Statistica Sinica Preprint No: SS-2021-0182	
Title	Joint Modeling of Change-Point Identification and Dependent Dynamic Community Detection
Manuscript ID	SS-2021-0182
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202021.0182
Complete List of Authors	Diqing Li, Yubai Yuan, Xinsheng Zhang and Annie Qu
Corresponding Authors	Annie Qu
E-mails	aqu2@uci.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

Joint modeling of change-point identification and dependent dynamic community detection

Diqing Li[1], Yubai Yuan[2], Xinsheng Zhang[3], Annie Qu[4]

[1] *Zhejiang Gongshang University* [2] *The Pennsylvania State University*

[3] *Fudan University* [4] *University of California Irvine*

Abstract: The field of dynamic network analysis has recently seen a surge of interest in community detection and evolution. However, existing methods for dynamic community detection do not consider dependencies between edges, which could lead to a loss of information when detecting community structures. In this study, we investigate the problem of identifying a change-point with abrupt changes in the community structure of a network. To do so, we propose an approximate likelihood approach for the change-point estimator and for identifying node membership that integrates marginal information and dependencies of network connectivities. We propose an expectation-maximization-type algorithm that maximizes the approximate likelihood jointly over change-point and community membership evolution. From a theoretical viewpoint, we establish estimation consistency under the regularity condition, and show that the proposed estimators achieve a higher convergence rate than those of their marginal likelihood counterparts, which do not incorporate dependencies between edges. We demonstrate the validity of the proposed method by applying it to the ADHD-200 data set to detect brain functional community changes over time.

Key words and phrases: Change-point detection, Community detection, Dynamic network, Stochastic block model.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

1. Introduction

Network data analysis has become an important tool for studying relationships and associations among subjects. In this study, we develop community detection for dynamic network data. Traditional network analysis assumes that network connectivities are independent and identically distributed (i.i.d.), and that networks are static over time. However, in practice, these assumptions fail when dynamic changes of sequential observed network data occur over time, such as in social networks, political networks, trading networks, brain networks, biological networks, among others.

Most existing dynamic networks do not use the latent community structure, mainly using the Gaussian graphic model to analyze the conditional correlation between variables. To deal with a time-varying graphical structure, typical assumptions for the covariance matrix require that the precision matrices either change smoothly or are piece-wise constant. For example, Kolar et al. (2010) study the structure recovery of time-varying Ising graphical models, assuming a smooth change of an underlying parameter. Gibberd and Nelson (2017) propose a group-fused graphical lasso method for estimating piece-wise constant Gaussian graphical models. Yang and Peng (2018) propose a local group graphical lasso estimation, assuming that the graph topology changes gradually over time. The aforementioned approaches focus on detecting changes in the precision matrix to explore the conditional correlation among nodes, which is not applicable for detecting community structures.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

Compared with static networks, dynamic community detection is much more challenging, because it involves both node classification and dynamic changes in the node grouping over time. Existing works analyze community extraction and community dynamics using two separate steps: first, a static analysis is applied to snapshots of the networks at given time points; then, community changes over time are detected (Toyoda and Kitsuregawa, 2003; Mei and Zhai, 2005; Palla et al., 2007). However, this strategy assumes that the networks are independent of each other at each discrete time point, thus failing to capture the community dynamics for the entire period, possibly making community detection unstable. In addition, the label-switching problem can arise between two successive time points (Matias and Miele, 2017).

The existing literature on dynamic community detection is based mainly on stochastic block models (SBMs), or variants of the SBM. For example, Xing et al. (2010), Yang et al. (2011) ,Xu and Hero (2014) use a transition matrix to describe the changes for nodes under Bayesian frameworks. Matias and Miele (2017) use an SBM for the static part, with Markov chains for the changes of node groups over time. As an alternative, latent space models can be applied to capture the probability of the dynamic connectivities of two nodes using the closeness of the corresponding latent points (Sarkar and Moore, 2006; Heaukulani and Ghahramani, 2013; Sewell and Chen, 2017).

Although the aforementioned methods incorporate dynamic changes of membership, they assume that the network communities follow a smooth change over time, and that

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

the underlying distribution of the community structure does not change over time, such as the stationary Markov chain. However, abrupt changes can occur in communities, owing to external events or changes in the locations or interests of individuals in the network. For example, a political network community may suffer a sudden and abrupt change after an election, individuals may join the social network of a new community because of a change in their geographical location, or a stock trading network may experience high volatility during an economic crisis or unforeseen disaster, such as a pandemic. Therefore, we cannot apply the dynamic network methods designed for smooth changes, requiring that we construct an appropriate model for an abrupt change that causes the underlying distribution of a community structure to change at a certain time point.

Few existing studies examine abrupt changes in dynamic community detection. Marangoni-Simonsen and Xie (2015) propose detecting the emergence of a community in large networks from sequential observations using Erdős–Renyi random graphs. Wang et al. (2017b) detect change-points in dynamic networks under the snapshot model at a given time. Wang et al. (2017a) study hierarchical change-point detection to differentiate between intra-community and inter-community evolution, assuming that the community structure is known in advance. In general, these approaches compare two sequential networks to determine the possible location of a change-point using the differences between the two networks. Dubey et al. (2020) propose a test statistic for inferring the presence of a change-point within a sequence of network distributions, but that cannot recover communities si-

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

multaneously. However, the estimation of a change-point and the community structure are joint processes, owing to their mutual influence on each other. In addition, the estimation of a change-point conditional on a given community structure could be biased if the given community structure is incorrect. For joint estimation, Bhattacharjee et al. (2020) consider an unknown community estimation with a single change-point using a spectral-based method, but their key assumption for a consistent estimation is difficult to verify. Wang et al. (2018) propose a weighted network aggregating method that detects change-points in sparse network sequences. Zhao et al. (2019) establish a nonparametric approach for detecting a change in the underlying network structure by using neighborhood smoothing.

The main limitations of the aforementioned methods are that they assume that connectivities are conditionally independent, given the membership of nodes, resulting in a loss of high-order network information on community structure change. In addition, this assumption is usually infeasible, because most networks are relational interdependent or interact with each other, in practice. Therefore, dependent edges should be incorporated. For example, friendships within social networks and functional connectivities in brain networks are highly correlated. Trust transitivity in a social network shows the dependency of trustworthiness among unknown participants (Liu et al., 2011). In addition, the correlations of brain network connectivity between patients with Alzheimer's disease and healthy controls could be quite different (Yapeng et al., 2013). To incorporate dependencies between edges in static networks, Cheng et al. (2014) propose a conditional distribution model of

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

binary network data, and incorporate covariate information into the Ising model. Park and Lee (2014) develope a clustering method that incorporates group dependence by using a geometric structure, and Yuan and Qu (2021) propose a truncated Bahadur representation to incorporate the underlying dependency structure between the connectivities. However, incorporating edge dependency has not been considered under dynamic network settings.

In this paper, we propose a simultaneous change-point identification and community detection method that incorporates abrupt changes and dependent connections. In contrast to static networks, we allow both the community structure and the dependence within communities to change. The change-point estimator and the corresponding estimated community membership are obtained by maximizing an approximate likelihood, providing a tighter lower bound to the true likelihood compared with that of the independent SBM likelihood approach. Specifically, the proposed approximate likelihood captures changes to the community structure from both marginal mean information and correlation information of the connectivities. In contrast to Yuan and Qu (2021), we allow the correlation coefficient between edges to change over time, which increases the computational complexity significantly, owing to the involvement of a change-point estimation. To avoid an intractable computation, we apply an expectation-maximization (EM)-type method for the proposed likelihood to solve the optimization problem. Under regularity conditions, we show that the estimators of both the community memberships and the change-points are consistent. Empirical studies confirm our theoretical findings in that the proposed method

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

outperforms existing approaches.

The main advantages and contributions of the proposed method can be summarized as follows. We consider community detection incorporating abrupt changes under the framework of a dependent dynamic SBM. In addition to the marginal mean difference between communities, the proposed method considers within-community dependency between connectivities during the dynamic evolution of the community structure. In practice, the concordance among within-community edges is an important intrinsic characteristic of communities. Consequently, incorporating within-community dependency could enhance the discriminative power of identifying the community memberships at change-points. This is especially helpful when the marginal mean information is not informative. Because of the mutual influence between the estimations of the change-point and the community structures, we propose an algorithm that jointly estimates the change-point and identifies the community structures. Establishing the consistency property of the change-point and community membership estimations simultaneously is nontrivial, particularly because we incorporate conditional dependencies between edges, whereas existing methods assume conditional independence. The remainder of this paper is organized as follows. Section 2 introduces the notation and presents the proposed method and the implementation algorithm. Section 3 illustrates the theoretical properties of the proposed method. Section 4 discusses our simulation studies, and in Section 5, we apply our method to real data on ADHD brain networks. The last section concludes the paper.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

2. Methodology

2.1 Formulation

In this section, we consider the problem of community detection with an abrupt change, where the members of some communities change at certain time points. Specifically, we consider a dynamic change-point under the SBM framework.

Let $\boldsymbol{Y} = \left\{(Y_{ij}^t)_{n\times n}\right\}_{t=1}^T$ be T symmetric and unweighted sample networks, where $\{Y_{ij}^t\}$ is one if there is an edge between nodes i and j at time point t, and is zero otherwise. Here, n is the number of nodes that belong to one of K communities, and the main diagonal $\{Y_{ii}^t\}_{i=1}^n$ is fixed to be zero. Suppose there exists an unknown change-point location $\tau^\star$, where some nodes change their community memberships after $\tau^\star$. In the following, we use a superscript d to distinguish variables before and after a change-point, where $d = 1$ before the change-point and $d = 2$ after change-point. We set $\boldsymbol{z}^d = (z_1^d, \ldots, z_n^d)^\top$ as the memberships for node $i = 1, \ldots, n$, where $z_i^d \in \{1, 2, \ldots, K\}$, such that

$$\boldsymbol{z}^1 = \boldsymbol{z}(1) = \ldots = \boldsymbol{z}(\tau^\star) \neq \boldsymbol{z}(\tau^\star + 1) = \ldots = \boldsymbol{z}(T) = \boldsymbol{z}^2, \tag{2.1}$$

where $\boldsymbol{z}(t)$ is the membership vector at time point t. We denote the membership assignment matrix $\boldsymbol{Z}^d = \left\{(Z_{ik}^d)_{n\times K}\right\} \in \{0,1\}^{n\times K}$, and $Z_{ik}^d = \mathbb{I}\{z_i^d = k\}$. Throughout this paper, we assume that the true membership $\boldsymbol{Z}^{\star d}$ is fixed. Given the membership of nodes, the observed edges between two nodes $\left\{(Y_{ij}^t)_{n\times n}\right\}_{t=1}^T$ follow a Bernoulli distribution with

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

parameter $\boldsymbol{\pi} = \{(\pi_{q,l})\}$ such that

$$P(Y_{ij}^t | z_i^d = q, z_j^d = l) \sim \textit{Bern}(\pi_{ql}), \quad \text{for } i, j = 1, \ldots, n, q, l = 1, \ldots, K, \tag{2.2}$$

where π_{ql} is the probability of nodes from communities q and l being connected. Additionally, we allow within-community connectivity dependency, and denote the correlation between edges $\left(Y_{i_1 j_1}^t, Y_{i_2 j_2}^t\right)$ within a community as $\text{corr}\left(Y_{i_1 j_1}^t, Y_{i_2 j_2}^t\right) = \rho_{i_1, j_1, i_2, j_2}^t$.

In the following, we propose identifying a change-point using global network information, through the probability matrix $\boldsymbol{\pi}$. The goal of this study is to identify $\tau^\star$ and recover the underlying community structures simultaneously.

2.2 Change-point identification and community detection

Without loss of generality, we assume that the change-point occurs at one time τ, although our method is not restricted to one time-point only. We define the parameter space as $\Theta = (\boldsymbol{\pi}, \tau, \boldsymbol{z}^1, \boldsymbol{z}^2)$. Let $\tau \in \mathbb{T} \equiv [t_0 T, t_0 T + 1, \ldots, t_1 T]$, where $0 < t_0 < t_1 < 1$ and $t_0 T$ and $t_1 T$ are integers. Let $\boldsymbol{\pi} \in \Pi \subset [0,1]^{K^2}$, where K is the number of communities, and $\mathbb{T}$ and Π are parameter spaces for τ and $\boldsymbol{\pi}$, respectively. The joint log-likelihood function $\log L(Y, \boldsymbol{\pi}, \tau, \boldsymbol{z}^1, \boldsymbol{z}^2)$ given a change-point τ can be decomposed into the summation of edge-wise terms, based on the conditional independence assumption

$$\begin{aligned} \log L(Y | \boldsymbol{\pi}, \tau, \boldsymbol{z}^1, \boldsymbol{z}^2) &= \sum_{t=1}^{\tau} \sum_{i<j} \left\{ Y_{ij}^t \log \pi_{z_i^1 z_j^1} + (1 - Y_{ij}^t) \log\left(1 - \pi_{z_i^1 z_j^1}\right) \right\} \\ &+ \sum_{t=\tau+1}^{T} \sum_{i<j} \left\{ Y_{ij}^t \log \pi_{z_i^2 z_j^2} + (1 - Y_{ij}^t) \log\left(1 - \pi_{z_i^2 z_j^2}\right) \right\}. \end{aligned} \tag{2.3}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

Then, the change-point and membership can be estimated simultaneously as a one-step maximizer, such that

$$(\widehat{\tau}, \widehat{\boldsymbol{z}}^1, \widehat{\boldsymbol{z}}^2) = \underset{\tau \in \mathbb{T} | \boldsymbol{z}^1, \boldsymbol{z}^2 \in \{1,\ldots,K\}^n, \boldsymbol{\pi} \in \Pi}{\operatorname{argmax}} \log L(Y | \boldsymbol{\pi}, \tau, \boldsymbol{z}^1, \boldsymbol{z}^2). \tag{2.4}$$

The community detection objective function in (2.3) assumes that connectivities are conditional independent, given the membership of nodes. Therefore it allows one to change the community structure using only the marginal information, in that the average connectivity rates within communities are different before and after the membership change. However, in most community detection problems, it is common for edges within communities to be more correlated, such as friendships in social networks. Therefore, the conditional independence assumption among connectivities is restrictive and practically infeasible, because it could lead to significant information loss related to the community structure change.

In this section, we propose an approximate likelihood function to incorporate within-community correlation, following Yuan and Qu (2021), thus improving the accuracy and efficiency of identifying a community structure and detecting a change-point. In addition to the marginal mean information of the edges, within-community dependency contains additional information about the membership of nodes. This is especially effective when the marginal mean is not informative in differentiating the connectivity between and within communities.

In contrast to Yuan and Qu (2021), we assume that the correlation coefficient be-

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

tween edges can also change over time, because the nodes' membership changes over time. Here, we use a homogeneous correlation structure, such that all pairwise correlations from each community are assumed to be the average within-community correlation. The rationale for this simplification is as follows. First, the pairwise correlation parameter $\rho^t_{i_1 i_2 j_1 j_2} = \text{corr}\left(Y^t_{i_1 j_1}, Y^t_{i_2 j_2}\right)$ is a nuisance correlation parameter to enhance clustering and identification. Second, Theorem 3.3 shows that the density of the pairwise correlations between within-community edges affects the clustering performance more than the intensity of the correlation does. Third, in practice, we use the sample correlation coefficient to estimate $\rho^t_{i_1 i_2 j_1 j_2}$. However, if the observed binary edges are all zeros or ones, then the corresponding sample correlation coefficient does not exist, owing to the zero variance. For example, if we assume that the change-point is $T/2$, it is possible that the sample correlation coefficients of certain pairs of edges are not estimable when only $T/2$ samples are used for the calculation. More critically, we have nearly $n^4/4$ correlation coefficients to estimate, which is infeasible. Therefore, we assume a homogeneous correlation structure such that all pairwise correlations in each community are assumed to be the average within-community correlation. Specifically, given that nodes i_1, j_1, i_2, and j_2 are in the same community q, $\text{corr}\left(Y^t_{i_1 j_1}, Y^t_{i_2 j_2}\right) = \rho^d_q$, where $d = 1$ if $t \leq \tau^\star$ and $d = 2$ if $t > \tau^\star$.

Inspired by the Bahadur representation (Bahadur, 1961), we propose an approximate likelihood function that combines information about node' membership from both marginal

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

mean information and within-community dependency, as follows

$$\log L(Y|\boldsymbol{\pi},\tau,\boldsymbol{\varrho},\boldsymbol{z}^1,\boldsymbol{z}^2) = \log L^1_{ind}(\tau) + \log L^1_{cor}(\tau) + \log L^2_{ind}(\tau) + \log L^2_{cor}(\tau), \quad (2.5)$$

where

$$\log L^d_{ind}(\tau) = \sum_{t\in\mathbb{T}_d}\sum_{i<j}\left\{Y^t_{ij}\log\pi_{z^d_i z^d_j} + (1-Y^t_{ij})\log\left(1-\pi_{z^d_i z^d_j}\right)\right\},$$

$$\log L^d_{cor}(\tau) = \sum_{t\in\mathbb{T}_d}\log\Big[1+\sum_{k=1}^{K}\frac{\rho^d_k}{2}\sum_{\substack{i<j;u<v\\(i,j)\neq(u,v)}}^{n} Z^d_{ik}Z^d_{jk}Z^d_{uk}Z^d_{vk}\widehat{Y}^t_{ij}\widehat{Y}^t_{uv}\Big],$$

for $d = 1, 2$. We denote the time-point sets before and after the change-point as $\mathbb{T}_1 = \{1,\ldots,\tau\}$ and $\mathbb{T}_2 = \{\tau+1,\ldots,T\}$, respectively, and $\widehat{Y}^t_{ij}$ is a normalized binary variable such that

$$\widehat{Y}^t_{ij} = \frac{Y^t_{ij} - E(Y^t_{ij})}{\sqrt{E(Y^t_{ij})(1-E(Y^t_{ij}))}}.$$

Specifically, the likelihood of the marginal mean is kept in $\log L^1_{ind}(\tau)+\log L^2_{ind}(\tau)$, which is consistent with the conditional independence model 2.3. In addition, we retain the second-order dependency information among edges within communities, and ignore the high-order dependency for computational efficiency. Intuitively, the second-order interaction term

$$\sum_{\substack{i<j;u<v\\(i,j)\neq(u,v)}}^{N} Z_{ik}Z_{jk}Z_{uk}Z_{vk}\widehat{Y}^t_{ij}\widehat{Y}^t_{uv}$$

measures the within-community concordance among edges, and therefore incorporates within-community dependency by encouraging clustering among nodes with associated edges that are correlated.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

Similar to (2.4), we can estimate $(\tau, \boldsymbol{z}^1, \boldsymbol{z}^2)$ by

$$(\widehat{\tau}, \widehat{\boldsymbol{z}}^1, \widehat{\boldsymbol{z}}^2) = \underset{\tau\in\mathbb{T},\boldsymbol{z}^1,\boldsymbol{z}^2\in\{1,\ldots,K\}^n,\boldsymbol{\pi}\in\Pi,\boldsymbol{\varrho}\in\mathcal{P}}{\operatorname{argmax}} \log L(Y|\boldsymbol{\pi}, \tau, \boldsymbol{\varrho}, \boldsymbol{z}^1, \boldsymbol{z}^2), \tag{2.6}$$

where $\mathcal{P} \subset [0,1]^{2K}$ is the parameter space for $\boldsymbol{\varrho} = \{\rho_1^1, \ldots, \rho_K^1, \rho_1^2, \ldots, \rho_K^2\}^\top$. To jointly maximize the proposed approximate likelihood, we use the profiling method such that we maximize $\log L(Y|\boldsymbol{\pi}, \tau, \boldsymbol{\varrho}, \boldsymbol{z}^1, \boldsymbol{z}^2)$ over the parameters $\boldsymbol{\pi}, \boldsymbol{\varrho}, \boldsymbol{z}^1, and \boldsymbol{z}^2$, given a fixed change-point τ; that is,

$$\log L(Y|\tau) = \underset{\boldsymbol{z}^1,\boldsymbol{z}^2\in\{1,\ldots,K\}^n,\boldsymbol{\pi}\in\Pi,\boldsymbol{\varrho}\in\mathcal{P}}{\operatorname{argmax}} \log L(Y|\boldsymbol{\pi}, \boldsymbol{\varrho}, \boldsymbol{z}^1, \boldsymbol{z}^2, \tau).$$

Then, we search τ from $[t_0 T, t_1 T]$ to maximize $\log L(Y|\tau)$ using an EM-type method. Specifically, although we assume that the underlying true node memberships $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ are fixed parameters, in the optimization procedure, we adopt the idea of the EM algorithm to treat $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ as missing variables that follows a multinomial distribution in the E step of algorithm 1. To obtain $\log L(Y|\tau)$ for any fixed τ, directly applying the EM method is computationally challenging, because the conditional distribution $P(\boldsymbol{z}^1, \boldsymbol{z}^2|Y)$ becomes intractable in the expectation step. Therefore, we modify the EM algorithm following variational methods that approximate the likelihood $P(\boldsymbol{z}^1, \boldsymbol{z}^2|Y)$ using a complete factorized distribution, $R(\boldsymbol{z}^1, \boldsymbol{z}^2, \boldsymbol{\mu}^1, \boldsymbol{\mu}^2) = \prod_{i=1}^{n} h(Z_i^1; \mu_i^1) h(Z_i^2; \mu_i^2)$, where $h(\cdot)$ denotes a multinomial distribution, $\boldsymbol{\mu}^d = (\mu_1^d, \cdots, \mu_N^d)$, and $\mu_i^d = (\mu_{i1}^d, \cdots, \mu_{iK}^d)$ is a probability vector such that $\sum_{q=1}^{K} \mu_{iq}^d = 1$, for $d = 1, 2$. Note that only in the optimization process are $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ treated as random. In our theoretical analysis, we still consider $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ to be fixed parameters.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

The key part of the proposed algorithm is to update the memberships of the nodes using a Bayes factor constructed using the proposed approximate likelihood function. Specifically, suppose the memberships of other nodes $\boldsymbol{Z}_{-i}$ are known. Then, we predict the membership of node i based on the following Bayes factor:

$$\frac{L^d(Y|\tau, \boldsymbol{Z}^d_{-i}, Z^d_{iq}=1)}{L^d(Y|\tau, \boldsymbol{Z}^d_{-i}, Z^d_{ik}=1)} = \frac{L^d_{ind}(Y|\tau, \boldsymbol{Z}^d_{-i}, Z^d_{iq}=1)L^d_{cor}(Y|\tau, \boldsymbol{Z}^d_{-i}, Z^d_{iq}=1)}{L^d_{ind}(Y|\tau, \boldsymbol{Z}^d_{-i}, Z^d_{ik}=1)L^d_{cor}(Y|\tau, \boldsymbol{Z}^d_{-i}, Z^d_{ik}=1)}. \tag{2.7}$$

If the above Bayes factor (2.7) > 1, then the probability of node i being in community q is larger than that of it being in community k. Here, the Bayes factor can be decomposed into ratios of the marginal and the correlation parts. When the marginal information is not informative in differentiating between two communities, the marginal part ratio is close to one. However if the correlation ratio is informative, the Bayes factor in (2.7) is still differentiable, thus enhancing the community detection. In addition, the correlation ratio serves as a correction to reduce the estimation bias.

In the maximization step, the point estimators for the model parameters are obtained by maximizing $\log L(Y, \boldsymbol{\pi}, \tau, \boldsymbol{\varrho}, \boldsymbol{\mu}^1, \boldsymbol{\mu}^2)$, and we obtain

$$\widehat{\pi}_{ql}(\tau) = \frac{\sum_{t=1}^{\tau}\sum_{i\neq j}^{n}\widehat{\mu}^1_{iq}\widehat{\mu}^1_{jl}Y^t_{ij} + \sum_{t=\tau+1}^{T}\sum_{i\neq j}^{n}\widehat{\mu}^2_{iq}\widehat{\mu}^2_{jl}Y^t_{ij}}{\sum_{t=1}^{\tau}\sum_{i\neq j}^{n}\widehat{\mu}^1_{iq}\widehat{\mu}^1_{jl} + \sum_{t=\tau+1}^{T}\sum_{i\neq j}^{n}\widehat{\mu}^2_{iq}\widehat{\mu}^2_{jl}}. \tag{2.8}$$

For each node, the membership $\boldsymbol{z}^d$ is determined by identifying the largest probability of

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

$\boldsymbol{z}_i^d(\tau)$ such that

$$\widehat{z}_{iq}^d(\tau) = \begin{cases} 1, & \text{if } q = \text{argmax}_{q'}\, \widehat{\mu}_{iq'}^d(\tau) \\ 0, & \text{otherwise.} \end{cases} \tag{2.9}$$

Algorithm 1 implements the change-point identification and community detection jointly.

Algorithm 1 EM type algorithm

Input: samples $\boldsymbol{Y}$, searching space $\mathbb{T}$, estimation accuracy ϵ.
Initialization: set the iteration counter $s = 0$, and initial parameters $\boldsymbol{\pi}^{(0)} \in \mathbb{R}^{K\times K}$, $\boldsymbol{\mu}^{1(0)} \in \mathbb{R}^{n\times K}$, $\boldsymbol{\mu}^{2(0)} \in \mathbb{R}^{n\times K}$, $\boldsymbol{Z}^{1(0)} \in \mathbb{R}^{n\times K}$, $\boldsymbol{Z}^{2(0)} \in \mathbb{R}^{n\times K}$.
for τ in $[t_1, t_2]$ **do**
  repeat
    M step:
    Obtain $\pi_{ql}^{(s)}$ through (2.8).
    E step:
    Obtain $\mu_{iq}^{1(s)}$ and $\mu_{iq}^{2(s)}$ through:

$$\mu_{iq}^{d(s)} = \frac{\mu_{iq}^{d(s-1)} L(Y|\tau, \boldsymbol{\mu}_{-i}^{d(s-1)}, Z_{iq}^d = 1)}{\sum_{k=1}^K \mu_{ik}^{d(s-1)} L(Y|\tau, \boldsymbol{\mu}_{-i}^{d(s-1)}, Z_{ik}^d = 1)}.$$

    Update $\rho_q^{d(s)}$ through the method of moments estimator.
    Update the iteration counter s=s+1.
  until $\boldsymbol{\mu}^{(s)}$ satisfies: $\max_i (|\boldsymbol{\mu}_i^{1(s)} - \boldsymbol{\mu}_i^{1(s-1)}|_1 + |\boldsymbol{\mu}_i^{2(s)} - \boldsymbol{\mu}_i^{2(s-1)}|_1) < \epsilon$.
  Calculate $z_i(\tau) = \text{argmax}_k\{\mu_{i1}^{d(s)}, \ldots, \mu_{iK}^{d(s)}\}$.
  Calculate $\mathcal{L}(\tau) = \log L(Y, \boldsymbol{\pi}^{(s)}, \tau, \boldsymbol{\varrho}^{(s)}, \boldsymbol{z}^{1(s)}, \boldsymbol{z}^{2(s)})$ as in (2.5).
end for
Output: $\widehat{\tau} = \text{argmax}_{\tau\in[t_1,t_2]}\, \mathcal{L}(\tau)$ and $z_i^d = z_i^{d(s)}$.

Remark 1. In order to ensure the nonnegativity of the correlations between edges, when we calculate the correlation coefficients in the E step of Algorithm 1, if the moment estimator of a correlation is negative, we set the correlation to be zero. This also ensures that the likelihood function is valid, because the log part is required to be positive.

The computational complexity of the proposed EM-type algorithm at each iteration

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

is $O(Tn^4K^2)$, which is determined mainly by updating the node memberships in the E step. Specifically, the complexity for calculating the Bayes factor for each single observed network is $O(n^4K^2)$, because it contains the correlation factor using the Bahadur representation and involves a second-order interaction between edges. The calculation is made additionally complex by incorporating the change-point, because the algorithm searches for the location of the change-point within a sequence of networks. The number of iterations of the proposed algorithm relies on the initial parameters, a common problem with the EM-type algorithm. A good initial value could lead to much faster algorithm convergence. Based on our numerical experiments, spectral clustering provides an excellent initial value for node memberships. In addition, we adopt a parallel computing strategy, because the node memberships can be updated in a parallel fashion, and estimations using different initial values can proceed independently. In the Supplemental Materials, we provide the computation times of the proposed algorithm and other competing algorithms.

Furthermore, we extend the proposed method to a multiple change-point scenario, and provide a speed-up algorithm based on the screening and ranking method. Further details about the algorithm for estimating multiple change-points can be found in the Supplemental Materials.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

3. Theoretical properties

In this section, we establish the consistency property of the membership estimations and the change-point detection. We require the following assumptions, which are common when using the SBM model (Celisse et al., 2012).

Assumption 1 (Identifiability for community)**.** For every $q \neq q'$, there exists $l \in \{1, \ldots, K\}$ such that

$$\pi_{q,l} \neq \pi_{q',l} \quad or \quad \pi_{l,q} \neq \pi_{l,q'}.$$

Assumption 1 requires that the nodes from different communities have unique probabilities for connecting with other nodes, which ensures that different communities can be identified by the marginal mean of the edges. Therefore, this assumption prevents the identifiability issue arising from a large community consisting of several smaller homogeneous communities with the same probabilities of connections between nodes.

Assumption 2 (Parameter space)**.** For every $q, l \in \{1, \ldots, K\}$, we have $\zeta = c\frac{\log n}{n}$ such that

$$\pi_{q,l} \in [\zeta, 1 - \zeta].$$

Assumption 2 imposes a bound for the node degree. We allow the probability of being close to zero for generating an edge, and Assumption 2 always holds with $\zeta > 0$ so that $\log \pi_{q,l}$ and $\log (1 - \pi_{q,l})$ are valid.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

Assumption 3 (No empty class)**.** For every $q \in \{1, \ldots, K\}$, there exists some constant $\gamma \in (0, 1/K)$ such that for any realization of an SBM with a dependence structure, the labeling vectors $\boldsymbol{z}^1$ and $\boldsymbol{z}^2$ satisfy

$$\#\{1 \leq i \leq n | z_i^d = q\} \geq \gamma n, \quad \text{for } q = 1, \ldots, K \text{ and } d = 1, 2.$$

This assumption implies that no community is empty. With this assumption, the number of nodes in each community is $O(n)$.

Assumption 4 (Searching space)**.** We assume $\tau^\star \in \mathbb{T} \equiv [t_0 T, t_0 T + 1, \ldots, t_1 T]$, which satisfies $0 < t_0 < t_1 < 1$, and $T \log n / n \to \infty$.

Assumption 4 requires that the true change-point location is not at the beginning or the end of the observations, which is a standard assumption in change-point analysis.

Assumption 5 (Sparsity of higher-order correlations)**.** The number of high-order correlations beyond the second-order correlations does not exceed the order of the number of second-order correlations. Specifically, for any t and any edge (i_1, j_1), $\#\{(i_2, j_2), (u_2, v_2) : corr(Y^t_{i_1 j_1}, Y^t_{i_2 j_2} Y^t_{u_2 v_2}) \neq 0\} \leq O(n^2)$, and for any edge pair $(i_1, j_1), (i_2, j_2)$, $\#\{(u_1, v_1), (u_2, v_2) : corr(Y^t_{i_1 j_1} Y^t_{u_1 v_1}, Y^t_{i_2 j_2} Y^t_{u_2 v_2}) \neq 0\} \leq O(n^2)$. In addition, we assume that any two edges $Y^t_{i_1 j_1}$ and $Y^t_{i_2 j_2}$ within a community have a nonnegative correlation $corr(Y^t_{i_1 j_1}, Y^t_{i_2 j_2}) \geq 0$, for $t = 1, \ldots, T$.

Assumption 5 characterizes the sparsity of high-order correlations between within-community edges. Here, we do not impose any restrictions on the second-order correla-

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

tions, because the pairwise interactions between edges are the most relevant. It is common to assume that high-order correlations are negligible in dependent network modeling. For example, Frank and Strauss (1986) introduce a Markov graph that allows a special correlation structure between edges, where only pairs of joint edges are dependent, and higher-order correlations between edges can be ignored. In the GEE2 approach (Garrett Fitzmaurice, 2009), the third-order and fourth-order correlations are assumed to be equal, and correlations beyond four are zero. We often need to simplify high-order correlations, because these can significantly increase the computational cost and instability, while providing negligible benefits in terms of improving membership identification and change-point detection. The nonnegative correlations between edges in the same community are also sensible in practice. For example, a positive pairwise correlation between edges is more likely to produce a star or triad relation in a network (Robins et al., 2007a,b).

In the following, we establish our theoretical results under the independent case ($\rho = 0$) and dependent case ($\rho > 0$).

When the network connectivities are independent, that is, $\rho_k^d = 0$, we define the regularized log likelihood as

$$L(\tau, \boldsymbol{Z}^1, \boldsymbol{Z}^2, \boldsymbol{\pi}) = \frac{1}{Tn(n-1)} \log L(\boldsymbol{Y}|\tau, \boldsymbol{\pi}, \boldsymbol{Z}^1, \boldsymbol{Z}^2), \tag{3.1}$$

where $L(\boldsymbol{Y}|\tau, \boldsymbol{\pi}, \boldsymbol{Z}^1, \boldsymbol{Z}^2)$ is defined as in (2.3). We obtain the estimators of the change-

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

point and the node memberships from

$$(\widehat{\tau}, \widehat{\boldsymbol{z}}^1, \widehat{\boldsymbol{z}}^2) = \underset{\tau\in\mathbb{T},\boldsymbol{z}^1,\boldsymbol{z}^2\in\{1,\ldots,K\}^n,\boldsymbol{\pi}\in[0,1]^{K^2}}{\operatorname{argmax}} L(\tau, \boldsymbol{Z}^1, \boldsymbol{Z}^2, \boldsymbol{\pi}). \tag{3.2}$$

Theorem 1 provides the consistency of the node memberships and change-point simultaneously under the independent case, because estimating a change-point and the community structure are joint processes, owing to their mutual influence on each other. Therefore, the consistency of the two estimators cannot be shown separately. Note the proof that the membership vectors $\boldsymbol{z}^{1\star}$ and $\boldsymbol{z}^{2\star}$ are nonrandom under the conditional SBM framework, and that the change-points and node memberships are treated as parameters of the space $\mathbb{T}\times\{1,\ldots,K\}^{2n}$. Therefore, we ccan establish the consistency of the two estimation processes simultaneously, following the consistency of M-estimators (Van der Vaart, 2000).

Theorem 1. Suppose assumptions 1 to 4 hold and $\rho = 0$, and $(\widehat{\tau}, \widehat{\boldsymbol{z}}^1, \widehat{\boldsymbol{z}}^2)$ is defined in (3.2). Then, $(\widehat{\tau}, \widehat{\boldsymbol{z}}^1, \widehat{\boldsymbol{z}}^2)$ converges to $(\tau^\star, \boldsymbol{z}^{1\star}, \boldsymbol{z}^{2\star})$ in probability as n and T go to infinity.

Theorem 1 indicates that the independent likelihood estimators are strongly consistent if there are no correlations between edges. In the independent case, the edges both within the same observation network and from different observation networks are considered independent samples from the underlying edge-generating model. Therefore, increases in the network size or the number of observation times contribute to the sample size, which is also supported by the rate of convergence in Theorem 3.

We incorporate the dependent part, and thus the number of terms in (2.5) is $O(Tn^4)$.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

Therefore we adjust the log-likelihood function by averaging over the total terms, and redefine the regularized log-likelihood function as

$$L(\tau, \boldsymbol{Z}^1, \boldsymbol{Z}^2, \boldsymbol{\varrho}, \boldsymbol{\pi}) = \frac{1}{Tn^4} \log L(\boldsymbol{Y}|\tau, \boldsymbol{\pi}, \boldsymbol{\varrho}, \boldsymbol{Z}^1, \boldsymbol{Z}^2), \tag{3.3}$$

where $L(\boldsymbol{Y}|\tau, \boldsymbol{\pi}, \boldsymbol{\varrho}, \boldsymbol{Z}^1, \boldsymbol{Z}^2)$ is defined as in (2.5). The estimators of nodes' membership and change-points are

$$(\widehat{\tau}, \widehat{\boldsymbol{z}}^1, \widehat{\boldsymbol{z}}^2) = \underset{\tau \in \mathbb{T}, \boldsymbol{z}^1, \boldsymbol{z}^2 \in \{1,\ldots,K\}^n, \boldsymbol{\pi} \in [0,1]^{K^2}, \boldsymbol{\varrho} \in [0,1]^K}{\operatorname{argmax}} L(\tau, \boldsymbol{Z}^1, \boldsymbol{Z}^2, \boldsymbol{\varrho}, \boldsymbol{\pi}). \tag{3.4}$$

The following theorem states the consistency for the estimators under dependent connectivities.

Theorem 2. Suppose assumptions 1 to 5 hold, and $(\widehat{\tau}, \widehat{\boldsymbol{z}}^1, \widehat{\boldsymbol{z}}^2)$ is defined in (3.4). Then, $(\widehat{\tau}, \widehat{\boldsymbol{z}}^1, \widehat{\boldsymbol{z}}^2)$ converges to $(\tau^\star, \boldsymbol{z}^{1\star}, \boldsymbol{z}^{2\star})$ in probability as n and T go to infinity.

In addition to the consistency of the proposed method, we establish its theoretical advantages over the independent likelihood function in (3.2). The proposed objective function is robust in the sense that it guarantees the memberships and change-point recovery, regardless of whether or not the dependencies exist between edges. In contrast, the independent likelihood approach cannot guarantee consistency for the memberships and change-point when the correlation $\rho > 0$. In addition to the guaranteed consistency property, the proposed method leads to faster convergence than that of its independent counterpart by incorporating the dependency of edges, as shown in the following theorem.

Let $P^\star := P(\boldsymbol{Y}|\tau = \tau^\star, \boldsymbol{z}^1 = \boldsymbol{z}^{1\star}, \boldsymbol{z}^2 = \boldsymbol{z}^{2\star}, \boldsymbol{\pi} = \boldsymbol{\pi}^\star, \boldsymbol{\varrho} = \boldsymbol{\varrho}^\star)$ denote the conditional

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

distribution of edges given the true change-point, the true membership of nodes, and true parameters. We define $T_r = c_1 T(r_1 + r_2) + |\tau^\star - \tau|\big(|r_1 - r| \wedge |r_2 - r|\big)$, where $r_1 = ||\boldsymbol{z}^1 - \boldsymbol{z}^{1\star}||_0$ and $r_2 = ||\boldsymbol{z}^2 - \boldsymbol{z}^{2\star}||_0$ are the numbers of misclassified nodes, and $r = ||\boldsymbol{z}^{1\star} - \boldsymbol{z}^{2\star}||_0$ is the number of nodes that change memberships. The constant ρ is the largest pairwise correlation among within-community edges, and C_1, C_2, c_1 and c_2 are some positive constants. Then, Theorem 3 provides the joint convergence rate for the change-point and memberships.

Theorem 3. Suppose assumptions 1 to 5 hold. Then for every $t > 0$ and $(\tau, \boldsymbol{z}^1, \boldsymbol{z}^2) \neq (\tau^\star, \boldsymbol{z}^{1\star}, \boldsymbol{z}^{2\star})$,

$$P^\star\Big\{\frac{L(Y|\boldsymbol{\pi}, \tau, \boldsymbol{Z}^1, \boldsymbol{Z}^2)}{L(Y|\boldsymbol{\pi}, \tau^\star, \boldsymbol{Z}^{1\star}, \boldsymbol{Z}^{2\star})} > t\Big\} = \mathcal{O}(\exp\Big\{-C_1 \frac{T_r n}{[3 + 2\rho\gamma(r_1 + r_2 + r)n]\log n}\Big\}), \tag{3.5}$$

where L denotes the independent approximate likelihood defined in (3.1), and

$$P^\star\Big\{\frac{L(Y|\boldsymbol{\pi}, \boldsymbol{\varrho}, \tau, \boldsymbol{Z}^1, \boldsymbol{Z}^2)}{L(Y|\boldsymbol{\pi}, \boldsymbol{\varrho}, \tau^\star, \boldsymbol{Z}^{1\star}, \boldsymbol{Z}^{2\star})} > t\Big\} = \mathcal{O}(\exp\Big\{-C_2 \frac{T_r(n + n^3 I\{\rho > 0\})}{[3 + c_2\rho\gamma(r_1 + r_2 + r)n]\log n}\Big\}), \tag{3.6}$$

where L denotes the dependent approximate likelihood defined in (3.3).

For the independent likelihood approach, if there is no pairwise correlation among the edges, that is, $\rho = 0$, then the convergence rate increases to $\mathcal{O}(\exp(-Tn))$ in (3.5), which implies the achieved consistency of the estimation. Additionally, if the true change-point is given such that $|\tau - \tau^\star| = 0$, then it degenerates to a community detection problem with multiple networks, where the convergence rate is consistent with the results in Yuan

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

and Qu (2021). However, when correlations among edges exist, that is, $\rho > 0$, even if the estimations of the memberships are correct, implying $r_1 = r_2 = 0$, and $T_r = |\tau - \tau^\star|r$, the convergence rate is $\mathcal{O}(\exp(-C_1\frac{|\tau-\tau^\star|rn}{3+2\rho\gamma rn}))$, implying that the convergence of the change-point estimation could fail. Intuitively, the independent likelihood approach uses the edges only as cumulative samples, and the dependencies between edges diminish the effective sample size. Therefore the convergence rate deteriorates and could lead to an inconsistent estimation. Under the most extreme scenario, $\pi_{l,q} = \pi_{l',q'}$, and there is no marginal information in the networks about the change-point. Then, the independent likelihood (2.3) does not change with respect to τ, and the estimator of the change-point can be any value in $\mathbb{T}$, implying that the location of the change-point cannot be identified if only marginal information is used.

Under the proposed likelihood approach, given the same number of networks T and node size n, the convergence in (3.6) is faster with (3.5). When dependencies between edges exist, (3.6) implies that the consistency of the independent likelihood approach can only be achieved by increasing the number of networks, but does not benefit from increasing the number of nodes. In contrast the consistency of the proposed method incorporating edge dependency leads to faster convergence, as indicated in (3.6), with an additional n^3 on the exponent part. Furthermore, the independent likelihood approach accumulates information from the first-order marginal mean of the edges, whereas the proposed method also incorporates the pairwise interactions between edges, thus using the second-order cor-

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

relation information between edges. It is not surprising that the increased effective sample size also leads to a faster convergence rate. In addition, given that the true memberships are obtained, the convergence rate of the change-point estimation is $\mathcal{O}(\exp(-n^2))$, indicating that the consistency of the change-point estimation is achieved.

4. Simulation study

In this section,we conduct simulation studies to confirm whether the properties of the proposed method hold in finite samples. To justify the broad applicability of our method, we consider two models. in which where the number of communities is two and four, respectively. In addition, we consider settings with various change-point locations, correlations, and membership-switching scenarios.

We draw T = 40 sample networks from a latent time-varying generative model, as in Section 2.1. We discuss three scenarios for $t > \tau^\star$:

- **(Balance)** Only partial nodes change their community memberships, and the number of nodes in each community remains the same after a change-point.
- **(Unbalance)** partial nodes change their community memberships, and the number of nodes in each community is not the same after a change-point.
- **(No change)** There is no change of membership.

Table 1 provides detailed membership-switching scenarios under different settings. To simplify the notation, we use the following notation to represent the size of each com-

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

munity, and the number in brackets represents the initial community label. For example, $\{20(1)|20(2)\} \rightarrow \{15(1), 5(2)|15(2), 5(1)\}$ indicates that there are 20 nodes in community 1 and community 2, respectively, at the beginning, and then five members of community 1 change their membership to community 2.

We set the correlation coefficient within a community as $\rho = 0$ and $\rho = 0.3$ to represent the independent and dependent cases, respectively. In general, there is a difference between inter-community connectivities and intra-community connectivities. Here, we set the block probability as $\pi_{i,i} = 0.6$ and $\pi_{i,j} = 0.3$, respectively, where $1 \leq i \neq j \leq K$. For the dependent case, we reduce the gap between two probabilities to show a weak signal case, and set the block probability to $\pi_{i,i} = 0.6$ and $\pi_{i,j} = 0.4$, respectively.

We set the change-point as $\tau^\star = rT$, where $r = 0.3, 0.4, 0.5$. In a change-point analysis, the change-point needs to be away from the endpoint of the time space, and we set $\mathbb{T} = [0.1T, 0.9T]$, where $T = 40$.

We compare the performance of the proposed method with that of three existing methods, namely the EM approach, applying the Bahadur approximation ($\text{EM}_{Bahadur}$), but not incorporating the dynamic feature (Yuan and Qu, 2021), the dynamic SBM (Matias and Miele, 2017), and dynamic network clustering using the variational Bayesian algorithm (Sewell and Chen, 2017). Note that Sewell and Chen (2017) provide two methods, namely the variational Bayesian algorithm and the Gibbs sampler, for the projection model. Here, we choose the former because of the similar performance of the two methods. Bhat-

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

tacharjee et al. (2020) propose a fast computational strategy that ignores the underlying community structure, and focuses on change-point estimation. This method use a spectral clustering algorithm, which leads to an inconsistent number of communities, and therefore a low adjusted rand index (ARI). Thus, we do not include this method in the comparison.

We use the ARI to measure the performance of clustering. The ARI takes a value between -1 and 1, and a higher ARI represents better clustering performance. The ARI can also yield negative values if the index is less than the expected one. We calculate the ARI at each time point and the corresponding average ARI value. All results are based on 100 replications.

Table 2 shows the average ARI for the four methods when $n = 40$, which is the average of 100 replications under various settings. The proposed method (EM_{cp}) outperforms other competing methods under all settings. The $\text{EM}_{Bahadur}$ assumes no change-point, and the estimated community memberships tend to be closer to the true membership after the change-point. This is because the change-point occurs in the first half of the time interval, and most samples are generated from the SBM after the change-point. The dynamic SBM performs well only when $\rho = 0$, but performs poorly when $\rho \neq 0$ and the edges are correlated. The variational Bayesian algorithm for the projection model performs inadequately, because the density of inter-community and intra-community edges is close in all settings, which leads to a low signal strength for the variational Bayesian algorithm. The proposed method demonstrates the advantage of incorporating within-community dependency. Note

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

that even if there is no change-point, the proposed method still performs well and is similar to the $\text{EM}_{Bahadur}$. This confirms the robustness of the proposed method against the correct specification of change-point.

Table 3 provides the ARI when $n = 100$. When the network data are dependent, the proposed method outperforms the other competing methods. Specifically, when the number of communities increases, the performance of the dynamic SBM deteriorates rapidly. In addition, when the connectivities between nodes are independent, the proposed method performs slightly worse than Dynsbm, but the ARI still exceeds 0.96 under all settings. This may be because that the percentage of nodes that change community membership is lower than $n = 40$, making the case closer to a smooth change case. However Table 4 shows that when the node size increases to $n = 200$, the performance of the proposed method is comparable to that of Dynsbm, even under the independent case, and the proposed method still performs best under the dependent case.

Owing to space limitations, the results for $n = 1000$ and the multiple change-point scenarios are provided in the Supplementary Material. along with an investigation of the robustness of the proposed method when the percentage of membership changes is small. In conclusion, the simulation studies support the theory in finite samples, and illustrate that the proposed method is effective under the SBM model framework with a potential change-point.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

5. Real-data analysis

In this section, we illustrate the proposed method using dynamic brain network data. Attention deficit hyperactivity disorder (ADHD) is one of the most commonly diagnosed child-onset neurodevelopmental disorders, and understanding the dynamics of brain function plays a significant role in diagnosing ADHD. We analyzed a data set from the ADHD-200 Global Competition, which includes demographic information and resting-state fMRI data for nearly 1000 children and adolescents, including on combined types of ADHD (ADHD-C) and typically developing control (TDC). The data were collected from eight participating sites. To avoid potential site bias, we focus our analysis on the fMRI data from the New York University site only. We removed subjects with a missing diagnostic status or missing scans. The final data set consists of 73 ADHD-C subjects and 98 TDC subjects. The data set contains 116 regions of interest (ROI), measured over 172 time points for each subject. Each node represents an ROI. The average empirical correlations of the connections between these regions are 0.166 and 0.171 for TDC and ADHD-C subjects, respectively. Considering that there are usually more between-communitie edges than there are within-community edges, in general, the within-community correlation coefficient is much larger, indicating that the connectivity dependency should not be ignored.

Because it is difficult to directly observe whether connectivities exist between the ROI, a common practice is to generate a functional connectivity network based on the correlations between all possible node pairs (Cary et al., 2017; He et al., 2018; Hilger and Fiebach,

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

2019). In this study, we apply the SPACE method of Peng et al. (2009) to generate networks at given time points. Specifically, we generate networks every five time points at time points 1,6,11, ..., 96 to reduce the time dependence and to estimate the precision matrix using the SPACE method. Thus, the number of nodes $n = 116$ and the number of time points $T = 20$. The number of edges, on average, over T is 825. To determine the number of communities, we use the Louvain method (Blondel et al., 2008) for community detection for a network from each individual. Then, we obtain the average of the calculated number of communities over 20 individuals, and round it to an integer, whichin this case is four.

Figure 1 displays the approximate likelihoods at different fixed time-points for the ADHD-C and TDC data sets, showing a possible change-point at $t = 86$ among TDC subjects, and at $t = 51$ among ADHD-C subjects. Figure 2 and Figure 3 provide visualizations of the ADHD networks before the change-point and immediately after the change-point, respectively, using the BrainNet Viewer (Xia et al., 2013). Note that we use different colors to identify the four communities to which the nodes belong.

The detailed memberships of the communities before and after the change-point can be found in the Supplementary Material. It is noticeable that the brain network for the ADHD-C subjects has a large community initially, but then there is a rapid change, and some of the ROIs from the largest community move to other communities. In addition, changes in TDC children's brain networks are smoother than those of ADHD-C children. This may

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

be because ADHD patients are more likely to be distracted during the experiment.

We compare the proposed method with the dynamic SBM (Matias and Miele, 2017) for dynamic ADHD-C brain networks. Table 6 provides the number of ROIs at different communities before and after each time point using the dynamic SBM, showing that only 22 ROIs changed communities. However, our method shows that 58 ROIs change their communities. In fact, at the 51 and 56 time points, excluding edges that have never been connected, 61.8% of the edges changed their connectivity status from zero to one or vice versa. This implies that the ADHD-C patients' brain network connectivities changed more dramatically, which is consistent with ADHD symptoms such as being inattentive and having a short attention span, whereas the dynamic SBM is not able to capture this phenomenon.

Table 7 shows the changed ROIs of the ADHD group, which are different from TDC group. The different ROIs mainly concentrate on the frontal gyrus, cingulate gyrus, cerebellum and cerebellar vermis, precentral gyrus, postcentral gyrus, and temporal gyrus. The prefrontal cortex is responsible for many more complex mental functions, including planning complex cognitive behavior, personality expression, decision-making, and moderating social behavior. ADHD is highly associated with alterations in the prefrontal cortex (Arnsten and Li, 2005). The cingulate gyrus is associated with cognitive processes such as emotional processing and the vocalization of emotions, and there is evidence of anterior cingulate dysfunctions in ADHD patients (Bush et al., 2005). The cerebellum coordinates

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

voluntary movements such as posture, balance, coordination, and speech. Dysfunction in the cerebellum and anomalies in the cerebellar vermis in ADHD patients have been established (Toplak et al., 2006). The precentral gyrus is the site of the primary motor cortex, and is involved in the planning, control, and execution of voluntary movements. The postcentral gyrus is the location of the primary somatosensory cortex, and is associated with ADHD (Fassbender et al., 2011). The temporal lobe consists of structures that are vital for declarative or long-term memory, and less temporal gray matter volume was found in ADHD children (Castellanos et al., 2002; Carmona et al., 2005). Compared with the TDC group, many ROIs of ADHD-C networks in brain anatomical regions have changed, which is, in general, consistent with the current clinical literature on ADHD, as mentioned above.

6. Conclusion

In this paper, we investigate the problem of simultaneous change-point identification in network community structures. We propose a new approximate likelihood method to integrate both marginal and correlation information among network communities to estimate the change-point and the corresponding community memberships. The proposed method provides flexible modeling of the underlying a joint distribution assumption.

Theoretically, we establish the consistency of both the change-point and the membership estimations for the proposed approximate likelihood under some regular conditions. In addition, we show the superiority of the proposed method compared with the indepen-

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

dent likelihood approach, because the membership estimator achieves a faster convergence rate, while obtaining the consistency of the change-point estimation.

The proposed method can be implemented efficiently, and numeric studies indicate that it can improve clustering performance over that of the independent model and other existing methods, even under moderate dependency within-community connectivities. In addition, in the application to dynamic fMRI brain network data, the proposed method detects brain functional community changes associated with ADHD that are not captured by other methods without incorporating within-community dependency among the functional connectivities.

We also consider balancing between saving on the computation cost and retaining high-order information, so that our method can deal with larger networks. This is common in real data, such as a brain network with ROIs as nodes. However, analyzing very large-scale networks can be time consuming, and thus improving the efficiency of this process is important, and is left to future research.

One advantage of the proposed method is that we can directly estimate the memberships at each time point, without pretesting the existence of the change-point. Both our numerical experiments (see the results in Table 2 to Table 4) and our theoretical findings (Yuan and Qu, 2021) show that even if the change-point does not occur, the community detection still performs well.

The proposed methed is based on the SBM framework, but the degree-corrected block

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

model (DCBM) can be used more broadly. However, even when there is no change-point, the DCBM is much more challenging than the SBM, both theoretically and practically (Chen et al., 2018; Gao et al., 2018; Wilson et al., 2019). In fact, the existence of a change-point and connectivity dependence introduce additional parameters, which make the theoretical analysis more challenging. In addition, in a functional brain network, the heterogeneity of the node degree is not strong, and the SBM can capture the community structure quite well (Le et al., 2018; Levin et al., 2019; Wang et al., 2019). The extension of the proposed method to the DCBM is left to future research. In addition, we can investigate community structure change detection by allowing a dynamic number of communities, where communities can devide or merge with each other. This could have significant value for evolution modeling of complex networks.

Acknowledgments

The authors thank the associate editor and two referees, for their helpful comments and suggestions. This research was supported by US NSF DMS 1952406 and DMS 2210640. Additionally, Li's work was supported by Zhejiang Provincial Natural Science Foundation of China (LQ23A010008), and Zhang's work was partially supported by NSF China (11971116).

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

Supplementary Material

The online Supplementary Material includes detailed proofs of the main theorems and lemmas, numerical results on a large-scale network, the robustness of the proposed method and multiple change-point estimation, and detailed tables related to the ADHD data analysis.

References

Arnsten, A. F. and Li, B.-M. (2005). Neurobiology of executive functions: Catecholamine influences on prefrontal cortical functions. *Biological Psychiatry*, 57(11):1377–1384.

Bahadur, R. R. (1961). A representation of the joint distribution of responses to n dichotomous items. *Studies in Item Analysis and Prediction* 169–176.

Bhattacharjee, M., Banerjee, M., and Michailidis, G. (2020). Change point estimation in a dynamic stochastic block model. *Journal of Machine Learning Research*, 21(107):1–59.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.

Bush, G., Valera, E. M., and Seidman, L. J. (2005). Functional neuroimaging of attention-deficit/hyperactivity disorder: A review and suggested future directions. *Biological Psychiatry*, 57(11):1273–1284.

Carmona, S., Vilarroya, O., Bielsa, A., Tremols, V., Soliva, J., Rovira, M., Tomas, J., Raheb, C., Gispert, J., Batlle, S., and Bulbena, A. (2005). Global and regional gray matter reductions in adhd: A voxel-based morphometric study. *Neuroscience Letters*, 389(2):88–93.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

Cary, R. P., Ray, S., Grayson, D. S., Painter, J., Carpenter, S., Maron, L., Sporns, O., Stevens, A. A., Nigg, J. T., and Fair, D. A. (2017). Network structure among brain systems in adult ADHD is uniquely modified by stimulant administration. *Cerebral Cortex*, 27(8):3970–3979.

Castellanos, F. X., Lee, P. P., Sharp, W., Jeffries, N. O., Greenstein, D. K., Clasen, L. S., Blumenthal, J. D., James, R. S., Ebens, C. L., and Walter, J. M. (2002). Developmental trajectories of brain volume abnormalities in children and adolescents with attention-deficit/hyperactivity disorder. *JAMA*, 288(14):1740–1748.

Celisse, A., Daudin, J.-J., and Pierre, L. (2012). Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electronic Journal of Statistics*, 6:1847–1899.

Chen, Y., Li, X., and Xu, J. (2018). Convexified modularity maximization for degree-corrected stochastic block models. *The Annals of Statistics*, 46(4):1573–1602.

Cheng, J., Levina, E., Wang, P., and Zhu, J. (2014). A sparse Ising model with covariates. *Biometrics*, 70(4):943–953.

Dubey, P., Müller, H.-G., et al. (2020). Fréchet change-point detection. *Annals of Statistics*, 48(6):3312–3335.

Fassbender, C., Schweitzer, J. B., Cortes, C. R., Tagamets, M. A., Windsor, T. A., Reeves, G. M., and Gullapalli, R. (2011). Working memory in attention deficit/hyperactivity disorder is characterized by a lack of specialization of brain function. *PloS one*, 6(11):e27240.

Frank, O. and Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, 81(395):832–842.

Gao, C., Ma, Z., Zhang, A. Y., and Zhou, H. H. (2018). Community detection in degree-corrected block models. *The Annals of Statistics*, 46(5):2153–2185.

Garrett Fitzmaurice, Marie Davidian, G. V. G. M., editor (2009). *Longitudinal Data Analysis: Handbooks of Modern*

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

Statistical Methods. CRC Press.

Gibberd, A. J. and Nelson, J. D. (2017). Regularized estimation of piecewise constant Gaussian graphical models: The group-fused graphical lasso. *Journal of Computational and Graphical Statistics*, 26(3):623–634.

He, Y., Lim, S., Fortunato, S., Sporns, O., Zhang, L., Qiu, J., Xie, P., and Zuo, X.-N. (2018). Reconfiguration of cortical networks in MDD uncovered by multiscale community detection with fMRI. *Cerebral Cortex*, 28(4):1383–1395.

Heaukulani, C. and Ghahramani, Z. (2013). Dynamic probabilistic models for latent feature propagation in social networks. In *International Conference on Machine Learning* 275–283.

Hilger, K. and Fiebach, C. J. (2019). ADHD symptoms are associated with the modular structure of intrinsic brain networks in a representative sample of healthy adults. *Network Neuroscience*, 3(2):567–588.

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

Le, C. M., Levin, K., and Levina, E. (2018). Estimating a network from multiple noisy realizations. *Electronic Journal of Statistics*, 12(2):4697–4740.

Levin, K., Lodhia, A., and Levina, E. (2019). Recovering low-rank structure from multiple networks with unknown edge distributions. *arXiv preprint arXiv:1906.07265*.

Liu, G., Wang, Y., and Orgun, M. A. (2011). Trust transitivity in complex social networks. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2011, San Francisco, California, USA, August 7-11, 2011*.

Marangoni-Simonsen, D. and Xie, Y. (2015). Sequential changepoint approach for online community detection. *IEEE Signal Process. Lett.*, 22(8):1035–1039.

Matias, C. and Miele, V. (2017). Statistical clustering of temporal networks through a dynamic stochastic block model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(4):1119–1141.

Mei, Q. and Zhai, C. (2005). Discovering evolutionary theme patterns from text: An exploration of temporal text mining. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining* 198–207. ACM.

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

Park, H. and Lee, K. (2014). Dependence clustering, a method revealing community structure with group dependence. *Knowledge-Based Systems*, 60:58–72.

Peng, J., Wang, P., Zhou, N., and Zhu, J. (2009). Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association*, 104(486):735–746.

Robins, G., Pattison, P., Kalish, Y., and Lusher, D. (2007a). An introduction to exponential random graph (p*) models for social networks. *Social networks*, 29(2):173–191.

Robins, G., Snijders, T., Wang, P., Handcock, M., and Pattison, P. (2007b). Recent developments in exponential random graph (p*) models for social networks. *Social Networks*, 29(2):192–215.

Sarkar, P. and Moore, A. W. (2006). Dynamic social network analysis using latent space models. In *Advances in Neural Information Processing Systems* 1145–1152.

Sewell, D. K. and Chen, Y. (2017). Latent space approaches to community detection in dynamic networks. *Bayesian Analysis*, 12(2):351–377.

Toplak, M. E., Dockstader, C., and Tannock, R. (2006). Temporal information processing in adhd: Findings to date and

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

new methods. *Journal of Neuroscience Methods*, 151(1):15–29.

Toyoda, M. and Kitsuregawa, M. (2003). Extracting evolution of web communities from a series of web archives. In *Proceedings of the Fourteenth ACM Conference on Hypertext and Hypermedia* 28–37. ACM.

Van der Vaart, A. W. (2000). *Asymptotic Statistics*, volume 3. Cambridge University Press.

Wang, D., Yu, Y., and Rinaldo, A. (2018). Optimal change point detection and localization in sparse dynamic networks. *arXiv preprint arXiv:1809.09602*.

Wang, S., Arroyo, J., Vogelstein, J. T., and Priebe, C. E. (2019). Joint embedding of graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(4):1324–1336.

Wang, Y., Chakrabarti, A., Sivakoff, D., and Parthasarathy, S. (2017a). Fast change point detection on dynamic social networks. *arXiv preprint arXiv:1705.07325*.

Wang, Y., Chakrabarti, A., Sivakoff, D., and Parthasarathy, S. (2017b). Hierarchical change point detection on dynamic networks. In *Proceedings of the 2017 ACM on Web Science Conference* 171–179. ACM.

Wilson, J. D., Stevens, N. T., and Woodall, W. H. (2019). Modeling and detecting change in temporal networks via the degree-corrected stochastic block model. *Quality and Reliability Engineering International*, 35(5):1363–1378.

Xia, M., Wang, J., and He, Y. (2013). BrainNet Viewer: A network visualization tool for human brain connectomics. *PloS one*, 8(7):e68910.

Xing, E. P., Fu, W., and Song, L. (2010). A state-space mixed membership blockmodel for dynamic network tomography. *The Annals of Applied Statistics*, 4(2):535–566.

Xu, K. S. and Hero, A. O. (2014). Dynamic stochastic blockmodels for time-evolving social networks. *IEEE Journal of*

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

Selected Topics in Signal Processing, 8(4):552–562.

Yang, J. and Peng, J. (2018). Estimating time-varying graphical models. *arXiv preprint arXiv:1804.03811*.

Yang, T., Chi, Y., Zhu, S., Gong, Y., and Jin, R. (2011). Detecting communities and their evolutions in dynamic social networks-a Bayesian approach. *Machine Learning*, 82(2):157–189.

Yapeng, L., Yuanyuan, Q., Xi, C., Wei, L., and Gianluigi, F. (2013). Exploring the functional brain network of alzheimer's disease: Based on the computational experiment. *Plos One*, 8(9):e73186.

Yuan, Y. and Qu, A. (2021). Community detection with dependent connectivity. *The Annals of Statistics*, 49(4):2378–2428.

Zhao, Z., Chen, L., and Lin, L. (2019). Change-point detection in dynamic networks via graphon estimation. *arXiv preprint arXiv:1908.01823*.

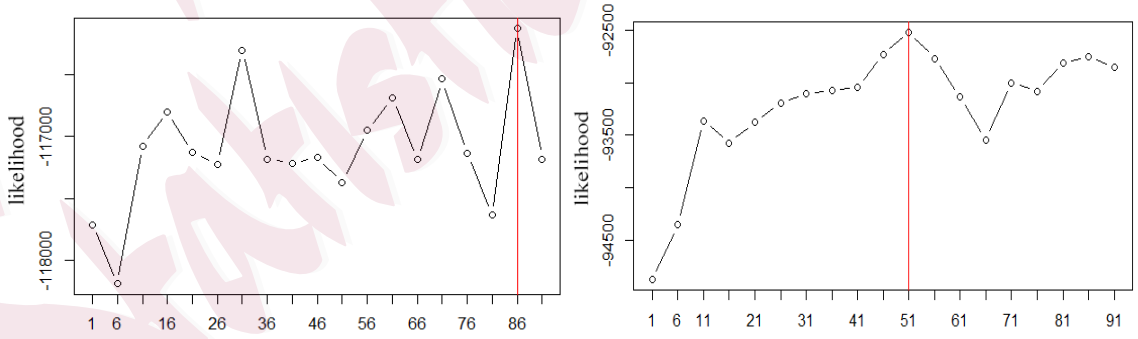
Table 1: Scenarios of membership change under different settings.

		Balance	Unbalance
K=2	n=40	$\{20(1)\|20(2)\}$ $\to \{15(1), 5(2)\|15(2), 5(1)\}$	$\{20(1)\|20(2)\}$ $\to \{15(1)\|20(2), 5(1)\}$
	n=100	$\{50(1)\|50(2)\}$ $\to \{40(1), 10(2)\|40(2), 10(1)\}$	$\{50(1)\|50(2)\}$ $\to \{40(1)\|50(2), 10(1)\}$
	n=200	$\{100(1)\|100(2)\}$ $\to \{80(1), 20(2)\|80(2), 20(1)\}$	$\{100(1)\|100(2)\}$ $\to \{80(1)\|100(2), 20(1)\}$
K=4	n=40	$\{10(1)\|10(2)\|10(3)\|10(4)\}$ $\to \{5(1), 5(4)\|5(2), 5(1)\|5(3), 5(2)\|5(4), 5(3)\}$	$\{10(1)\|10(2)\|10(3)\|10(4)\}$ $\to \{5(1)\|10(2), 5(1)\|5(3)\|10(4), 5(3)\}$
	n=100	$\{25(1)\|25(2)\|25(3)\|25(4)\}$ $\to \{15(1), 10(4)\|15(2), 10(1)\|15(3), 10(2)\|15(4), 10(3)\}$	$\{25(1)\|25(2)\|25(3)\|25(4)\}$ $\to \{15(1)\|25(2), 10(1)\|15(3)\|25(4), 10(3)\}$
	n=200	$\{50(1)\|50(2)\|50(3)\|50(4)\}$ $\to \{30(1), 20(4)\|30(2), 20(1)\|30(3), 20(2)\|30(4), 20(3)\}$	$\{50(1)\|50(2)\|50(3)\|50(4)\}$ $\to \{30(1)\|50(2), 20(1)\|30(3)\|50(4), 20(3)\}$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

Table 2: Adjusted rand index between the estimated memberships and the true memberships for networks with $n = 40$.

			$K = 2$				$K = 4$			
	$\rho^\star$	$\tau^\star/T$	EM_{cp}	$\text{EM}_{Bahadur}$	Dynsbm	VB	EM_{cp}	$\text{EM}_{Bahadur}$	Dynsbm	VB
Balanced	0	0.3	1	0.828	0.952	0.414	1	0.805	0.564	0.748
		0.4	1	0.771	0.953	0.441	1	0.715	0.365	0.747
		0.5	1	0.713	0.949	0.467	1	0.652	0.247	0.755
	0.3	0.3	0.997	0.825	0.401	0.372	0.992	0.788	0.212	0.378
		0.4	0.999	0.770	0.392	0.375	0.995	0.706	0.187	0.375
		0.5	1	0.713	0.382	0.356	0.995	0.645	0.176	0.380
Unbalanced	0	0.3	1	0.724	0.971	0.466	1	0.867	0.763	0.750
		0.4	1	0.632	0.969	0.444	1	0.816	0.691	0.743
		0.5	1	0.540	0.968	0.425	1	0.757	0.657	0.758
	0.3	0.3	0.984	0.712	0.444	0.334	0.988	0.843	0.349	0.360
		0.4	0.993	0.630	0.465	0.345	0.993	0.805	0.333	0.354
		0.5	0.992	0.538	0.454	0.306	0.994	0.728	0.318	0.361
No change	0	—	1	1	0.982	0.661	1	1	0.869	0.826
	0.3	—	1	1	0.430	0.339	1	1	0.272	0.375



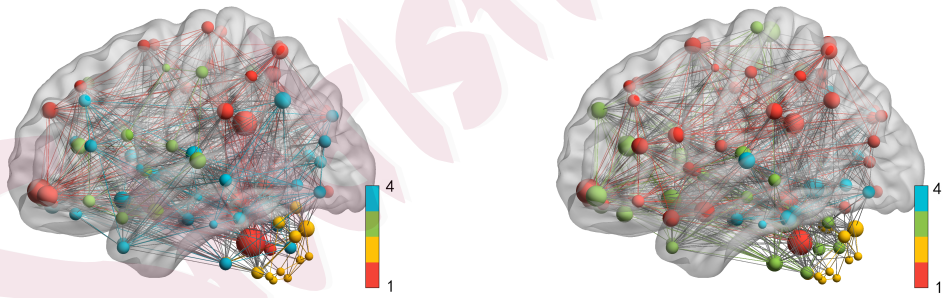
TDC children ADHD children

Figure 1: Likelihood at different time points

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

Table 3: Adjusted rand index between the estimated memberships and the true memberships for networks with $n = 100$.

			$K = 2$				$K = 4$			
	$\rho^\star$	$\tau^\star/T$	EM_{cp}	$\text{EM}_{Bahadur}$	Dynsbm	VB	EM_{cp}	$\text{EM}_{Bahadur}$	Dynsbm	VB
Balanced	0	0.3	0.994	0.823	0.999	0.689	0.991	0.805	0.986	0.988
		0.4	0.997	0.744	1	0.733	0.987	0.746	0.987	0.989
		0.5	1	0.685	0.999	0.767	0.995	0.680	0.986	0.989
	0.3	0.3	0.992	0.799	0.714	0.354	0.969	0.797	0.547	0.354
		0.4	0.996	0.734	0.718	0.376	0.976	0.751	0.541	0.376
		0.5	0.997	0.670	0.717	0.327	0.998	0.654	0.532	0.327
Unbalanced	0	0.3	0.965	0.903	0.992	0.805	0.942	0.882	0.991	0.994
		0.4	0.979	0.864	0.992	0.812	0.984	0.830	0.991	0.994
		0.5	0.981	0.821	0.991	0.809	0.997	0.808	0.991	0.994
	0.3	0.3	0.964	0.884	0.707	0.263	0.940	0.863	0.707	0.303
		0.4	0.976	0.848	0.713	0.303	0.969	0.830	0.713	0.263
		0.5	0.98	0.811	0.705	0.282	0.984	0.787	0.705	0.282
No change	0	—	1	1	1	0.979	1	1	0.994	0.997
	0.3	—	0.997	1	0.744	0.237	0.998	1	0.564	0.611



Before change After change

Figure 2: Network for TDC children

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182

Table 4: Adjusted rand index between the estimated memberships and the true memberships for networks with $n = 200$.

			$K = 2$				$K = 4$			
	$\rho^\star$	$\tau^\star/T$	EM_{cp}	$\text{EM}_{Bahadur}$	Dynsbm	VB	EM_{cp}	$\text{EM}_{Bahadur}$	Dynsbm	VB
Balanced	0	0.3	1	0.822	1	0.918	1	0.814	0.979	0.999
		0.4	1	0.762	1	0.927	1	0.739	0.978	0.999
		0.5	1	0.669	1	0.941	1	0.694	0.978	0.999
	0.3	0.3	1	0.800	0.796	0.433	1	0.812	0.648	0.670
		0.4	1	0.736	0.781	0.379	1	0.732	0.641	0.665
		0.5	1	0.671	0.778	0.375	1	0.681	0.634	0.675
Unbalanced	0	0.3	1	0.891	1	0.974	1	0.894	0.985	1
		0.4	1	0.864	1	0.976	1	0.852	0.985	1
		0.5	1	0.835	1	0.976	1	0.805	0.985	1
	0.3	0.3	0.982	0.884	0.743	0.366	0.992	0.867	0.680	0.594
		0.4	0.992	0.848	0.749	0.363	0.996	0.833	0.671	0.621
		0.5	0.995	0.812	0.754	0.269	1	0.797	0.670	0.646
No change	0	—	1	1	1	1	1	1	1	1
	0.3	—	0.998	1	0.792	0.251	1	1	0.681	0.704

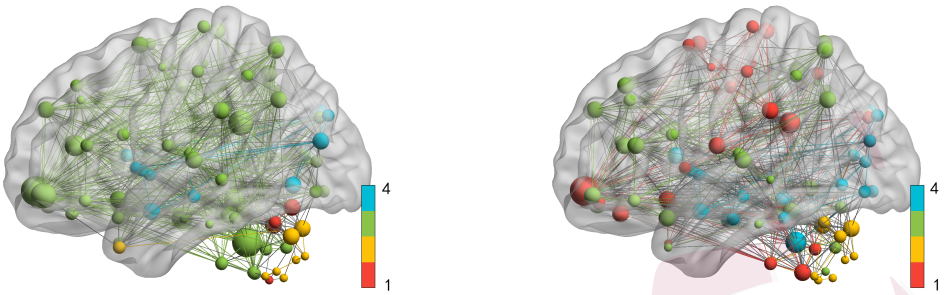
Table 5: Adjusted rand index between the estimated memberships and the true memberships for networks with $n = 40$.

		$K = 2$				$K = 4$			
	$\rho^\star$	EM_{cp}	$\text{EM}_{Bahadur}$	Dynsbm	VB	EM_{cp}	$\text{EM}_{Bahadur}$	Dynsbm	VB
Balanced	0	0.941	0.828	0.834	0.595	0.905	0.730	0.340	0.719
	0.3	0.937	0.819	0.704	0.470	0.917	0.668	0.473	0.379
Unbalanced	0	0.929	0.718	0.887	0.549	0.842	0.646	0.672	0.775
	0.3	0.888	0.703	0.701	0.421	0.899	0.603	0.321	0.401

Table 6: The number of ROIs that belong to different communities before and after each time point using the dynamic SBM.

time point	1	6	11	16	21	26	31	36	41	46	51	56	61	66	71	76	81	86	91	96
numbers	26	27	27	23	23	25	20	28	20	22	22	19	24	16	19	25	17	22	15	

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0182



Before change After change

Figure 3: Network for ADHD-C children

Table 7: Changed regions of ADHD group.

ROI	community	ROI	community	ROI	community
Cerebelum_Crus1_R	1→2	Cingulum_Mid_R	3→1	Occipital_Inf_R	3→4
Cerebelum_4_5_R	1→4	Cingulum_Post_L	3→1	Fusiform_L	3→4
Cerebelum_8_R	1→2	Cingulum_Post_R	3→1	Fusiform_R	3→4
Cerebelum_7b_L	2→3	ParaHippocampal_L	3→4	SupraMarginal_L	3→1
Precentral_R	3→1	ParaHippocampal_R	3→4	Precuneus_L	3→1
Frontal_Sup_Orb_L	3→1	Calcarine_R	3→4	Temporal_Inf_R	3→4
Supp_Motor_Area_L	3→1	Cuneus_L	3→1	Cerebelum_9_R	3→2
Supp_Motor_Area_R	3→1	Lingual_R	3→4	Vermis_3	3→4
Olfactory_R	3→1	Occipital_Mid_R	3→4	Vermis_10	3→4
Cingulum_Mid_L	3→1	Occipital_Inf_L	3→4	Insula_R	4→3