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
Notice: Accepted version subject to English editing.	

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

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

[1] *Zhejiang Gongshang University* [2] *University of California Irvine* [3] *Fudan University*

Abstract: In dynamic network analysis, there has been a surge of interest in community detection and evolution. However, existing methods for dynamic community detection do not consider the dependency among edges, which could lead to information loss in detecting community structures. In this paper, we investigate the problem of identifying a change-point with abrupt changes in community structure of a network. We propose an approximate likelihood approach for the change-point estimator as well as node membership identification through integrating both marginal information and dependency of network connectivities. An EM-type algorithm is proposed for maximizing the approximate likelihood jointly over change-point and community membership evolution. In theory, we establish estimation consistency under the regularity condition and show that the proposed estimators achieve a higher convergence rate compared with their marginal likelihood counterpart without incorporating dependency among edges. The validity of the proposed method is supported via application on the ADHD-200 dataset for detecting brain functional community change over time.

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

1. Introduction

Network data analysis has arisen as an important tool to study relationships and associations among subjects. In this paper, we develop community detection for dynamic network data. Traditional network analysis assumes that the network connectivities are independently and identically distributed, and networks are static over time. However, in practice, these assumptions fail while dynamic changes of sequential observed network data over time are ubiquitous such as in social networks, political networks, trading networks, brain networks, biological networks, among others.

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

Compared to static networks, dynamic community detection is much more challenging since it involves not only node classification, but also the dynamic change of the node grouping over time. Existing works analyze community extraction and community dynamics involving two separate steps: first, static analysis is applied to snapshots of the networks at given time points, and then community change over time is detected afterwards (Toyoda and Kitsuregawa, 2003; Mei and Zhai, 2005; Palla et al., 2007). However, this strategy assumes that networks are independent from each other at each discrete time point, which fails to capture the community dynamics for the entire period of time, and therefore community detection could be unstable. In addition, the label switching problem could arise between two successive time points (Matias and Miele, 2017).

The existing literature on dynamic community detection is mainly based on stochastic block models (SBM), 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) combines an SBM for its static part with Markov chains for the change of node groups through time. Alternatively, latent space models are applied to capture the probability of dynamic connectivities of two nodes via the closeness of the corresponding latent points (Sarkar and Moore, 2006; Heaukulani and Ghahramani, 2013; Sewell and Chen, 2017).

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

the underlying distribution of community structure will not change over time, such as the stationary Markov chain. However, abrupt changes of communities could occur in practice due to the sudden occurrence of external events, or changes of 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 takes place. In the social network, individuals join a new community due to their changes of geographical locations. The network of stock trading could experience high volatility during an economic crisis or unforeseen disaster such as pandemic. Therefore, the dynamic network methods designed for a smooth change are not applicable under these scenarios, and it is desirable to construct an appropriate model for an abrupt change in that the underlying distribution of community structure changes at a certain time point.

There are limited studies on abrupt changes in dynamic community detection. Among them, Marangoni-Simonsen and Xie (2015) proposed detecting the emergence of a community in large networks from sequential observations using Erdős-Renyi random graphs. Wang et al. (2017b) detected change-points in dynamic networks under the snapshot model at a given time. Wang et al. (2017a) studied hierarchical change-point detection to differentiate intra-community and inter-community evolution assuming that the community structure is known in advance. In general, these approaches mainly compare two sequential networks to determine the possible location of a change-point through the differences between two networks. Dubey et al. (2020) proposed a test statistic to infer the presence

of change-point within a sequence of network distributions, but is incapable of recovering communities simultaneously. However, the estimation of change-point and community structure are joint processes due to their mutual influence on each other. In addition, the estimation of change-point conditional on given community structures could be biased if a given community structure is incorrect. For joint estimation, Bhattacharjee et al. (2020) considered an unknown community estimation with a single change-point via a spectral-based method, but the key assumption for a consistent estimation is hard to verify. Wang et al. (2018) proposed a weighted network aggregating method to detect the change-point in sparse network sequences. Zhao et al. (2019) established a non-parametric approach for detecting the change of underlying network structure via neighborhood smoothing.

The main limitations of the aforementioned methods are that they assume that connectivities are conditionally independent given the membership of nodes, which lead to the loss of high-order network information on community structure change. In addition, this assumption is practically infeasible as most of 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 social network shows the dependency of trustworthiness among unknown participants (Liu et al., 2011). In addition, correlations of brain networks connectivity between patients with Alzheimer's Disease and healthy controls could be very different (Yapeng et al., 2013). To incorporate dependency among

edges in static networks, Cheng et al. (2014) proposed a conditional distribution model of binary network data, and incorporated covariate information into the Ising model. Park and Lee (2014) developed a clustering method which incorporates group dependence through geometric structure, and Yuan and Qu (2021) proposed a truncated Bahadur representation to incorporate the underlying dependency structure among connectivities. However, incorporating edge dependency has not been considered under dynamic network settings.

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

in that the proposed method outperforms existing approaches.

The main advantages and contributions of the proposed method can be summarized as follows. We consider community detection with abrupt changes under the framework of dependent dynamic stochastic block model. Apart from the marginal mean difference between communities, the proposed method also considers within-community dependency among connectivities during the dynamic evolution of community structure. In practice, the concordance among within-community edges is an important intrinsic characteristic of communities. Consequently, incorporating within-community dependency could enhance discriminative power to identify the community memberships at change-points. This is especially helpful when the marginal mean information is not informative. Due to mutual influence between the estimations of change-point and community structures, we propose an algorithm to estimate change-point and identify community structures jointly. Establishing the consistency property of both change-point and community membership estimations simultaneously is non-trivial, and much more challenging for incorporating conditional dependency among edges, compared to the existing ones assuming conditional independence.

This paper is organized as follows: Section 2 introduces notation and presents the proposed method and the implementation algorithm; Section 3 illustrates the theoretical properties of the proposed method. Section 4 provides simulation studies, and Section 5 illustrates an application to real data on ADHD brain networks. The last section provides

conclusions and some further discussion.

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 the dynamic change-point under the framework of the stochastic block model.

Let $\boldsymbol{Y} = \{(Y_{ij}^t)_{n\times n}\}_{t=1}^T$ be T symmetric and unweighted sample networks, and $\{Y_{ij}^t\}$ is 1 if there is an edge between nodes i and j at time point t and 0 otherwise, where n is the number of nodes which belong to one of K communities. Here, 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 superscript d to distinguish variables before and after a change-point, where $d = 1$ stands for before change-point and $d = 2$ stands for 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 $\boldsymbol{Z}^d = \{(Z_{ik}^d)_{n\times K}\} \in \{0,1\}^{n\times K}$, 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 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 the within-community connectivities dependency, and denote the correlation among edges $\left(Y_{i_1j_1}^t, Y_{i_2j_2}^t\right)$ within a community to be $\text{corr}\left(Y_{i_1j_1}^t, Y_{i_2j_2}^t\right) = \rho_{i_1,j_1,i_2,j_2}^t$.

In the following, we propose change-point identification utilizing global network information, through the probability matrix $\boldsymbol{\pi}$. The goal of this paper is to identify $\tau^\star$ as well as 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 restrictive to one time-point only. We define the parameter space as $\Theta = (\boldsymbol{\pi}, \tau, \boldsymbol{z}^1, \boldsymbol{z}^2)$. Let $\tau \in \mathbb{T} \equiv [t_0T, t_0T + 1, \ldots, t_1T]$, where $0 < t_0 < t_1 < 1$ and t_0T and t_1T 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:

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

Then the change-point and membership can be estimated simultaneously as the 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, and therefore allows one to change community structure via utilizing 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 that edges within communities are more correlated, such as friendship in social networks. Therefore, assuming the conditional independence assumption among connectivities is restrictive and practically infeasible because it could lead to significant information loss of the community structure change.

In this section, we propose an approximate likelihood function to incorporate within-community correlation following Yuan and Qu (2021), and therefore improve accuracy and efficiency in identifying community structure and change-point detection. In addition to the edges' marginal mean information, within-community dependency contains addi-

tional information regarding the membership of nodes. This is especially effective when the marginal mean is not informative in differentiating between and within communities' connectivity.

In contrast to Yuan and Qu (2021), we assume that the correlation coefficient between edges can also change over time since the nodes' membership changes over time. Here, we use a homogeneous correlation structure such that all the pairwise correlations from each community are assumed to be the average within-community correlation. The rationales for this simplification are based on the following. Firstly, the pairwise correlation parameter $\rho^t_{i_1 i_2 j_1 j_2} = \operatorname{corr}\left(Y^t_{i_1 j_1}, Y^t_{i_2 j_2}\right)$ is a nuisance correlation parameter to enhance clustering and identification. Secondly, Theorem 3.3 shows that the density of pairwise correlation among within-community edges plays a more important role than the intensity of the correlation in affecting clustering performance. Thirdly, 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 due 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 calculation. More critically, we have nearly $n^4/4$ correlation coefficients to estimate, for which it is infeasible to estimate all sample correlation coefficients. Therefore, we assume a homogeneous correlation structure such that all the pairwise correlations in each community are assumed to be the average

within-community correlation. Specifically, given nodes i_1, j_1, i_2 and j_2 are in the same community q, $\text{corr}\left(Y^t_{i_1j_1}, Y^t_{i_2j_2}\right) = \rho^d_q$, 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 to combine information about nodes' membership from both marginal mean information and within-community dependency in the following:

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

where

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

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

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 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 regarding the marginal mean is kept in $\log L^1_{ind}(\tau)+\log L^2_{ind}(\tau)$, which is consistent with 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 interac-

tion term

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

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

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 utilize the method of profiling such that maximizing $\log L(Y|\boldsymbol{\pi}, \tau, \boldsymbol{\varrho}, \boldsymbol{z}^1, \boldsymbol{z}^2)$ over parameters $\boldsymbol{\pi}, \boldsymbol{\varrho}, \boldsymbol{z}^1, \boldsymbol{z}^2$ given a fixed change-point τ, i.e.,

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

Then we search τ from $[t_0T, t_1T]$ 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 following 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 due to the fact that the conditional distribution $P(\boldsymbol{z}^1, \boldsymbol{z}^2|Y)$ becomes intractable in the expectation step. Therefore, we modify the EM algorithm fol-

lowing variational methods which approximate the likelihood $P(\boldsymbol{z}^1, \boldsymbol{z}^2|Y)$ by 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 $\sum_{q=1}^{K} \mu_{iq}^d = 1, d = 1, 2$. Notice that only in the optimization process, $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ are treated as random. In our theoretical analysis, we still consider $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ to be the fixed parameters.

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

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

If the above Bayes factor (2.7) > 1, then the probability of node i in community q is larger than that of 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 two communities, the marginal part ratio is close to 1. However if the correlation ratio is informative, the Bayes factor in (2.7) is still differentiable to enhance 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

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

For each node, the membership $\boldsymbol{z}^d$ is determined by identifying the largest probability of $\boldsymbol{z}_i^d(\tau)$ such that

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

Algorithm 1 provides the implementation algorithm for the change-point identification and community detection jointly.

Remark 2.1. In order to ensure its nonnegativity of correlation among edges, when we calculate the correlation coefficients in E step of Algorithm 1, if the moment estimator of correlation is negative, we set the correlation to be zero. This also ensures that the likelihood function valid since the log part is required to be positive.

The computational complexity of the proposed EM-type algorithm at each iteration is $O(Tn^4K^2)$, which is mainly determined through the nodes' memberships updating at the E step. Specifically, the complexity for calculating the Bayes factor for each single observed network is $O(n^4K^2)$, since it contains the correlation factor using the Bahadur representation and involves a second-order interaction among edges. Additional computa-

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: $\max_i \left(|\boldsymbol{\mu}_i^{1(s)} - \boldsymbol{\mu}_i^{1(s-1)}|_1 + |\boldsymbol{\mu}_i^{2(s)} - \boldsymbol{\mu}_i^{2(s-1)}|_1\right) < \epsilon$.
  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 $z_i^d = z_i^{d(s)}$.

tion complexity is due to incorporating the change-point, since the algorithm searches for the location of the change-point from a sequence of networks. The number of iterations of the proposed algorithm relies on the initial parameters, a common problem that arises from the EM-type algorithm. A good initial value could lead to a much faster algorithm convergence. Based on our numerical experiments, spectral clustering provides an excellent initial value for nodes' memberships. In addition, we adopt a parallel computing strategy to accelerate the proposed algorithm, since the nodes' memberships can be updated in a parallel fashion and estimations using different initials can proceed independently. In the supplemental materials, we provide computation time of the proposed algorithm and other

competing algorithms.

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

3. Theoretical properties

In this section, we establish the consistency property of the membership estimations as well as the change-point detection. We require the following assumptions, which are common in 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 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 among 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.

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 SBM with dependence structure, the labeling vectors $\boldsymbol{z}^1$ and $\boldsymbol{z}^2$ satisfy

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

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 imposes 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 in that the potential change-point is away from the endpoints of the time interval.

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), $\#\{(i_2, j_2), (u_2, v_2) :$

$corr(Y^t_{i_1j_1}, Y^t_{i_2j_2}Y^t_{u_2v_2}) \neq 0\} \leq O(n^2)$, and for any edge pair $(i_1, j_1), (i_2, j_2)$, $\#\{(u_1, v_1), (u_2, v_2) : corr(Y^t_{i_1j_1}Y^t_{u_1v_1}, Y^t_{i_2j_2}Y^t_{u_2v_2}) \neq 0\} \leq O(n^2)$. In addition, we assume that any two edges $Y^t_{i_1j_1}$ and $Y^t_{i_2j_2}$ within a community have a non-negative correlation $corr(Y^t_{i_1j_1}, Y^t_{i_2j_2}) \geq 0$ for $t = 1, \ldots, T$.

Assumption 5 characterizes the sparsity of high-order correlation among within-community edges. Here we do not impose any restrictions on the second-order correlations since the pairwise interactions among 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) introduces a Markov graph to allow a special correlation structure among edges where only pairs of joint edges are dependent and higher-order correlations among edges can be ignored. In GEE2 approach (Garrett Fitzmaurice, 2009), the third-order and fourth-order correlations are assumed to be equal while correlations beyond four are zeros. Simplifying high-order correlations are often needed, since higher-order correlations could significantly increase computational cost and instability while the benefit of improving membership identification and change-point detection is negligible. The non-negative correlations among edges in the same community are also sensible in practice. For example, the positive pairwise correlation among edges is more likely to produce star or triad relations in network (Robins et al., 2007a,b).

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

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

$$L(\tau, \boldsymbol{Z}^1, \boldsymbol{Z}^2, \boldsymbol{\pi}) = \frac{1}{Tn(n-1)} \log L(\boldsymbol{Y}|\tau, \boldsymbol{\pi}, \boldsymbol{Z}^1, \boldsymbol{Z}^2), \tag{3.10}$$

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

$$(\widehat{\tau}, \widehat{\boldsymbol{z}}^1, \widehat{\boldsymbol{z}}^2) = \underset{\tau \in \mathbb{T}, \boldsymbol{z}^1, \boldsymbol{z}^2 \in \{1,\ldots,K\}^n, \boldsymbol{\pi} \in [0,1]^{K^2}}{\operatorname{argmax}} L(\tau, \boldsymbol{Z}^1, \boldsymbol{Z}^2, \boldsymbol{\pi}). \tag{3.11}$$

Theorem 3.1 provides consistency of memberships and change-point simultaneously under the independent case, since the estimation of change-point and community structure are joint processes due to their mutual influence on each other. Therefore, the consistency of the two estimators cannot be shown separately. Note the proof that the membership vector $\boldsymbol{z}^{1\star}$ and $\boldsymbol{z}^{2\star}$ is non-random under the conditional SBM framework, and the change-points and memberships of nodes are treated as parameters of space $\mathbb{T} \times \{1, \ldots, K\}^{2n}$. Therefore the consistency of the two estimation processes can be established simultaneously following the consistency of M-estimators (Van der Vaart, 2000).

Theorem 3.1. Suppose assumptions 1 to 4 hold and $\rho = 0$, and $(\widehat{\tau}, \widehat{\boldsymbol{z}}^1, \widehat{\boldsymbol{z}}^2)$ is defined in (3.11). 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 3.1 indicates that the independent likelihood estimators are strongly consistent if there are no correlations among edges. In the independent case, both the edges within the same observation network and from different observation networks are con-

sidered as independent samples from the underlying edge-generating model. Therefore, either the increase of network size or the number of observation times contributes to the sample size, which is also supported by the rate of convergence in Theorem 3.3.

We incorporate the dependent part and thus the number of terms in (2.5) is $O(Tn^4)$. So we adjust the log likelihood function through averaging over total terms and redefine the regularized log likelihood function as

$$L(\tau, \boldsymbol{Z}^1, \boldsymbol{Z}^2, \boldsymbol{\varrho}, \boldsymbol{\pi}) = \frac{1}{Tn^4} \log L(\boldsymbol{Y}|\tau, \boldsymbol{\pi}, \boldsymbol{\varrho}, \boldsymbol{Z}^1, \boldsymbol{Z}^2), \tag{3.12}$$

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

$$(\widehat{\tau}, \widehat{\boldsymbol{z}}^1, \widehat{\boldsymbol{z}}^2) = \underset{\tau\in\mathbb{T}, \boldsymbol{z}^1, \boldsymbol{z}^2\in\{1,\ldots,K\}^n, \boldsymbol{\pi}\in[0,1]^{K^2}, \boldsymbol{\varrho}\in[0,1]^K}{\operatorname{argmax}} L(\tau, \boldsymbol{Z}^1, \boldsymbol{Z}^2, \boldsymbol{\varrho}, \boldsymbol{\pi}). \tag{3.13}$$

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

Theorem 3.2. Suppose assumptions 1 to 5 holds, and $(\widehat{\tau}, \widehat{\boldsymbol{z}}^1, \widehat{\boldsymbol{z}}^2)$ is defined in (3.13). 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.

Besides the consistency of the proposed method, we establish the theoretical advantages over the independent likelihood function in (3.11). The proposed objective function is robust in the sense that it guarantees the memberships and change-point recovery whether the dependency among edges exists or not. In contrast, the independent likelihood approach cannot guarantee consistency regarding the memberships and change-point

when correlation $\rho > 0$. In addition to the guaranteed consistency property approach, the proposed method leads to faster convergence compared to the independent counterpart through incorporating dependency of edges, which are illustrated by 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 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 $r_1 = ||\boldsymbol{z}^1 - \boldsymbol{z}^{1\star}||_0$, $r_2 = ||\boldsymbol{z}^2 - \boldsymbol{z}^{2\star}||_0$ are the numbers of misclassified nodes, and $r = ||\boldsymbol{z}^{1\star} - \boldsymbol{z}^{2\star}||_0$ is the number of nodes which change memberships. The constant ρ is the largest pairwise correlation among within-community edges, and C_1, C_2, c_1, c_2 are some positive constants. Then the following Theorem 3.3 provides the joint convergence rate for the change-point and memberships.

Theorem 3.3. Suppose assumptions 1 to 5 hold, then for every $t > 0$ and $(\tau, \boldsymbol{z}^1, \boldsymbol{z}^2) \neq (\tau^\star, \boldsymbol{z}^{1\star}, \boldsymbol{z}^{2\star})$,

$$P^\star\Big\{\frac{L(Y|\boldsymbol{\pi}, \tau, \boldsymbol{Z}^1, \boldsymbol{Z}^2)}{L(Y|\boldsymbol{\pi}, \tau^\star, \boldsymbol{Z}^{1\star}, \boldsymbol{Z}^{2\star})} > t\Big\} = \mathcal{O}(\exp\Big\{-C_1 \frac{T_r n}{[3 + 2\rho\gamma(r_1 + r_2 + r)n]\log n}\Big\}), \tag{3.14}$$

where L denotes the independent approximate likelihood defined in (3.10), and

$$P^\star\Big\{\frac{L(Y|\boldsymbol{\pi}, \boldsymbol{\varrho}, \tau, \boldsymbol{Z}^1, \boldsymbol{Z}^2)}{L(Y|\boldsymbol{\pi}, \boldsymbol{\varrho}, \tau^\star, \boldsymbol{Z}^{1\star}, \boldsymbol{Z}^{2\star})} > t\Big\} = \mathcal{O}(\exp\Big\{-C_2 \frac{T_r(n + n^3 I\{\rho > 0\})}{[3 + c_2\rho\gamma(r_1 + r_2 + r)n]\log n}\Big\}), \tag{3.15}$$

where L denotes the dependent approximate likelihood defined in (3.12).

For the independent likelihood approach, if there is no pairwise correlation among edges, i.e, $\rho = 0$, then the convergence rate increases to $\mathcal{O}(\exp(-Tn))$ in (3.14), and implies the achieved consistency of 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 and Qu (2021). However, when the correlations among edges exist, i.e, $\rho > 0$, even if estimations of the memberships are correct, implying $r_1 = r_2 = 0$, and $T_r = |\tau - \tau^\star|r$, the convergence rate is of the order $\mathcal{O}(\exp(-C_1 \frac{|\tau-\tau^\star|rn}{3+2\rho\gamma rn}))$, implying that the convergence of change-point estimation could fail. Intuitively, the independent likelihood approach only utilizes the edges as cumulative samples, and the dependency among edges diminishes the effective sample size, and therefore the convergence rate deteriorates and could lead to inconsistency estimation. Under the most extreme scenario, when $\pi_{l,q} = \pi_{l',q'}$, and there is no marginal information of networks regarding 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 network T and node size n, the convergence in (3.15) is faster with (3.14). When the dependency between edges exists, (3.15) 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 numbers of nodes, while the consistency of the proposed method incorporating edge dependency leads to a faster convergence indicated in (3.15) with an additional n^3 on the exponent part. In contrast to the independent likelihood approach which accumulates information from the first-order marginal mean of edges, the proposed method also incorporates the pairwise interactions among edges, and therefore the second-order correlation information among edges is utilized. It is not surprising that the increased effective sample size also leads to a faster convergence rate. In addition, given the true memberships are obtained, the convergence rate of the change-point estimation has the order of $\mathcal{O}(\exp(-n^2))$, indicating that the consistency of change-point estimation is achieved.

4. Simulation study

In this section,we conduct simulation studies, and confirm whether the properties of the proposed method hold in finite samples. To justify the broad applicability of our method, we consider two models where the number of communities is 2 and 4, respectively. In addition, we consider settings with various change-point locations, correlations as well as different scenarios of membership switching.

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 change-point.

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

Table 1 provides the detailed scenarios of membership switching under different settings. To simplify the notation, We use the following notation to represent the size of each community, 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 5 members in community 1 change their community memberships and become community 2.

We set the correlation coefficient within community as $\rho = 0$ and $\rho = 0.3$ to represent independent and dependent cases respectively. In general, there is a difference between inter-community connectivities and intra-community connectivities, and we set the block probability as $\pi_{i,i} = 0.6$ and $\pi_{i,j} = 0.3$ respectively, where $1 \leq i \neq j \leq K$. While for the dependent case, we reduce the gap between two probabilities to show a weak signal case and set the block probability $\pi_{i,i} = 0.6$ 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 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 three existing methods, including the EM approach applying Bahadur approximation ($\text{EM}_{Bahadur}$) but not incor-

porating dynamic feature (Yuan and Qu, 2021), Dynamic SBM (Matias and Miele, 2017) and Dynamic Network Clustering using the variational Bayesian algorithm (Sewell and Chen, 2017). Note that there are two methods, the variational Bayesian algorithm and the Gibbs sampler for the projection model in Sewell and Chen (2017), and we only choose the former one due to similar performance of the two methods. Bhattacharjee et al. (2020) proposed a fast computational strategy, which ignores the underlying community structure and focus on change-point estimation. This method use a spectral clustering algorithm, which leads to the inconsistent number of communities with the preset and therefore a low ARI. Thus, we do not take this method as comparison.

We use the adjusted rand index (ARI) to measure the performance of clustering. The ARI takes a value between -1 and 1, and a higher ARI represents a 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}) performs better than 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 that the change-point occurs in the first half of the time interval, and most samples are generated from the SBM after the change-point. Dynamic

SBM only performs well when $\rho = 0$, but has poor performance when $\rho \neq 0$ and edges are correlated. The variational Bayesian algorithm for the projection model performs inadequately since the density of edges from inter-community and intra-community 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. It is worth noting that even if there is no change-point, the proposed method still performs well and is similar to the $\text{EM}_{Bahadur}$. This also 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 is dependent, the proposed method performs better than the other competing methods. Specifically, when the number of communities increases, the performance of the dynamic SBM deteriorates rapidly. In addition when connectivities among nodes are independent, the proposed method performs slightly worse than Dynsbm, but the ARI still exceeds 0.96 under all settings. This may be explained by that the percentage of nodes changing community memberships is lower than $n = 40$, making the case closer to a smooth change case. But the result in Table 4 shows that when the node size increases to $n = 200$, the performance of the proposed method is comparable to Dynsbm even under independent case, and the proposed method still performs best under dependent case.

Due to space limitations, the results for $n = 1000$ and multiple change-points scenarios are provided in the supplementary material. In addition, we investigate the robustness

of the proposed method when the percentage of membership changes is small in the supplementary material. 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.

5. Real data analysis

In this section, we illustrate the proposed method for dynamic brain network data. Attention deficit hyperactivity disorder (ADHD) is one of the most commonly diagnosed child-onset neurodevelopmental disorders, and understanding dynamics of brain function plays a significant role in diagnosing ADHD. We analyzed a dataset from the ADHD-200 Global Competition, which includes demographical information and resting-state fMRI of nearly one thousand children and adolescents, including 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 the subjects with missing diagnostic status or missing scans. The resulting dataset 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 a region of interest (ROI). The average empirical correlations of the connections among 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 within-community edges, the within-community correlation coefficient is generally much larger, indicating that the connectivity dependency should not be ignored.

Since it is difficult to directly observe whether there is a connectivity between the ROI, a common practice is to generate a functionally connectivity network based on the correlations between all possible node-pairs (Cary et al., 2017; He et al., 2018; Hilger and Fiebach, 2019). In this study, we apply the SPACE method in Peng et al. (2009) to generate networks at given time points. Specifically, we generate networks every 5 time points at time point 1,6,11, ..., 96 to reduce time dependence and estimate the precision matrix using SPACE. 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 perform the Louvain method (Blondel et al., 2008) for community detection for a network from each individual, and obtain the number of communities. Then we obtain the average of the calculated number of communities over 20 individuals and round it to an integer, which is 4 here.

Figure 1 displays the approximate likelihoods at different fixed time-points for the ADHD-C and TDC datasets, showing that there is a possible change-point at $t = 86$ among TDC subjects and at $t = 51$ among ADHD-C subjects. Figure 2 and Figure 3 provide the visualization of the ADHD networks before the change-point and right after the change-point via the BrainNet Viewer (Xia et al., 2013). Note that we use four different colors to distinguish the four communities where each node belongs.

The detailed memberships of 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-size 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 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, and only 22 ROIs changed their communities. However, the result of our method shows 58 ROIs changing their communities. In fact, at the 51 and 56 time points excluding edges which have never been connected, 61.8% of the edges changed their connectivity status, from 0 to 1 or vice versa, implying that the ADHD-C patients' brain network connectivities appeared to change more dramatically, which is consistent with ADHD symptoms such as being inattentive and having a short attention span, while the Dynamic SBM is not able to capture this phenomenon.

Table 7 shows the changed ROIs of 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 pre-frontal 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 process including emotional processing and vocalization of emotions, and there are evidences of anterior cingulate dysfunctions in ADHD patients (Bush et al., 2005). The cerebellum coordinates voluntary movements such as posture, balance, coordination, and speech, and dysfunction in the cerebellum and anomaly 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 on the underlying within-community dependency structure among connectivities without requiring specification of a joint distribution assumption.

Theoretically, we establish consistency of both change-point and membership estimations for the proposed approximate likelihood under some regular conditions. In addition, we show the superiority of the proposed method compared to the independent likelihood approach as 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 the proposed method is able to improve clustering performance compared to 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 change associated with ADHD, which is not captured by other methods without incorporating within-community dependency among functional connectivities.

We also consider to balance between saving computation cost and retaining high-order information so that our method can deal with larger networks. This could be common in real data, e.g. a brain network with ROIs as nodes. However, analyzing very large-scale networks can be time-consuming and should be one of important future direction to achieve efficiency in computing algorithm.

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 the numerical experiments (see results in Table 2 to Table 4) and theoretical findings (Yuan and Qu, 2021) show that even if the change-point does not occur, the community detection can still perform well.

The proposed methed is based on the framework of the SBM. Indeed, the Degree-Corrected Block Model (DCBM) can be used more broadly. However, even in the cases where there are no change-point, the DCBM is much more challenging than the SBM from the theoretical perspective as well as the implementations (Chen et al., 2018; Gao et al., 2018; Wilson et al., 2019). In fact, the existence of the change-point and connectivity dependence introduce more parameters, which make the theoretical analysis more challenging. In addition, in the functional brain network, the heterogeneity of 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 DCBM is still a worth future research direction. In additon, we can further investigate community structure change detection allowing a dynamic number of communities, in that communities can be divided among themselves or merge with each other. This could have a significant value in evolution modeling for complex networks.

7. Acknowledgements

The authors thank an associate editor and two referees, whose helpful comments and suggestions led to a much improved presentation. This research is supported by US NSF DMS 1952406 and DMS 2210640. Additionally, Zhang's work is partially supported by NSF China (11971116).

8. Supplementary Materials

Supplementary material available at *Statistica Sinica* online includes detailed proofs of main Theorems and necessary Lemmas, numerical results on large scale network, robustness of the proposed method and multiple change-points estimation, and detailed tables in 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.

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 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 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 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.

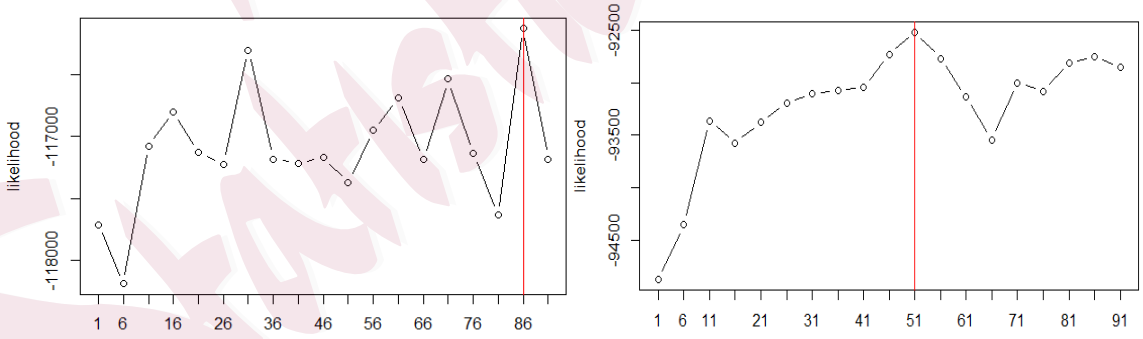
		Balance	Unbalance
K=2	n=40	$\{20(1)\|20(2)\}$ $\to \{15(1), 5(2)\|15(2), 5(1)\}$	$\{20(1)\|20(2)\}$ $\to \{15(1)\|20(2), 5(1)\}$
	n=100	$\{50(1)\|50(2)\}$ $\to \{40(1), 10(2)\|40(2), 10(1)\}$	$\{50(1)\|50(2)\}$ $\to \{40(1)\|50(2), 10(1)\}$
	n=200	$\{100(1)\|100(2)\}$ $\to \{80(1), 20(2)\|80(2), 20(1)\}$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

Table 2: Adjusted Rand Index between estimated memberships and 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

Table 3: Adjusted Rand Index between estimated memberships and 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



TDC children ADHD children

Figure 1: Likelihood at different time points

Table 4: Adjusted Rand Index between estimated memberships and true memberships for networks with $n = 200$.

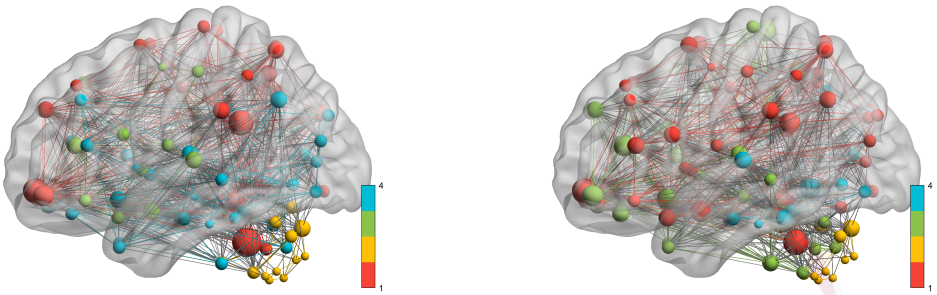
			$K = 2$				$K = 4$			
	$\rho^\star$	$\tau^\star/T$	EM_{cp}	$\text{EM}_{Bahadur}$	Dynsbm	VB	EM_{cp}	$\text{EM}_{Bahadur}$	Dynsbm	VB
Balanced	0	0.3	1	0.822	1	0.918	1	0.814	0.979	0.999
		0.4	1	0.762	1	0.927	1	0.739	0.978	0.999
		0.5	1	0.669	1	0.941	1	0.694	0.978	0.999
	0.3	0.3	1	0.800	0.796	0.433	1	0.812	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 estimated memberships and 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

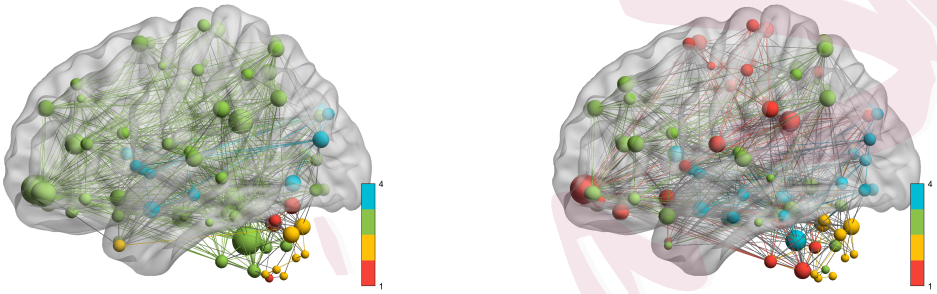
Table 6: The number of ROIs which belong to different communities before and after each time point using Dynamic SBM.

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	



Before change After change

Figure 2: Network for TDC children



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