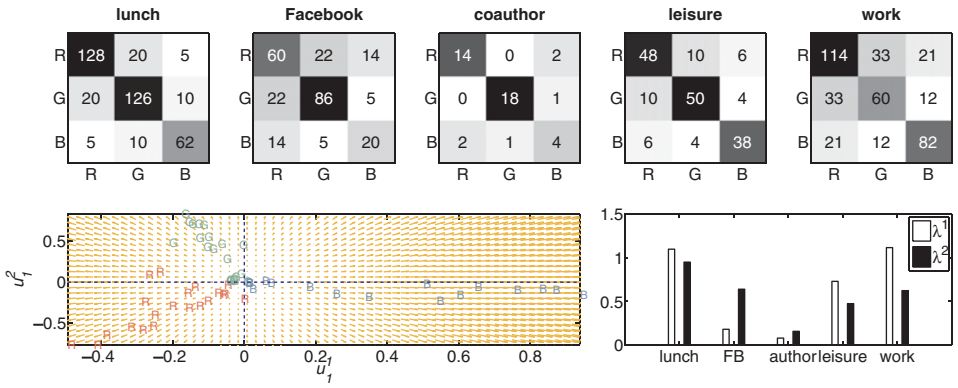

Fig. 5. Common social groups inferred from five-layer online and offline interaction records among 61 members of a computer science department. The top panels indicate group-to-group connection matrices, where the number on each block represents the number of links between the corresponding group pair. The bottom left panel shows the recovered node traits with vector fields representing λ_t^r for the network of dyads having lunch together. The bottom right panel shows two-dimensional λ_t^r of each layer. We conduct 20,000 MCMC simulations and 10,000 burn-in trials. (Color online)

node coordinates, λ_t^1 is larger than λ_t^2 for most layers (plots in the bottom right). A notable outlier is Facebook friend relationships (FB), the only online interaction records included in the study. We can conclude that while the proximity in offline relationships is largely explained by the variation of node positions on the first dimension, online relationships are more likely to be explained by the second-dimensional coordinates.

4.4 Changepoint analysis of congressional voting records

The historical roll call voting record of United States (U.S.) Senate is an attractive dataset to study transitions in network generating processes over time. One of the most important scholarly debates in American politics is whether and when the voting similarity of legislators between and within parties has changed dramatically. We conduct a changepoint analysis, using voting profile similarity network data of U.S. Senate members from 1911 to 2009. The analysis unveils endogenous historical regimes representing distinct constellations of ideological coalitions among U.S. states.

Each network layer represents a roll call voting similarity profile among 48 states in a given session. We build statewise voting agreement intensity networks for 49 sessions by taking the state average of voting agreement scores among Senators. Specifically, we compute voting agreement intensity scores for each senator pair at session t (W_{ijt}) by defining $W_{ijt} = \sum_k D_{ikt}D_{jkt}/(\sum_l D_{lkt} - 1)$ where $D_{ikt} = 1$ if and only if person i voted yea to kth bill and otherwise $D_{ikt} = 0$. Then, we take the state average of W_{ijt} to construct the statewise voting agreement intensity network: $A_{s_i,s_j,t} = \frac{1}{4}\sum_{k\in s_i, l\in s_j} W_{k,l,t}$ where s_i indicates the state i belongs to.

Our changepoint analysis finds three regimes as shown in Figure 6. Matrices in the top panel denote regime-specific voting similarities among 48 states weighted by regime probabilities of each session. States are realigned based on their voting

https://doi.org/10.1017/nws.2017.13 Published online by Cambridge University Press

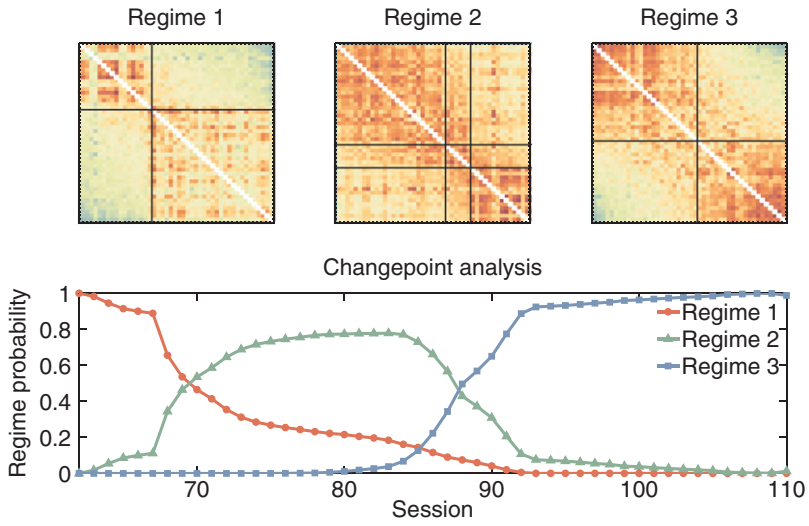

Fig. 6. Endogenous regime changes in the U.S. Senate voting records. For 49 sessions ranging from the 62nd to the 110th Senate sessions, we construct a 48 by 48 network for each layer representing the roll call voting similarity profiles among Senators from 48 states, excluding Alaska and Hawaii. Matrices in the top panel denote regime-specific voting similarities among 48 states weighted by regime probabilities of each session. The red color in the matrices indicates a high level of the voting overlap among the states whereas the blue color indicates a low level of voting overlap. Black lines indicate group boundaries inferred by applying the k-means clustering on two-dimensional regime-specific node trait distributions. State names are omitted due to the space limitation. We conduct 20,000 MCMC simulations and 10,000 burn-in trials for the changepoint detection. (Color online)

similarity for a presentational purpose. As clearly shown by the clustering of high values near the diagonal of the agreement matrices, positive λ_t^r for every dimension of the multilayer networks indicates the presence of assortative associations within coalitions over the entire time period. The plot in the bottom panel displays regime change probabilities. The three regimes can be labeled as the pre-New-Deal regime, the southern realignment regime, and the two-party polarization regime, respectively. The timings of regime changes are highly consistent with historical periods identified by scholars in American politics (McCarty et al., 2006). These changes are strongly associated with the historical trend of southern realignment.

The transition from the first to the second regime coincides with the time when the mean ideological difference between southern Democrats and northern Democrats started to grow (McCarty et al., 2006). The transition from the second to the third regime corresponds to the period when their mean ideological difference reached its maximum value and started to fall. Also, these two breaks respectively correspond to the beginning of the declining phase of the party polarization (i.e. the level of disagreement between Democrats and Republicans) and the period from which the party polarization level started to grow rapidly.

Another striking observation is on the coalitional origin of the extreme party polarization in the recent Congress, illustrated by the final regime probability of each session. While the extreme party polarization is observed in the very recent

https://doi.org/10.1017/nws.2017.13 Published online by Cambridge University Press

sessions (McCarty et al., 2006), its corresponding state coalition structure started to emerge after the 80th session (1947–1949) and became a dominant group structure explaining the majority of roll call voting patterns among the Senators around the 90th session (1967–1969). On the basis of this result, we can argue that there has not been a significant change in the state coalition structure over the last 40 years, but there has been a fixation of the group structure leading to an extreme level of party polarization in the contemporary Congress. While these results provide more direct statistical evidence of changes than Mucha et al. (2010), in which they provided descriptive results without parametric modelings of the timing of changes and changes in the data generating process, do, Figure 6 and Mucha et al. (2010)'s findings are largely in accordance with each other. The largest changes in community assignments of Mucha et al. (2010) took place around the structural breaks depicted in Figure 6.

5 Summary and discussion

Although there have been significant advances in techniques to discover large-scale organizations of complex networks, a practical statistical scheme for the stochastic blockmodeling of multilayer networks and the network times series analysis has been missing. In this article, we discuss a Bayesian framework for the stochastic blockmodeling of multiple networks, and introduce a technique to discover changepoints in multilayer networks with ordered layers.

While empirical examples in the study consist of social and political data, our methodology can be applied to any natural or artificial network with multiple layers. For instance, in order to know the organizational principle of a neural network, one can apply the multilayer approach to analyze a two-layer network of a connectome, in which each layer consists of an excitatory synaptic network and an inhibitory synaptic network, respectively. Also, one can use the method to examine the geographical role differentiation in a traffic network time series dataset where the network on each layer represents traffic flows between geographical regions at a given time point.

Many real-world networks may contain node centrality distributions and other non-group-based properties as their key attributes. Although our study focuses on the stochastic blockmodel recovery, a different specification of the null model recovers substances other than group related traits. For example, setting $\mathbf{P_t} = \mathbf{0}$ allows the first dimension of the recovered trait space to represent the node centrality. If the changepoint detection algorithm is added, the model detects changes in node centrality ranking over time. Other representations of networks, such as graph Laplacians and modified modularity matrices (Expert et al., 2011), can be implemented as $\mathbf{B_t}$ in the Bayesian rank reduction process.

Technical development of the multilayer modeling framework is a promising area for a further scrutiny. An immediate improvement of the present algorithm can be made by generalizing it for the directed network trait recovery. Also, instead of fixing the number of changepoints, an algorithm can be developed to endogenously determine the number of changepoints. Not limited to these examples, establishing statistical criteria for understanding interconnectivity between layers would provide a useful means to understand hyper-dimensional network datasets.

https://doi.org/10.1017/nws.2017.13 Published online by Cambridge University Press

Acknowledgments

Yunkyu Sohn was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2012M3C4A7033342). Jong Hee Park was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2013S1A3A2053683). We are grateful to Matteo Magnani and Keith Poole for providing the empirical datasets used in the study. We thank John Ahlquist, Yong-Yeol Ahn, James Fowler, Kosuke Imai, Gary Jacobson, Koji Kagotani, Thad Kousser, Sang Hoon Lee, Jacob Montgomery, Sue Moon, Magaret Roberts, Jon Rogowski, Arturas Rozenas, and Michael Ward for helpful comments on an earlier draft.

Supplementary materials

For supplementary material for this article, please visit https://doi.org/10.1017/nws.2017.13

References

Ahn, Y.-Y., Bagrow, J. P., & Lehmann, S. (2010). Link communities reveal multiscale complexity in networks. *Nature*, **466**(7307), 761–764.

Airoldi, E. M., Blei, D. M., Fienberg, S. E., & Xing, E. P. (2009). Mixed membership stochastic blockmodels. *Advances in Neural Information Processing Systems*, **9**, 33–40.

Akoglu, L., Tong, H., & Koutra, D. (2015). Graph-based anomaly detection and description: A survey. *Data Mining and Knowledge Discovery*, **29**(3), 626–688.

Araujo, M., Papadimitriou, S., GŁünnemann, S., Faloutsos, C., Basu, P., Swami, A., . . . Koutra, D. (2014). Com2: Fast automatic discovery of temporal ('comet') communities. In V. S. Tseng, T. B. Ho, Z.-H. Zhou, A. L. P. Chen, & H.-Y. Kao (Eds.), *Advances in Knowledge Discovery and Data Mining: 18th Pacific-Asia Conference, PAKDD 2014, Tainan, Taiwan, May 13–16, 2014. Proceedings, Part II* (pp. 271–283). New York, NY: Springer International Publishing.

Bickel, P. J., & Chen, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proceedings of the National Academy of Sciences*, **106**(50), 21068–21073.

Boccaletti, S., Bianconi, G., Criado, R., Del Genio, C., Gómez-Gardeñes, J., Romance, M., . . . Zanin, M. (2014). The structure and dynamics of multilayer networks. *Physics Reports*, **544**(1), 1–122.

Borgatti, S. P., & Everett, M. G. (2000). Models of core/periphery structures. *Social Networks*, **21**(4), 375–395.

Chaudhuri, K., Chung, F., & Tsiatas, A. (2012). Spectral clustering of graphs with general degrees in the extended planted partition model. *Journal of Machine Learning Research*, **23**(35), 1–23.

Chib, S. (1996). Calculating posterior distributions and modal estimates in markov mixture models. *Journal of Econometrics*, **75**, 79–98.

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

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

Expert, P., Evans, T. S., Blondel, V. D., & Lambiotte, R. (2011). Uncovering space-independent communities in spatial networks. *Proceedings of the National Academy of Sciences*, **108**(19), 7663–7668.

Faust, K. (1988). Comparison of methods for positional analysis: Structural and general equivalences. *Social Networks*, **10**(4), 313–341.
Fishkind, D. E., Sussman, D. L., Tang, M., Vogelstein, J. T., & Priebe, C. E. (2013). Consistent adjacency-spectral partitioning for the stochastic block model when the model parameters are unknown. *SIAM Journal on Matrix Analysis and Applications*, **34**(1), 23–39.
Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, **486**(3), 75–174.
Fortunato, S., & Barthelemy, M. (2007). Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, **104**(1), 36–41.
Gelfand, A. E., & Smith, A. F. M. (1990, June). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**(410), 398–409.
Girvan, M., & Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, **99**(12), 7821–7826.
Gopalan, P. K., & Blei, D. M. (2013). Efficient discovery of overlapping communities in massive networks. *Proceedings of the National Academy of Sciences*, **110**(36), 14534–14539.
Guimera, R., & Amaral, L. A. N. (2005). Functional cartography of complex metabolic networks. *Nature*, **433**(7028), 895–900.
Han, Q., Xu, K., & Airoldi, E. (2015). Consistent estimation of dynamic and multi-layer block models. *Proceedings of the 32nd International Conference on Machine Learning*, **37**, 1511–1520.
Handcock, M. S., Raftery, A. E., & Tantrum, J. M. (2007). Model-based clustering for social networks. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **170**(2), 301–354.
Hoff, P. D. (2005). Bilinear mixed-effects models for dyadic data. *Journal of the American Statistical Association*, **100**(469), 286–295.
Hoff, P. D. (2008). Modeling homophily and stochastic equivalence in symmetric relational data. In J. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in neural information processing systems* (vol. 20, pp. 657–664). Cambridge: Cambridge University Press.
Hoff, P. D. (2009). Simulation of the matrix bingham-von mises-fisher distribution, with applications to multivariate and relational data. *Journal of Computational and Graphical Statistics*, **18**(2), 438–456.
Hoff, P. D. (2011). Hierarchical multilinear models for multiway data. *Computational Statistics & Data Analysis*, **55**, 530–543.
Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, **97**(460), 1090–1098.
Hofman, J. M., & Wiggins, C. H. (2008). Bayesian approach to network modularity. *Physical Review Letters,* **100**(25), 258701.
Holland, P. W., Laskey, K. B., & Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks*, **5**(2), 109–137.
Holme, P., & Saramäki, J. (2012). Temporal networks. *Physics Reports*, **519**(3), 97–125.
Karrer, B., & Newman, M. E. (2011). Stochastic blockmodels and community structure in networks. *Physical Review E*, **83**(1), 016107.
Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., & Porter, M. A. (2014). Multilayer networks. *Journal of Complex Networks*, **2**(3), 203–271.
Kolar, M., Song, L., Ahmed, A., & Xing, E. P. (2010). Estimating time-varying networks. *The Annals of Applied Statistics*, **4**(1), 94–123.
Kolda, T. G., & Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review*, **51**(3), 455–500.
Krzakala, F., Moore, C., Mossel, E., Neeman, J., Sly, A., & Zdeborová, L. (2013). Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences*, **110**(52), 20935–20940.
Lancichinetti, A., Radicchi, F., Ramasco, J. J., & Fortunato, S. (2011). Finding statistically significant communities in networks. *PloS One*, **6**(4), e18961.
Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, **401**(6755), 788–791.
Lei, J., & Rinaldo, A. (2014). Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, **43**(1), 215–237.

Liu, J. S. (1994). The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, **89**(427), 958–966.

Magnani, M., Micenkova, B., & Rossi, L. (2013). Combinatorial analysis of multiple networks. *preprint, arXiv:1303.4986*.

McCarty, N., Poole, K. T., & Rosenthal, H. (2006). *Polarized America: The dance of ideology and unequal riches*, Vol. 5. Cambridge, MA: MIT Press.

Meila, M. (2003). Comparing clusterings by the variation of information. In B. Schoelkopf & M. K. Warmuth (Eds.), *Learning Theory and Kernel Machines: 16th Annual Conference on Computational Learning Theory and 7th Kernel workshop, CLOT/Kernel 2003, Washington, DC, USA, August 24–27, 2003, Proceedings* (vol. 2777, pp. 173–187). Berlin: Springer.

Moody, J., & White, D. R. (2003). Structural cohesion and embeddedness: A hierarchical concept of social groups. *American Sociological Review*, **68**(1), 103–127.

Mucha, P. J., Richardson, T., Macon, K., Porter, M. A., & Onnela, J.-P. (2010). Community structure in time-dependent, multiscale, and multiplex networks. *Science*, **328**(5980), 876–878.

Nadakuditi, R. R., & Newman, M. E. (2013). Spectra of random graphs with arbitrary expected degrees. *Physical Review E*, **87**(1), 012803.

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

Newman, M. E. (2013, Oct). Spectral methods for community detection and graph partitioning. *Physical Review E,* **88**, 042822. doi: 10.1103/Phys-RevE.88.042822

Newman, M. E., & Leicht, E. A. (2007). Mixture models and exploratory analysis in networks. *Proceedings of the National Academy of Sciences*, **104**(23), 9564–9569.

Ng, A. Y., Jordan, M. I., & Weiss, Y. (2002). On spectral clustering: Analysis and an algorithm. *Advances in Neural Information Processing Systems*, **2**, 849–856.

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

Palla, G., Barabási, A.-L., & Vicsek, T. (2007). Quantifying social group evolution. *Nature*, **446**(7136), 664–667.

Paul, S., & Chen, Y. (2016). Consistent community detection in multi-relational data through restricted multi-layer stochastic blockmodel. *Electronic Journal of Statistics*, **10**(2), 3807–3870.

Peel, L., & Clauset, A. (2015). Detecting change points in the large-scale structure of evolving networks. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence* (pp. 2914–2920). Austin, TX: AAAI Press.

Peixoto, T. P. (2013). Eigenvalue spectra of modular networks. *Physical Review Letters*, **111**(9), 098701.

Peixoto, T. P. (2015). Inferring the mesoscale structure of layered, edge-valued, and time-varying networks. *Physical Review E*, **92**(4), 042807.

Riolo, M. A., & Newman, M. E. (2014). First-principles multiway spectral partitioning of graphs. *Journal of Complex Networks*, **2**(2), 121–140.

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

Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, **105**(4), 1118–1123.

Smith, A. F. M., & Roberts, G. O. (1993). Bayesian computation via the gibbs sampler and related markov chain monte carlo methods. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, **55**(1), 3–23.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, **64**(4), 583–639.

Stanley, N., Shai, S., Taylor, D., & Mucha, P. (2016). Clustering network layers with the strata multilayer stochastic block model. *IEEE Transactions on Network Science and Engineering*, **3**(2), 95–105.

Sussman, D. L., Tang, M., Fishkind, D. E., & Priebe, C. E. (2012). A consistent adjacency spectral embedding for stochastic blockmodel graphs. *Journal of the American Statistical Association*, **107**(499), 1119–1128.

Ward, M. D., Ahlquist, J. S., & Rozenas, A. (2013). Gravity's rainbow: A dynamic latent space model for the world trade network. *Network Science*, **1**(1), 95–118.

https://doi.org/10.1017/nws.2017.13 Published online by Cambridge University Press