

---

# Changepoint Estimation in Sparse Dynamic Stochastic Block Models under Near-Optimal Signal Strength

---

**Shirshendu Chatterjee**  
City University of New York,  
Graduate Center and City College,  
New York, NY, USA

**Soumendu Sundar Mukherjee**  
Statistics and Mathematics Unit  
Indian Statistical Institute,  
Kolkata, WB, India

**Tamojit Sadhukhan**  
Statistics and Mathematics Unit  
Indian Statistical Institute,  
Kolkata, WB, India

## Abstract

We consider the offline changepoint estimation problem in the context of multilayer stochastic block models. We develop an algorithm involving suitably chosen CUSUM statistics based on the adjacency matrices of the observed networks for estimating a single changepoint present in the input data. We provide rigorous theoretical guarantees on the performance of the proposed method when one or more of the following phenomena occur at the changepoint: (a) merging of communities, (b) splitting of communities, and (c) changes in the connection probabilities among the communities. We derive a lower bound on the minimax detectability threshold involving the relevant signal strength parameter and show that the proposed algorithm can estimate the changepoint consistently when the signal strength is above a small multiplicative factor times the minimax detectability threshold. We do not make any a priori assumption on the sparsity of the underlying networks and only require that the overall average degree goes to infinity. Via simulation experiments, we empirically show that the proposed algorithm works in regimes of signal strength where global network changepoint estimation algorithms that do not take into account the community structure, fail to estimate an existing changepoint correctly. Finally, we apply our algorithm to a series of networks constructed using roll call data from the US senate and obtain changepoint(s) which align with those reported in the political science literature regarding the phenomenon of increasing political polarization.

## 1 INTRODUCTION

Sequences of networks are now widely available as the main observable or derived data in many fields of research, including time-series of social networks (Panisson et al. (2013); Stopczynski et al. (2014); Rocha et al. (2010)), epidemiological networks (Salathé et al. (2010); Rocha et al. (2011)), animal networks (Gates and Woolhouse (2015); Lahiri and Berger-Wolf (2007)), mobile and online communication networks (Krings et al. (2012); Ferraz Costa et al. (2015); Jacobs et al. (2015)), economic networks (Popović et al. (2014); Zhang et al. (2014)), brain networks (Park and Friston (2013); Sporns (2013)), genetic networks (Rigbolt et al. (2011)), and ecological networks (Blonder et al. (2012)). In addition to more classical statistical problems on temporal networks (e.g., modeling, analysis of statistical behavior and dynamics, community detection (Holme and Saramäki (2012); Holme (2015); Peixoto (2015); Sikdar et al. (2016); Peixoto and Gauvin (2018))), estimation of changepoints have garnered emerging interest recently. Potential applications are in, for instance, brain imaging, where one has brain scans of individuals collected over time and is looking for abnormalities, ecological networks observed over time, where one wonders if there is a structural change, etc. Motivated by these observations, in this article, we consider the problem of estimating changepoint(s) in a finite sequence of networks having community structure. Our aims are (a) to obtain the information-theoretic detectability threshold (in the minimax sense) for the signal strength associated with multilayer stochastic block models (MSBM) having changepoints, and (b) to develop an efficient polynomial-time changepoint estimation procedure that performs consistently even when the input MSBM is extremely sparse (e.g., when the degrees of the nodes are sub-logarithmic in the number of nodes) and has very low signal strength.

Estimation of changepoints in sequentially observed data is a classical problem in statistics going back to the early days of statistical quality control. We refer the readers to Bhattacharyya et al. (2020) for pointers to classical works on changepoint estimation as well as a discussion of recent works on the network changepoint problem. Recent works

that are most relevant to us, include Mukherjee (2018); Wang et al. (2021); Bhattacharjee et al. (2020); Zhao et al. (2019); Bhattacharyya et al. (2020). These works propose methods for changepoint estimation in networks generated from several related models and also provide theoretical guarantees on the consistency of the proposed methods. Bhattacharjee et al. (2020) consider changepoint problems in the context of a dynamic stochastic block model but only tackle the setting of dense graphs. Mukherjee (2018); Wang et al. (2021) (resp. Bhattacharyya et al. (2020)) have proposed changepoint estimation methods for sequences of independent and semi-dense inhomogeneous Erdős-Rényi random graphs, which are consistent when the associated signal strength is near-optimal (resp. optimal). The method proposed by Zhao et al. (2019) is based on graphon estimation and therefore only relevant for dense graphs. Our results in the present article show that when networks have community structures, all of the above-mentioned methods are sub-optimal either in terms of the required signal strength, or in terms of the required assumptions on network sparsity. For example, Mukherjee (2018); Wang et al. (2021); Zhao et al. (2019); Bhattacharyya et al. (2020) require much higher signal strengths for consistency than the minimax threshold for consistent estimation derived in this paper. Our proposed estimation method achieves consistency above this threshold (up to logarithmic factors) when the number of communities is polylogarithmic in the network sizes and the number of networks. Moreover, unlike the present paper, Mukherjee (2018); Wang et al. (2021); Bhattacharjee et al. (2020); Zhao et al. (2019) require the input networks to be dense (or semi-dense) for their theoretical results to hold. We explain these in more detail in Section 2.3.

In this article, our main focus is on obtaining the correct order of signal strength required for consistent estimation of changepoints in the presence of community structures. We have therefore restricted our attention to models with a single changepoint, where this can be achieved in the cleanest possible manner. Our proposed method can be coupled with off-the-shelf meta-algorithms such as *wild binary segmentation* (Fryzlewicz et al. (2014)) to estimate multiple changepoints (see Section 5.2 for an empirical illustration of this). We also note that our method may be applied within sliding time-windows to detect changepoints in an online fashion. These will be the focus of a future work.

## 2 CONTRIBUTIONS AND COMPARISON WITH RELATED WORKS

We begin by describing the changepoint estimation problem for networks with community structure.

### 2.1 Setup

A *multilayer network* having  $n$  nodes and  $T$  layers consists of a finite sequence of undirected graphs  $G^{(t)}$  for  $t \in [T] := 1, \dots, T$  on the same vertex set  $[n] := \{1, 2, \dots, n\}$ , which possibly have different sets of edges at different layers. We refer to  $G^{(t)}$  as the  $t$ -th *network layer* and represent it by the corresponding  $n \times n$  adjacency matrix  $A^{(t)}$ , whose elements are  $A_{ij}^{(t)} \in \{0, 1\}$  for  $i, j \in [n]$ , with  $A_{ij}^{(t)} = 1$  if there is an edge between node  $i$  and node  $j$  in the  $t$ -th layer, and  $A_{ij}^{(t)} = 0$  otherwise.

**Definition 2.1.** A *Multilayer Stochastic Block Model* (MSBM) having  $n$  nodes,  $T$  layers, and  $K$  communities is a model for generating multilayer networks that can be described in terms of the following components.

- (i) The community membership vector  $(z_1, \dots, z_n)$ , where  $z_i \in [K] := \{1, \dots, K\}$  denotes the community index of vertex  $i \in [n]$ , given by the following specifications:

$$z_1, \dots, z_n \stackrel{\text{i.i.d.}}{\sim} \text{Multinomial}(1; \pi_1, \dots, \pi_K),$$

where the  $K \times 1$  vector  $(\pi_1, \dots, \pi_K)$  consists of the probabilities of allocation to different communities.

- (ii) The  $n \times n$  adjacency matrices  $A^{(t)}$ ,  $t \in [T]$ , given by the following specifications:

$$\begin{aligned} \mathbb{P}(A_{ij}^{(t)} = 1 \mid z_i, z_j) &= B_{z_i z_j}^{(t)} \text{ for } i < j, i, j \in [n], \\ A_{i_0 j_0}^{(t_0)} &\perp\!\!\!\perp A_{i_1 j_1}^{(t_1)} \text{ for } (i_0, j_0) \neq (i_1, j_1) \in [n] \times [n], \\ A^{(t_0)} &\perp\!\!\!\perp A^{(t_1)} \text{ for } t_0 \neq t_1 \in [T], \end{aligned}$$

where the  $K \times K$  matrices  $B^{(t)}$ ,  $t \in [T]$  are the community-wise connection probability matrices (to be referred to as the *block-connectivity matrices*).

In an MSBM, the community membership vector  $(z_1, \dots, z_n)$  can be equivalently expressed using the  $n \times K$  *community membership matrix*  $Z$ , defined as

$$Z_{ik} = \mathbf{1}_{\{z_i=k\}} \text{ for } i \in [n] \text{ and } k \in [K].$$

Using  $Z$ , the edge-probability matrix (i.e. the expected adjacency matrix) of the  $t$ -th layer can be written as

$$P^{(t)} := \mathbb{E} A^{(t)} = Z B^{(t)} Z^T, \quad t = 1, \dots, T.$$

Our theoretical results assume that we observe a multi-relational network having  $T$  layers,  $n$  nodes, and community structures that change at a single changepoint  $\tau \in [T]$ . At the changepoint  $\tau$ , one or more of the following phenomena occur:

- (a) some of the existing communities split into multiple communities (in which case the number of communities increases),

- (b) some of the existing communities merge into bigger communities (in which case the number of communities decreases),
- (c) the connectivity probabilities between some pairs of communities change.

The segments before and after the changepoint  $\tau$  are assumed to be samples from MSBMs with constant block-connectivity matrices.

More elaborately, let  $K_0$  (resp.  $K_1$ ) denote the number of communities and

$$C_1^0, \dots, C_{K_0}^0 \subset [n] \quad (\text{resp. } C_1^1, \dots, C_{K_1}^1 \subset [n])$$

denote the communities before (resp. after) the changepoint  $\tau$ . Let the  $n \times K_0$  (resp.  $n \times K_1$ ) community membership matrix before (resp. after) the changepoint  $\tau$  be  $\tilde{Z}_0$  (resp.  $\tilde{Z}_1$ ), given by  $(\tilde{Z}_0)_{ia} = \mathbf{1}_{\{i \in C_a^0\}}$  (resp.  $(\tilde{Z}_1)_{ia} = \mathbf{1}_{\{i \in C_a^1\}}$ ) for  $i \in [n]$  and  $a \in [K_0]$  (resp.  $a \in [K_1]$ ). Let  $\tilde{B}^{(t)}$  denote the associated block-connectivity matrix for the  $t$ -th layer. In our setup, the matrices  $\tilde{B}^{(t)}$ ,  $t \in [T]$  change at  $t = \tau$ ,  $1 \leq \tau \leq T$ . More precisely, there exists  $K_0 \times K_0$  (resp.  $K_1 \times K_1$ ) probability matrix  $\tilde{B}_0$  (resp.  $\tilde{B}_1$ ) such that

$$\tilde{B}^{(t)} = \begin{cases} \tilde{B}_0, & 1 \leq t \leq \tau, \\ \tilde{B}_1, & \tau + 1 \leq t \leq T. \end{cases}$$

The expected adjacency matrix before (resp. after)  $\tau$  can be written as  $P_0 = \tilde{Z}_0 \tilde{B}_0 \tilde{Z}_0^T$  (resp.  $P_1 = \tilde{Z}_1 \tilde{B}_1 \tilde{Z}_1^T$ ).

One specialty of the above setup is that it can be recast as an instance of a single MSBM. In order to see this, consider a directed weighted bipartite graph  $\mathcal{G} = (V_0, V_1, E)$  with  $|V_0| = K_0$  and  $|V_1| = K_1$ . For  $i \in [K_0]$ , the  $i$ -th node in  $V_0$  corresponds to  $C_i^0$ , the  $i$ -th pre-change community and for  $j \in [K_1]$ , the  $j$ -th node in  $V_1$  corresponds to  $C_j^1$ , the  $j$ -th post-change community. A node is called a *parent node* if at the changepoint, the corresponding community either i) splits into two or more communities, or ii) is formed by merging two or more communities, or iii) does not undergo any change and belongs to  $V_0$ . Non-parent nodes are called *child nodes*. We identify the parent node (resp. child node(s)) of a child node (resp. parent node) in the obvious way. We put directed weighted edge(s) from a parent node to its child node(s), with the weights equal to the size of the community corresponding to the child node. Let  $K$  be the number of child nodes in  $\mathcal{G}$ . For  $a \in [K]$ , let  $C_a$  be the community corresponding to the  $a$ -th child node in  $\mathcal{G}$ . The community structure formed by the  $K$ -communities  $C_1, \dots, C_K$  is the *lifted community structure*. We define the *lifted*  $n \times K$  membership matrix  $Z$  by

$$Z_{ia} = \mathbf{1}_{\{i \in C_a\}} \text{ for } i \in [n] \text{ and } a \in [K].$$

Let  $\mathfrak{C} := \{C_k^0 \cap C_l^1 : k \in [K_0] \text{ and } l \in [K_1]\}$ . It is easy to see that  $K = |\mathfrak{C}|$  and  $C_1, \dots, C_K$  is an enumeration of  $\mathfrak{C}$ .

However,  $K$  is typically much smaller than its maximum possible value  $K_0 K_1$ . Observe that for each  $a \in [K]$ , we know the following about  $C_a$ , the  $a$ -th child node in  $\mathcal{G}$ :

- whether it belongs to  $V_0$  or  $V_1$ ;
- if it belongs to  $V_0$  (resp.  $V_1$ ), its original index in  $V_0$  (resp.  $V_1$ ), which we denote by  $\gamma_a \in [K_0]$  (resp.  $[K_1]$ );
- if it belongs to  $V_0$  (resp.  $V_1$ ), the index of its parent node in  $V_1$  (resp.  $V_0$ ), which we denote by  $\pi_a \in [K_1]$  (resp.  $[K_0]$ ).

Using these information, we define the pre-change (resp. post-change) *lifted*  $K \times K$  block-connectivity matrix  $B_0$  (resp.  $B_1$ ) in the following way. Fix  $k, l \in [K]$ . For  $a \in [K]$ , set

$$u_a = \begin{cases} \gamma_a & \text{if } a \in V_0, \\ \pi_a & \text{if } a \in V_1; \end{cases} \quad v_a = \begin{cases} \pi_a & \text{if } a \in V_0, \\ \gamma_a & \text{if } a \in V_1. \end{cases}$$

Then

$$(B_0)_{kl} := (\tilde{B}_0)_{u_k u_l}, \quad (B_1)_{kl} := (\tilde{B}_1)_{v_k v_l}.$$

The purpose of lifting the number of communities in this way is to consistently describe the communities and the block-connectivity matrices before and after the changepoint. Under this equivalent lifted community structure, the lifted membership matrix  $Z$  remains the same throughout but the lifted block-connectivity matrices  $B^{(t)}$ ,  $t \in [T]$  change at  $t = \tau$ :

$$B^{(t)} = \begin{cases} B_0 & \text{if } 1 \leq t \leq \tau, \\ B_1 & \text{if } \tau + 1 \leq t \leq T. \end{cases}$$

Also the expected adjacency matrix before (resp. after)  $\tau$  can be written as  $P_0 = Z B_0 Z^T$  (resp.  $P_1 = Z B_1 Z^T$ ).

In Figure 1, we depict an example with  $K_0 = 3$ ,  $K_1 = 4$ ,  $K = 5$  and  $n = 5m$  with  $m = 2$ . Take

$$\tilde{B}_0 = \begin{pmatrix} p & q & q \\ q & p & q \\ q & q & p \end{pmatrix}; \quad \tilde{B}_1 = \begin{pmatrix} p & q & q & q \\ q & p & q & q \\ q & q & p & q \\ q & q & q & p \end{pmatrix};$$

$$\tilde{Z}_0 = \begin{pmatrix} \mathbf{1}_{3m} e_1^\top \\ \mathbf{1}_m e_2^\top \\ \mathbf{1}_m e_3^\top \end{pmatrix}; \quad \tilde{Z}_1 = \begin{pmatrix} \mathbf{1}_m e_1^\top \\ \mathbf{1}_m e_2^\top \\ \mathbf{1}_m e_3^\top \\ \mathbf{1}_{2m} e_4^\top \end{pmatrix}.$$

We can then lift the community structures to a common  $Z$  (cf. Figure 1) as follows. Take

$$B_0 = \begin{pmatrix} p & p & p & q & q \\ p & p & p & q & q \\ p & p & p & q & q \\ q & q & q & p & q \\ q & q & q & q & p \end{pmatrix}; \quad B_1 = \begin{pmatrix} p & q & q & q & q \\ q & p & q & q & q \\ q & q & p & q & q \\ q & q & q & p & p \\ q & q & q & p & p \end{pmatrix};$$

$$Z = \begin{pmatrix} \mathbf{1}_m e_1^\top \\ \mathbf{1}_m e_2^\top \\ \mathbf{1}_m e_3^\top \\ \mathbf{1}_m e_4^\top \\ \mathbf{1}_m e_5^\top \end{pmatrix}.$$

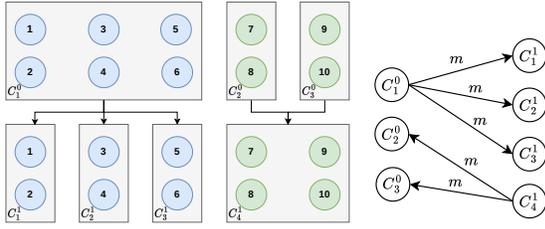


Figure 1: An example of simultaneous merging and splitting of communities. There are 10 nodes with three communities  $C_1^0, C_2^0, C_3^0$  before the changepoint  $\tau$  and four communities  $C_1^1, C_2^1, C_3^1, C_4^1$  after the changepoint  $\tau$ . The figure on the right panel shows the corresponding directed bipartite graph  $\mathcal{G}$  (here  $m = 2$ ).

It is then clear that  $ZB_0Z^\top = \tilde{Z}_0\tilde{B}_0\tilde{Z}_0^\top$  and  $ZB_1Z^\top = \tilde{Z}_1\tilde{B}_1\tilde{Z}_1^\top$ .

The primary goal of this article is to estimate the location of the changepoint  $\tau$  consistently. The complexity of the changepoint detection problem depends on the following quantities:

- i) The *sparsity* of the networks, given by  $d := n\rho$ , with

$$\rho := \|B_0\|_\infty \vee \|B_1\|_\infty,$$

where  $\|M\|_\infty := \max_{i,j} |M_{ij}|$ .

- ii) The *signal strength*, which is a measure of the difference between  $B_0$  and  $B_1$ . In this article, we take the Frobenius norm  $\|B_0 - B_1\|_F$  as the signal strength.
- iii) The *cushion*  $\kappa = \min\{\tau, T - \tau\}$ , the minimum length of the segments without change.

## 2.2 Our Contributions

Under Assumptions 3.1 and 4.2, we obtain the following.

1. A lower bound for the changepoint detectability threshold (in the minimax sense) in terms of the signal strength  $\|B_0 - B_1\|_F$ , which is given by

$$\theta_0 = \theta_0(n, T, K, d, \kappa) := \sqrt{\left(\frac{K}{n}\right)^3 \frac{d}{\kappa}}.$$

2. An upper bound for the changepoint detectability threshold by devising a polynomial-time changepoint estimation algorithm. For any  $\Lambda \in [T]$ , our algorithm can detect the changepoint consistently and give a high-probability confidence interval of length  $\Lambda$  for  $\tau$ , provided

$$(a) \|B_0 - B_1\|_F \gg \sqrt{K \log(KL)} \sqrt{\left(\frac{K}{n}\right)^3 \frac{d}{\Lambda \wedge \kappa}},$$

where  $L = \lceil \frac{3T}{\Lambda} \rceil - 2$ .

$$(b) Td\lambda \gg 1, \text{ where } \lambda := \rho^{-1} \lambda_{\min}(\bar{B}) \text{ is the sparsity-normalized smallest eigenvalue of the average connection probability matrix } \bar{B} := \frac{\tau}{T} B_0 + \left(1 - \frac{\tau}{T}\right) B_1 \text{ of the pre- and post-changepoint block-connectivity matrices } B_0 \text{ and } B_1.$$

*Remark 2.2.* Note that if  $\max\{K_0, K_1\}$  (and therefore  $K$ ) is at most poly-logarithmic in  $\max\{n, T\}$ , then our lower and upper bounds match up to logarithmic factors.

## 2.3 Summary of Related Theoretical Results

We now summarize, using the notation of the present paper, the assumptions required in previous works in order to ensure that the methods proposed therein are consistent.

- [Mukherjee \(2018\)](#) requires that either  $d \gg \sqrt{\log T}$  and  $\|B_0 - B_1\|_F \gg n\sqrt{\frac{T}{\kappa K}}\theta_0$ , or  $d \gg T(\log(nT))^2$  and  $\|B_0 - B_1\|_{\text{op}} \gg \sqrt{\frac{nT}{\kappa K}}\theta_0$ .
- [Bhattacharjee et al. \(2020\)](#) require that (see pages 6, 7, and 10 of their paper)  $d \gg n^{3/5}, K \ll d^{1/3}, \|B_0 - B_1\|_F \gg \sqrt{\frac{KT}{\kappa}}\theta_0, T = O(\frac{d^2}{K^4})$ , and that the minimum between the smallest eigenvalues of  $B_0$  and  $B_1$  is positive.
- [Wang et al. \(2021\)](#) require that (see their Assumptions 1 and 4)  $d \gg \log n, \|B_0 - B_1\|_F \gg \sqrt{n} \log^{1+\epsilon}(T) \theta_0$  for some  $\epsilon > 0$ , and some additional restrictions (see their Assumption 5) on the diagonal and off-diagonal entries of  $B_0$  and  $B_1$ . Furthermore, the algorithm in [Wang et al. \(2021\)](#) requires two independent copies of the network time series, which limits its usefulness.
- [Zhao et al. \(2019\)](#) require  $d \geq cn$  for some constant  $c > 0, \|B_0 - B_1\|_F \gg (n^3 T / K^2)^{1/4} \text{polylog}(n) \theta_0$  and  $T = O(\text{poly}(n))$ .
- [Bhattacharyya et al. \(2020\)](#) require that  $\|B_0 - B_1\|_F \gg \sqrt{\frac{n}{K}} \theta_0$ .

Although [Bhattacharjee et al. \(2020\)](#) is better in terms of signal strength by a factor  $\sqrt{\log K}$ , they require  $d \gg n^{3/5}$  for consistency, whereas we only require  $d \gg \frac{1}{T\lambda}$  (see Proposition 4.5). Thus we can operate in a much broader regime of sparsity. Also, our method requires a significantly smaller signal strength compared to that required by the global changepoint estimation algorithm of [Bhattacharyya et al. \(2020\)](#), provided  $K^2 \log K \ll n$ , which is a rather mild assumption. We empirically demonstrate these advantages in Section 5.1.

### 3 PROPOSED METHODOLOGY

#### 3.1 Estimation of the Community Membership Matrix

In the context of MSBM, many estimators of the community membership matrix  $Z$  are readily available in the literature, see, e.g., [Bhattacharyya and Chatterjee \(2020a\)](#). Our algorithm works with any such estimate  $\widehat{Z}$  of  $Z$ . For our theoretical results, we need the proportion of nodes misclassified by  $\widehat{Z}$  to be small with high probability.

**Assumption 3.1.** Let  $\mathcal{M}(\widehat{Z})$  denote the number of nodes misclassified by the estimator  $\widehat{Z}$ . Then, for any  $\epsilon > 0$ ,

$$\mathbb{P}(\mathcal{M}(\widehat{Z}) \leq \epsilon n) \geq 1 - \mathcal{P}(\epsilon, K, n, d, T), \quad (1)$$

where  $\mathcal{P}(\epsilon, K, n, d, T)$  is a probability function of  $\epsilon, K, n, d$  and  $T$  which depends on the algorithm used to obtain  $\widehat{Z}$  and is decreasing in  $\epsilon$ .

See [Proposition 4.5](#) for an illustration of such an estimator. For  $m \in [K]$ , let  $\widehat{C}_m = \{k : \widehat{Z}_{km} = 1\}$  be the  $m$ -th community estimated using  $\widehat{Z}$  and  $\widehat{n}_m = |\widehat{C}_m|$  be its size.

#### 3.2 Estimation of the Block-Connectivity Matrices

For  $k, l \in [K]$ , we estimate the  $(k, l)$ -th entry of the block-connectivity matrix  $B^{(t)}$  of the  $t$ -th layer by

$$\widehat{B}_{kl}^{(t)} = \frac{1}{\widehat{n}_{kl}^2} \sum_{\substack{i \in \widehat{C}_k \\ j \in \widehat{C}_l \\ i < j}} A_{ij}^{(t)}, \text{ where } \widehat{n}_{kl}^2 = \begin{cases} \binom{\widehat{n}_k}{2} & \text{if } k = l, \\ \widehat{n}_k \widehat{n}_l & \text{if } k \neq l. \end{cases} \quad (2)$$

Note that if  $\widehat{Z} \equiv Z$ , then  $\widehat{B}_{kl}^{(t)}$  is the Maximum Likelihood Estimate (MLE) of  $B_{kl}^{(t)}$ .

#### 3.3 Estimation of the Changepoint

First, we describe an oracle version of our changepoint estimation algorithm. Given the oracle parameter  $\Lambda$ , we divide the entire sequence  $[T]$  into multiple smaller windows of length  $\Lambda$  and search for a candidate for the changepoint estimate in each of these windows. The collection of smaller windows consists of intervals of the form  $\mathcal{I}_l = \mathcal{I}_l(\Lambda) := (T_l, T_l + \Lambda]$ , where

$$T_l = T_l(\Lambda) := (l-1) \left\lceil \frac{\Lambda}{3} \right\rceil \quad (3)$$

$$\text{for } l = 1, \dots, L = L(\Lambda) := \left\lceil \frac{3T}{\Lambda} \right\rceil - 2.$$

For each  $l \in [L]$ , we use the estimated block-connectivity matrices  $(\widehat{B}^{(T_l+v)})_{v=1}^\Lambda$  to construct the cumulative sum

(CUSUM) statistics corresponding to window  $\mathcal{I}_l$ , given by,

$$\widehat{G}_l^{(T_l+u)} = \widehat{G}_l^{(T_l+u)}(\Lambda) := \left( \frac{u}{\Lambda} \left( 1 - \frac{u}{\Lambda} \right) \right)^\delta \quad (4)$$

$$\times \left( \frac{1}{u} \sum_{v=1}^u \widehat{B}^{(T_l+v)} - \frac{1}{\Lambda - u} \sum_{v=u+1}^\Lambda \widehat{B}^{(T_l+v)} \right),$$

$$\text{for } u \in S(\Lambda) := \left[ \left\lceil \frac{\Lambda}{3} \wedge \frac{\kappa}{2} \right\rceil + 1, \dots, \Lambda - \left\lceil \frac{\Lambda}{3} \wedge \frac{\kappa}{2} \right\rceil \right],$$

where  $\delta \in [0, 1]$  is an input of our algorithm. In each window  $\mathcal{I}_l$ , we maximize the Frobenius norm of the corresponding CUSUM statistic w.r.t.  $u \in S(\Lambda)$  to obtain a candidate changepoint estimate  $\tau_l$ , given by

$$\tau_l = \tau_l(\Lambda) := T_l + \operatorname{argmax}_{u \in S(\Lambda)} \left\| \widehat{G}_l^{(T_l+u)} \right\|_F.$$

If the maximum of Frobenius norm of the CUSUM statistics evaluated at these candidates exceeds a threshold

$$\max_{l \in [L]} \left\| \widehat{G}_l^{(\tau_l)} \right\|_F > C_s \sqrt{\left( \frac{K}{n} \right)^3 \frac{\widehat{d}}{\frac{\Lambda}{3} \wedge \frac{\kappa}{2}}},$$

we set the maximizing candidate as the changepoint estimate

$$l(\Lambda) = \operatorname{argmax}_{l \in [L]} \left\| \widehat{G}_l^{(\tau_l)} \right\|_F \text{ and } \widehat{\tau}(\Lambda) = \tau_{l(\Lambda)}.$$

In the above threshold,  $C_s$  is an input of our algorithm, which controls the type-I and type-II errors of the estimate (see [Theorem 4.4](#)), and  $\widehat{d}$  is given by

$$\widehat{d} = \frac{1}{nT} \sum_{t \in [T]} \sum_{i, j \in [n]} A_{ij}^{(t)}.$$

In practice,  $S(\Lambda)$  is unknown (as  $\kappa$  is unknown) and we need to use a proxy for that. We can simply work with

$$\widetilde{S}(\Lambda) = \left[ \left\lceil \frac{\Lambda}{3} \right\rceil + 1, \dots, \Lambda - \left\lceil \frac{\Lambda}{3} \right\rceil \right],$$

which equals  $S(\Lambda)$  provided  $\Lambda \leq 3\kappa/2$ . (this is the reason the choice of  $\Lambda$  is oracle). So, in practice, the length of the smaller windows  $\bar{\Lambda}$  is a tuning parameter of our estimation algorithm, which we would like to be as minimum as possible. Our algorithm repeats the entire above procedure, replacing  $\Lambda$  by  $\bar{\Lambda}$  and  $S(\Lambda)$  by  $\widetilde{S}(\bar{\Lambda})$ , for different  $\bar{\Lambda}$ , starting from  $\bar{\Lambda} = T$  to  $\bar{\Lambda} = \Lambda_{\min}$  (an input of our algorithm). Let  $\bar{\Lambda}_0$  be the minimum  $\bar{\Lambda}$  for which a  $\widehat{\tau}(\bar{\Lambda})$  is obtained. We set  $\widehat{\tau} := \widehat{\tau}(\bar{\Lambda}_0)$  with

$$\left[ (l(\bar{\Lambda}_0) - 1) \left\lceil \frac{\bar{\Lambda}_0}{3} \right\rceil + 1, (l(\bar{\Lambda}_0) - 1) \left\lceil \frac{\bar{\Lambda}_0}{3} \right\rceil + \bar{\Lambda}_0 \right]$$

as a confidence interval for  $\tau$ . Note that we may not get such a  $\bar{\Lambda}_0$  in which case our algorithm gives no output. The entire estimation procedure is described in [Algorithm 1](#).

**Algorithm 1:** Estimation Procedure

**Input :**  $(A^{(t)})_{t \in [T]}$ ,  $\widehat{Z}_{n \times K}$ ,  $\delta \in [0, 1]$ ,  $C_s$ ,  $\Lambda_{\min}$ .

**Output :**  $\bar{\Lambda}_0$ ,  $l(\bar{\Lambda}_0)$ ,  $\widehat{\tau}$ .

1. For all  $m \in [K]$ , set

$$\widehat{C}_m = \{k : \widehat{Z}_{km} = 1\}, \quad \widehat{n}_m = |\widehat{C}_m|.$$

2. For all  $t \in [T]$  and  $k, l \in [K]$ , obtain  $\widehat{B}_{kl}^{(t)}$ , given by (2).

3. Set  $\widehat{d} = \frac{1}{nT} \sum_{t \in [T]} \sum_{i, j \in [n]} A_{ij}^{(t)}$ .

4. For each  $\bar{\Lambda} = T, \dots, \Lambda_{\min}$ , repeat steps 5, 6, 7.

5. For  $l = 1, \dots, \lceil 3T/\bar{\Lambda} \rceil - 2$  and for  $u \in \widetilde{S}(\bar{\Lambda})$ , obtain  $T_l(\bar{\Lambda})$  and  $\widehat{G}_l^{(T_l+u)}(\bar{\Lambda})$ , given by (3) and (4).

6. For all  $l = 1, \dots, L(\bar{\Lambda}) := \lceil 3T/\bar{\Lambda} \rceil - 2$ , set

$$\tau_l = T_l(\bar{\Lambda}) + \operatorname{argmax}_{u \in \widetilde{S}(\bar{\Lambda})} \|\widehat{G}_l^{(T_l+u)}(\bar{\Lambda})\|_F.$$

7. If  $\max_{l \in [L(\bar{\Lambda})]} \|\widehat{G}_l^{(\tau_l)}(\bar{\Lambda})\|_F > C_s \sqrt{\left(\frac{K}{n}\right)^3 \widehat{d}}$ , set

$$l(\bar{\Lambda}) = \operatorname{argmax}_{l \in [L(\bar{\Lambda})]} \|\widehat{G}_l^{(\tau_l)}(\bar{\Lambda})\|_F \text{ and } \widehat{\tau}(\bar{\Lambda}) = \tau_{l(\bar{\Lambda})}.$$

8. Let  $\bar{\Lambda}_0$  be the minimum  $\bar{\Lambda}$  for which  $\widehat{\tau}(\bar{\Lambda})$  is obtained. Declare  $\widehat{\tau} := \widehat{\tau}(\bar{\Lambda}_0)$  with

$$\left[ (l(\bar{\Lambda}_0) - 1) \left\lfloor \frac{\bar{\Lambda}_0}{3} \right\rfloor + 1, (l(\bar{\Lambda}_0) - 1) \left\lfloor \frac{\bar{\Lambda}_0}{3} \right\rfloor + \bar{\Lambda}_0 \right]$$

as a confidence interval for  $\tau$ .

Algorithm 1 is a polynomial (in  $n$  and  $T$ ) time algorithm. Step 1 takes  $O(nK)$  time. Steps 2 and 3 take  $O(n^2T)$  time. In each of the  $O(T)$  iterations of Step 4, for a fixed  $\Lambda$ , Steps 5 and 6 take  $O\left(\frac{T}{\Lambda} \cdot \Lambda\right) = O(T)$  time and Step 7 takes  $O\left(\frac{T}{\Lambda}\right)$  time. So, overall Step 4 takes  $O(T^2)$  time. Hence, Algorithm 1 has time-complexity  $O(T \max\{n^2, T\})$ .

*Remark 3.2.* In [Bhattacharyya et al. \(2020\)](#), the following CUSUM statistics, given by, for  $l \in [L(\Lambda)]$  and  $u \in S(\Lambda)$ ,

$$\widehat{F}_l^{(T_l+u)} = \widehat{F}_l^{(T_l+u)}(\Lambda) := \left( \frac{u}{\Lambda} \left( 1 - \frac{u}{\Lambda} \right) \right)^\delta \times \left( \frac{1}{u} \sum_{v=1}^u A^{(T_l+v)} - \frac{1}{\Lambda - u} \sum_{v=u+1}^{\Lambda} A^{(T_l+v)} \right), \quad (5)$$

is used. It does not take into account the community structure and so the corresponding changepoint estimate  $\widehat{\tau}$  requires higher signal strength, as mentioned in Section 2.3. In Section 5.1, we also illustrate this empirically.

### 3.4 Motivation behind the Estimation Algorithm

Fix any  $\Lambda \in [T]$ . For each  $l \in [L]$  and  $u \in [\Lambda]$ , consider the population version of the CUSUM statistic, given by

$$\mathbf{G}_l^{(T_l+u)} := \left( \frac{u}{\Lambda} \left( 1 - \frac{u}{\Lambda} \right) \right)^\delta \times \left( \frac{1}{t} \sum_{v=1}^u B^{(T_l+v)} - \frac{1}{\Lambda - u} \sum_{v=u+1}^{\Lambda} B^{(T_l+v)} \right).$$

Observe that

$$\left( \frac{u}{\Lambda} \left( 1 - \frac{u}{\Lambda} \right) \right)^{-\delta} \mathbf{G}_l^{(T_l+u)} = \begin{cases} 0 & \text{if } \tau \notin \mathcal{I}_l, \\ \frac{T_l + \Lambda - \tau}{\Lambda - u} (B_0 - B_1) & \text{if } \tau \in \mathcal{I}_l, u \leq \tau - T_l, \\ \frac{\tau - T_l}{u} (B_0 - B_1) & \text{if } \tau \in \mathcal{I}_l, u > \tau - T_l. \end{cases}$$

For any  $\delta \in [0, 1]$ , the function

$$u \mapsto \begin{cases} \left( \frac{u}{\Lambda} \left( 1 - \frac{u}{\Lambda} \right) \right)^\delta \frac{T_l + \Lambda - \tau}{\Lambda - u}, & u \leq \tau - T_l, \\ \left( \frac{u}{\Lambda} \left( 1 - \frac{u}{\Lambda} \right) \right)^\delta \frac{\tau - T_l}{u}, & u > \tau - T_l, \end{cases}$$

has a unique maxima at  $u = \tau - T_l$ . Thus for any  $l, l'$  such that  $\tau \in \mathcal{I}_l$  but  $\tau \notin \mathcal{I}_{l'}$ , we get

$$0 = \max_{u \in [\Lambda]} \|\mathbf{G}_{l'}^{(T_{l'}+u)}\|_F \leq \max_{u \in [\Lambda]} \|\mathbf{G}_l^{(T_l+u)}\|_F, \text{ and}$$

$$\tau = T_l + \operatorname{argmax}_{u \in [\Lambda]} \|\mathbf{G}_l^{(T_l+u)}\|_F.$$

This is the primary motivation for maximizing the CUSUM statistic to get the changepoint estimate.

*Remark 3.3.* The CUSUM statistic(s) (for any particular window) tends to get maximized near the boundary of the interval due to a lack of concentration. So, we omit  $2 \lfloor \frac{\Lambda}{3} \wedge \kappa \rfloor$ -many points from the boundary while maximizing them.

*Remark 3.4.* In the classical univariate changepoint problem, the factor  $\left( \frac{u}{\Lambda} \left( 1 - \frac{u}{\Lambda} \right) \right)^\delta$ ,  $\delta \neq 0$  plays an important role in controlling the type-I and type-II errors for detection of changepoint and also the estimation error. See, for example, Chapter 3 of [Brodsky and Darkhovsky \(2013\)](#). Also, it is empirically beneficial as it curbs down the tendency of the CUSUM statistic to get maximized near the boundary of the intervals, as stated in Remark 3.3.

## 4 MAIN RESULTS

We state and discuss our main theoretical results in this section. The proofs are deferred to the supplementary material.

#### 4.1 Lower Bound

We begin with a lower bound result that shows that if  $\|B_0 - B_1\|_F = O\left(\sqrt{\left(\frac{K}{n}\right)^3 \frac{d}{\kappa}}\right)$ , then it is not possible to estimate  $\tau$  consistently.

Let  $\mathbb{P}_B^t = \otimes_{s=1}^t \mathbb{P}_B$  denote the joint distribution of  $t$  networks generated from an MSBM having  $t$  layers,  $n$  nodes and  $B$  as the common  $K \times K$  connectivity matrix for each of the  $t$  layers. Let  $\mathcal{M}_K$  be the set of all symmetric  $K \times K$  matrices with entries in  $[0, 1]$ . For  $B_0, B_1 \in \mathcal{M}_K$ , let

$$\mathbb{P}_{\tau, B_0, B_1}^T := \otimes_{t=1}^{\tau} \mathbb{P}_{B_0} \otimes_{t=\tau+1}^T \mathbb{P}_{B_1}$$

denote the joint distribution of  $T$  networks generated from an MSBM with  $T$  layers,  $n$  nodes and the  $K \times K$  block connection probability matrix changes from  $B_0$  to  $B_1$  at the changepoint  $\tau$ . For any two distributions  $\nu_0$  and  $\nu_1$  on  $\mathcal{M}_K$ , let

$$\mathbb{P}_{\tau, \nu_0, \nu_1}^T(\cdot) := \mathbb{E}_{\nu_0, \nu_1} \mathbb{P}_{\tau, B_0, B_1}^T(\cdot).$$

For  $\kappa \in [T]$ , and  $\gamma > 0$ , let

$$\begin{aligned} \mathcal{P}_{\gamma, \kappa} := & \left\{ \mathbb{P}_{\tau, \nu_0, \nu_1}^T : \right. \\ & (a) \nu_0 \text{ and } \nu_1 \text{ are distributions on } \mathcal{M}_K, \\ & (b) (B_0, B_1) \sim (\nu_0, \nu_1) \text{ satisfies} \\ & \quad \|B_0 - B_1\|_F^2 \leq \gamma \left(\frac{K}{n}\right)^3 \frac{d}{\kappa}, \\ & \quad \text{with } d = n(\|B_0\|_\infty \vee \|B_1\|_\infty), \\ & (c) \kappa \leq \tau \leq T - \kappa \left. \right\}. \end{aligned}$$

We are now ready to state our lower bound result.

**Theorem 4.1 (Lower Bound).** *For  $\mathbb{P}_0 \in \mathcal{P}_{\gamma, \kappa}$ , let  $\tau(\mathbb{P}_0)$  denote the changepoint under the probability distribution  $\mathbb{P}_0$ . For any  $\varepsilon > 0$ , there is a constant  $\gamma(\varepsilon) > 0$  such that for any  $r \in [n]$  and  $\kappa \leq T/4$ ,*

$$\inf_{\hat{\tau}} \sup_{\mathbb{P}_0 \in \mathcal{P}_{\gamma, \kappa}} \mathbb{E}_0 |\hat{\tau} - \tau(\mathbb{P}_0)| \geq \frac{T}{3}(1 - \varepsilon).$$

This means, in the class  $\mathcal{P}_{\gamma, \kappa}$  of models with signal strength of the order of  $\left(\frac{K}{n}\right)^{3/2} \left(\frac{d}{\kappa}\right)^{1/2}$ , there are models for which no algorithm can estimate the changepoint consistently.

#### 4.2 Upper Bound

In the other direction, we establish that if the signal strength

$$\|B_0 - B_1\|_F \gg \sqrt{K \log(KL)} \theta_1,$$

where  $\theta_1 := \sqrt{\left(\frac{K}{n}\right)^3 \frac{d}{\Lambda \wedge \kappa}}$ , then the oracle algorithm estimates  $\tau$  consistently. It may be noted that this is almost

optimal (up to the dependence on  $K$ ) The main reasons behind the improvement over prior works are two-fold. First, we are estimating the underlying community structure using averaged adjacency matrices (see Proposition 4.5). Optimal rates for community estimation using this method are available in Bhattacharyya and Chatterjee (2020b). Further, our method uses CUSUM statistics based on estimated block-connectivity matrices, as opposed to adjacency matrices, thus incorporating the low-rank nature of the edge-probability matrices for networks with community structure.

The following assumptions are standard in the literature.

**Assumption 4.2.** There exist constants  $C_1, C_2$  with  $C_1 \geq C_2 \vee 1$  such that  $C_2 \left(\frac{n}{K}\right) \leq n_m \leq C_1 \left(\frac{n}{K}\right)$  for all  $m \in [K]$ .

**Assumption 4.3.** There exists a constant  $0 < c_0 < 1$  such that the average degree of the networks  $\bar{d}$  satisfy

$$\bar{d} := \frac{1}{nT} \sum_{t \in [T]} \sum_{i, j \in [n]} \mathbb{E} \left( A_{ij}^{(t)} \right) = c_0 d.$$

We now state our upper bound result.

**Theorem 4.4 (Upper Bound).** *Assume that the Assumptions 3.1, 4.2 are satisfied and let*

$$w := \theta_1^{-1} \|B_0 - B_1\|_F \left( \frac{\Lambda \wedge \kappa}{3\Lambda} \left( 1 - \frac{\Lambda \wedge \kappa}{3\Lambda} \right) \right)^\delta.$$

*Then there exist absolute constants  $C_1, C_2, C_3, C_4 > 0$  such that the following results hold.*

i) (Type-I error bound) *Assume that Assumption 4.3 holds,  $B_0 = B_1$  and  $\Delta_1 = C_s \sqrt{c_0}/2$ . Then the algorithm produces an output  $\hat{\tau}$  with probability at most*

$$\begin{aligned} & C_1 K^2 L \exp \left( - \frac{C_2 \Delta_1^2}{K} \right) \\ & + C_3 \mathcal{P} \left( \frac{C_4 \Delta_1 \theta_1}{K^2}, K, n, d, T \right) + \exp \left( - \frac{c_0^2}{16} n T d \right). \end{aligned}$$

ii) (Type-II error bound) *Assume that the Assumption 4.3 is satisfied and  $\Delta_2 := w - 3C_s \sqrt{3c_0}/2 > 0$ . Then the algorithm produces no output with probability at most*

$$\begin{aligned} & C_1 K^2 L \exp \left( - \frac{C_2 \Delta_2^2}{K} \right) \\ & + C_3 \mathcal{P} \left( \frac{C_4 \Delta_2 \theta_1}{K^2}, K, n, d, T \right) + \exp \left( - \frac{c_0^2}{16} n T d \right). \end{aligned}$$

iii) (Localization bound) *If Algorithm 1 gives an output  $\hat{\tau}$ , then for all  $\eta \in (0, 1)$  and  $\epsilon_0 = (C_4 \eta w \theta_1)/K^2$ ,*

$$\begin{aligned} \mathbb{P} \left( \frac{1}{\Lambda} |\hat{\tau} - \tau| \geq \eta \right) & \leq C_1 K^2 L \exp \left\{ - \frac{C_2 \eta^2 w^2}{K} \right\} \\ & + C_3 \mathcal{P}(\epsilon_0, K, n, d, T). \quad (6) \end{aligned}$$

We now provide a concrete example of the probability function  $\mathcal{P}$  considered in (1) and describe the localization bound of  $\hat{\tau}$  obtained using it.

**Proposition 4.5.** *Assume that  $K = \text{rank}(\frac{\tau}{T}B_0 + \frac{T-\tau}{T}B_1)$  and that  $\hat{Z}$  be obtained using spectral clustering of the sum of the adjacency matrices (Algorithm 1 of [Bhattacharyya and Chatterjee \(2020a\)](#)). Let  $\lambda$  be the smallest eigenvalue of the average of normalized connectivity matrices and  $\lambda = \lambda_K(\frac{n}{d}(\frac{\tau}{T}B_0 + \frac{T-\tau}{T}B_1)) > 0$ . By Theorem 1 of [Bhattacharyya and Chatterjee \(2020a\)](#), there are constants  $c_1, c_2, c_3 > 0$  such that if  $Td \geq c_2(K/\lambda)^{9/8}$ ,  $n \geq 3K$  and  $n_{\min} := \text{smallest community size} > 2/\lambda$ , then*

$$\begin{aligned} \mathbb{P}\left(\mathcal{M}(\hat{Z}) \geq c_1 \left(\frac{K^{12}}{Td\lambda^8}\right)^{1/4} n\right) \\ \leq 5 \exp\left(-c_3 \min\left\{Td\lambda, (Td)^{1/2} \log n\right\}\right). \end{aligned}$$

This result combined with (6) gives the following localization bound for  $\hat{\tau}$ . Assume that the Assumptions 3.1, 4.2, 4.3 are satisfied and  $w, C_1, C_2, C_3, C_4$  are as in Theorem 4.4. Then, for all  $\eta \in (0, 1)$ , as long as  $(Td)^{1/4} \geq (c_1 K^5)/(C_4 \eta w \theta_1)$ , we have

$$\begin{aligned} \mathbb{P}\left(\frac{1}{\Lambda} |\hat{\tau} - \tau| \geq \eta\right) = C_1 K^2 L \exp\left\{-\frac{C_2 \eta^2 w^2}{K}\right\} \\ + 5C_3 \exp\left\{-c_3 \min\left\{Td\lambda, (Td)^{1/2} \log n\right\}\right\}. \end{aligned} \quad (7)$$

*Remark 4.6.* From (7), it follows that if  $\|B_0 - B_1\|_F \gg \sqrt{K \log(KL)}\theta_1$  and  $Td\lambda \gg 1$ , then  $\hat{\tau}$  is consistent.

*Remark 4.7.* In Algorithm 1 of [Bhattacharyya and Chatterjee \(2020a\)](#),  $K$  is considered to be known. However,  $K$  can be estimated based on the eigenvalues of the sum of squared adjacency matrices. See Algorithm 3 of [Bhattacharyya and Chatterjee \(2020c\)](#) and the discussions there.

## 5 EMPIRICAL RESULTS

All the analyses reported here were performed on Kaggle notebooks (without any additional accelerator) using R. All the figures and tables can be reproduced using the codes available at <https://gitlab.com/soumendu041/cpd-sbm>.

### 5.1 Simulation Experiments

Our simulation studies involve two experiments. Experiment I empirically demonstrates the performance of  $\hat{\tau}$  by comparing it with  $\tilde{\tau}$  (see Remark 3.2), which is minimax optimal in unstructured networks changepoint problems in all regimes of sparsity. This experiment is performed under four setups, designated as (a), (b), (c) and (d). In Experiment II, we compare the performance of our algorithm with the methods of [Bhattacharjee et al. \(2020\)](#), [Wang et al.](#)

Table 1: Experiment I: Mean Relative Error

Setup	(a)	(b)	(c)	(d)
$ \hat{\tau} - \tau /\bar{\Lambda}_0$	0.008	0.005	0.009	0.021
$ \tilde{\tau} - \tau /\bar{\Lambda}_0$	0.447	0.513	0.306	0.433

Table 2: Experiment I: Empirical Coverage Probability

Setup	(a)	(b)	(c)	(d)
ECP	1.0	1.0	1.0	0.9
ECP	0.2	0.0	0.2	0.2

(2021), [Zhao et al. \(2019\)](#) and [Bhattacharyya et al. \(2020\)](#) across varying level of sparsity. We perform Experiment II also under four setups, designated again as (a), (b), (c) and (d).

In setup (a) of both experiments, there is one community initially that splits into two communities at the changepoint (so,  $K_0 = 1, K_1 = 2, K = 2$ ). In setup (b) of both experiments, there are two communities throughout and only the connectivity probability matrix changes at the changepoint (so,  $K_0 = K_1 = K = 2$ ). In setup (c) of both experiments, there are two communities initially that merge into one community at the changepoint (so,  $K_0 = 2, K_1 = 1, K = 2$ ). In setup (d) of both experiments, there are three communities at first. At the changepoint, one of them splits into three communities and the other two communities merge into one, as in Figure 1 (so,  $K_0 = 3, K_1 = 4, K = 5$ ).

For our algorithm, we take  $\delta = 0.5$ ,  $C_s = 6$  and  $\Lambda_{\min} = 100$ .  $\hat{Z}$  is obtained using spectral clustering of the sum of adjacency matrices with  $K = 2$ . For each experiment and setup, we perform 10 Monte-Carlo (MC) runs. For other competing methods, the default values of the tuning parameters are used.

In setups (a), (b), (c) and (d) of Experiment I, we take  $d = 10, 12, 5$  and 10 respectively and  $\|B_0 - B_1\|_F = 0.028, 0.021, 0.014$  and 0.057 respectively. For both  $\hat{\tau}$ ,  $\tilde{\tau}$  and for each MC run in each setup, we get  $\bar{\Lambda}_0 = 100$  (note that  $L(\bar{\Lambda}_0) = 4$ ). The other outputs are summarized in Tables 1 and 2, where we report the mean (across 10 MC runs) relative errors ( $|\hat{\tau} - \tau|/\bar{\Lambda}_0$  and  $|\tilde{\tau} - \tau|/\bar{\Lambda}_0$ ) and empirical coverage probabilities (ECP) (fraction of times among the 10 MC runs where the output interval contains  $\tau$ , i.e.  $l(\bar{\Lambda}_0) = 3$ ), respectively. In Figure 2, we illustrate the behavior of the CUSUM statistics, corresponding to both  $\hat{\tau}$  and  $\tilde{\tau}$ , in setup (b). In Figures 4, 5 and 6 of the Supplementary Material, the same is illustrated in setups (a), (c) and (d), respectively.

Each setup of Experiment II is performed under different values of  $d$ . We summarize the values of the parameter  $d$  and

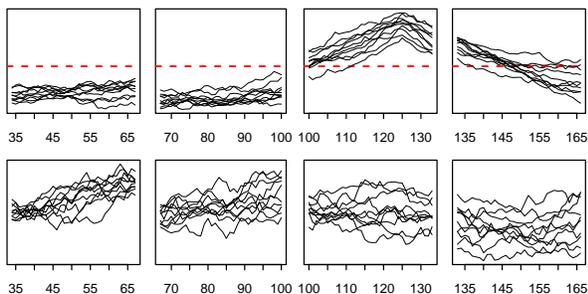


Figure 2: Experiment I-(b). Top row: Observed CUSUM statistics (4) (with  $l = 1, 2, 3, 4$  and  $\Lambda = 100$ ) for 10 MC runs. The dashed lines display the mean threshold value. Bottom row: Observed adjacency-based CUSUM statistics (5) (with  $l = 1, 2, 3, 4$  and  $\Lambda = 100$ ) for 10 MC runs.

the corresponding mean (across 10 MC runs) relative errors ( $|\hat{\tau}_{\text{est}} - \tau|/T$ , where  $\hat{\tau}_{\text{est}}$  stands for a generic estimate) in each setup for every method under comparison in Tables 3, 4, 5 and 6 of the Supplementary Material. In the case of Wang et al. (2021) and Zhao et al. (2019), while computing the mean relative error, if multiple changepoints are detected, we consider the one closest to the true changepoint and if no change point is detected, we consider the estimated changepoint to be  $T = 200$ .

## 5.2 A Real-World Example of Community Splitting

We use the US Senate roll call data (1979 - 2023) from Lewis et al. (2024) to illustrate the phenomenon of community splitting. From data on votes taken on time-stamped bills, we construct a network time-series. Senate seats constitute nodes. Each of the 50 states has 2 senate seats, so the networks all have size  $n = 100$ . The votes are “yay/nay”, and sometimes missing/present — such cases are imputed by taking the majority stand in the respective party, or if that is not possible, taking the winning majority stand. We discard roll calls wherein the majority  $\geq 75$ , because such unanimous roll calls do not reveal much structure. This leaves us with 9770 roll calls. We convert these roll calls into networks as follows. We consider 49 consecutive roll calls to construct a single network. In each network, we put an edge between two seats if they take the same stand on the proposed bill of at least 34 ( $\sim 70\%$ ) roll calls among the 49 consecutive roll calls under consideration. This results in  $T = 199$  networks. We plot in the left panel of Figure 3 the eigenvalue distributions of these networks. Typically two eigenvalues stick outside of the bulk spectrum, hinting at the presence of two communities. For each network  $A^{(t)}$ , we also plot  $\frac{\lambda_2(A^{(t)})^2}{\lambda_1(A^{(t)})}$  which is a measure of the strength of the community structure. We observe an overall increasing

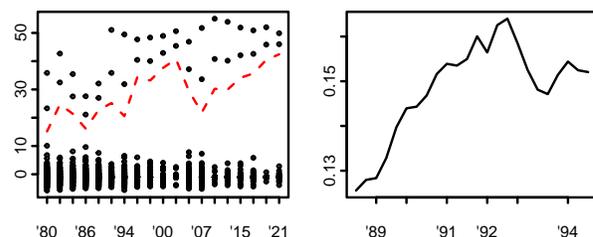


Figure 3: Left: Eigenvalues of the roll call networks across time. The dashed line shows the ratio of the squared second largest eigenvalue and the largest eigenvalue of these networks, a measure of the strength of the community structure. Right: Observed CUSUM statistics (4) (with  $l = 2$  and  $\Lambda = 70$ ) for the period 1988 – 1994. The peak corresponds to 49 consecutive roll calls from 1992/09/10 to 1993/02/04.

trend in this measure, which aligns with the increase in party polarization in US politics (Moody and Mucha (2013)).

We apply our changepoint estimation algorithm to this time-series, taking  $\delta = 0.5$ ,  $C_s = 6$  and  $\Lambda_{\min} = 70$ , and using spectral clustering of the sum of adjacency matrices with  $K = 2$  to obtain  $\hat{Z}$ . The outputs of our algorithm are  $\bar{\Lambda}_0 = 70$ ,  $l(\bar{\Lambda}_0) = 2$ ,  $\hat{\tau} = 62$ . The estimated changepoint at  $\hat{\tau} = 62$  corresponds to 49 roll calls from September 10, 1992 to February 04, 1993 (see the right panel of Figure 3). This is near the November 1994 election, when the Republican Party regained control of the House of Representatives for the first time since 1956. A sharp increase in party polarisation had been reported around this period in Moody and Mucha (2013).

Temporal variations in the eigenvalue distributions of these networks suggest the presence of multiple changepoints. In an analysis reported in Table 7 of the Supplementary Material, we combine our method for single changepoint estimation with wild binary segmentation (Fryzlewicz et al., 2014) to detect several potentially important changepoints in these roll call networks.

## 6 FUTURE DIRECTIONS

Although our method addresses many important challenges, it has some limitations. For instance, it is not minimax-optimal when  $K$  grows with  $n$ . Our method needs to be improvised when there are multiple changepoints, individuals may switch communities at changepoints, or there are correlations among the network layers. Further, it will be of interest to extend our methodology to the setting of online changepoints. We postpone the analysis of such extensions to future work, keeping in mind the space restriction of the current venue.

## Acknowledgments

The authors thank the anonymous reviewers for extensive and helpful comments on an earlier version of this manuscript. SSM was partially supported by the INSPIRE research grant DST/INSPIRE/04/2018/002193 from the Dept. of Science and Technology, Govt. of India, and a Start-Up Grant from Indian Statistical Institute. The research of S. C. was supported by NSF grant DMS-2154564. T. S. is partially supported by the Breakout Graduate Fellowship from International Mathematical Union (IMU).

## References

- Bhattacharjee, M., Banerjee, M., and Michailidis, G. (2020). Change point estimation in a dynamic stochastic block model. *The Journal of Machine Learning Research*, 21(1):4330–4388.
- Bhattacharyya, S. and Chatterjee, S. (2020a). Consistent recovery of communities from sparse multi-relational networks: A scalable algorithm with optimal recovery conditions. In Barbosa, H., Gomez-Gardenes, J., Gonçalves, B., Mangioni, G., Menezes, R., and Oliveira, M., editors, *Complex Networks XI*, pages 92–103, Cham. Springer International Publishing.
- Bhattacharyya, S. and Chatterjee, S. (2020b). General community detection with optimal recovery conditions for multi-relational sparse networks with dependent layers.
- Bhattacharyya, S. and Chatterjee, S. (2020c). General community detection with optimal recovery conditions for multi-relational sparse networks with dependent layers. *ArXiv*. <https://doi.org/10.48550/arXiv.2004.03480>.
- Bhattacharyya, S., Chatterjee, S., and Mukherjee, S. S. (2020). Consistent detection and optimal localization of all detectable change points in piecewise stationary arbitrarily sparse network-sequences. *arXiv preprint arXiv:2009.02112*.
- Blonder, B., Wey, T. W., Dornhaus, A., James, R., and Sih, A. (2012). Temporal dynamics and network analysis. *Methods in Ecology and Evolution*, 3(6):958–972.
- Brodsky, E. and Darkhovsky, B. S. (2013). *Nonparametric methods in change point problems*, volume 243. Springer Science & Business Media.
- Ferraz Costa, A., Yamaguchi, Y., Juci Machado Traina, A., Traina Jr, C., and Faloutsos, C. (2015). Rsc: Mining and modeling temporal activity in social media. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 269–278. ACM.
- Fryzlewicz, P. et al. (2014). Wild binary segmentation for multiple change-point detection. *The Annals of Statistics*, 42(6):2243–2281.
- Gates, M. C. and Woolhouse, M. E. (2015). Controlling infectious disease through the targeted manipulation of contact network structure. *Epidemics*, 12:11–19.
- Holme, P. (2015). Modern temporal network theory: a colloquium. *The European Physical Journal B*, 88(9):234.
- Holme, P. and Saramäki, J. (2012). Temporal networks. *Physics reports*, 519(3):97–125.
- Jacobs, A. Z., Way, S. F., Ugander, J., and Clauset, A. (2015). Assembling the facebook: Using heterogeneity to understand online social network assembly. In *Proceedings of the ACM Web Science Conference*, page 18. ACM.
- Krings, G., Karsai, M., Bernhardsson, S., Blondel, V. D., and Saramäki, J. (2012). Effects of time window size and placement on the structure of an aggregated communication network. *EPJ Data Science*, 1(1):4.
- Lahiri, M. and Berger-Wolf, T. Y. (2007). Structure prediction in temporal networks using frequent subgraphs. In *2007 IEEE Symposium on Computational Intelligence and Data Mining*, pages 35–42. IEEE.
- Lewis, J. B., Poole, K., Rosenthal, H., Boche, A., Rudkin, A., and Sonnet, L. (2024). Voteview: Congressional roll-call votes database.
- Moody, J. and Mucha, P. J. (2013). Portrait of political party polarization. *Network Science*, 1(1):119–121.
- Mukherjee, S. S. (2018). *On Some Inference Problems for Networks*. PhD thesis, University of California, Berkeley.
- Panisson, A., Gauvin, L., Barrat, A., and Cattuto, C. (2013). Fingerprinting temporal networks of close-range human proximity. In *2013 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, pages 261–266. IEEE.
- Park, H.-J. and Friston, K. (2013). Structural and functional brain networks: from connections to cognition. *Science*, 342(6158):1238411.
- Peixoto, T. P. (2015). Inferring the mesoscale structure of layered, edge-valued, and time-varying networks. *Physical Review E*, 92(4):042807.
- Peixoto, T. P. and Gauvin, L. (2018). Change points, memory and epidemic spreading in temporal networks. *Scientific reports*, 8(1):15511.
- Popović, M., Štefančić, H., Sluban, B., Novak, P. K., Grčar, M., Mozetič, I., Puliga, M., and Zlatič, V. (2014). Extraction of temporal networks from term co-occurrences in online textual sources. *PloS one*, 9(12):e99515.
- Rigbolt, K. T., Prokhorova, T. A., Akimov, V., Henningsen, J., Johansen, P. T., Kratchmarova, I., Kassem, M., Mann, M., Olsen, J. V., and Blagoev, B. (2011). System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. *Sci. Signal.*, 4(164):rs3–rs3.

- Rocha, L. E., Liljeros, F., and Holme, P. (2010). Information dynamics shape the sexual networks of internet-mediated prostitution. *Proceedings of the National Academy of Sciences*, 107(13):5706–5711.
- Rocha, L. E., Liljeros, F., and Holme, P. (2011). Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts. *PLoS computational biology*, 7(3):e1001109.
- Salathé, M., Kazandjieva, M., Lee, J. W., Levis, P., Feldman, M. W., and Jones, J. H. (2010). A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences*, 107(51):22020–22025.
- Sikdar, S., Ganguly, N., and Mukherjee, A. (2016). Time series analysis of temporal networks. *The European Physical Journal B*, 89(1):11.
- Sporns, O. (2013). Structure and function of complex brain networks. *Dialogues in clinical neuroscience*, 15(3):247.
- Stopczynski, A., Sekara, V., Sapiezynski, P., Cuttone, A., Madsen, M. M., Larsen, J. E., and Lehmann, S. (2014). Measuring large-scale social networks with high resolution. *PloS one*, 9(4):e95978.
- Tsybakov, A. B. (2008). *Introduction to nonparametric estimation*. Springer Science & Business Media.
- Wang, D., Yu, Y., and Rinaldo, A. (2021). Optimal change point detection and localization in sparse dynamic networks. *The Annals of Statistics*, 49(1):203–232.
- Zhang, X., Shao, S., Stanley, H. E., and Havlin, S. (2014). Dynamic motifs in socio-economic networks. *EPL (Europhysics Letters)*, 108(5):58001.
- Zhao, Z., Chen, L., and Lin, L. (2019). Change-point detection in dynamic networks via graphon estimation. *arXiv preprint arXiv:1908.01823*.
3. For all figures and tables that present empirical results, check if you include:
    - (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [**Yes**]
    - (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [**Not Applicable**]
    - (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [**Yes**]
    - (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [**Yes**]
  4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
    - (a) Citations of the creator If your work uses existing assets. [**Yes**]
    - (b) The license information of the assets, if applicable. [**Not Applicable**]
    - (c) New assets either in the supplemental material or as a URL, if applicable. [**Not Applicable**]
    - (d) Information about consent from data providers/curators. [**Not Applicable**]
    - (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [**Not Applicable**]
  5. If you used crowdsourcing or conducted research with human subjects, check if you include:
    - (a) The full text of instructions given to participants and screenshots. [**Not Applicable**]
    - (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [**Not Applicable**]
    - (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [**Not Applicable**]

## Checklist

1. For all models and algorithms presented, check if you include:
  - (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [**Yes**]
  - (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [**Yes**]
  - (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [**Yes**]
2. For any theoretical claim, check if you include:
  - (a) Statements of the full set of assumptions of all theoretical results. [**Yes**]
  - (b) Complete proofs of all theoretical results. [**Yes**]
  - (c) Clear explanations of any assumptions. [**Yes**]

---

## Supplementary Material

---

### A PROOFS OF THE MAIN RESULTS

In this section, we present the detailed proofs of our main results.

#### A.1 Proof of the Lower Bound

In this section, we prove Theorem 4.1.

Let  $Z$  be the community membership matrix of an equal-blocks SBM with  $K$  communities. Thus each block has size  $n/K$ . Let  $\Theta_0 = ZB_0Z^\top$  and  $\Theta_1 = ZB_1Z^\top$  where  $B_i = \rho S_i$ ,  $\min_{u,v}(S_i)_{uv} \asymp \max_{u,v}(S_i)_{uv}$ , and

$$B_1 - B_0 = \rho \alpha U U^\top, \quad (8)$$

where  $U$  is a  $K \times 1$  random  $\pm 1$  vector. We first choose  $B_0$ , and perturb it according to (8) to obtain  $B_1$ , ensuring that it is a probability matrix by choosing  $\alpha$  small enough.

Recall that the total variation distance  $d_{\text{TV}}(\mu, \nu)$  between two probability measures on  $\mathbb{R}$  is given by

$$d_{\text{TV}}(\mu, \nu) = \sup_{A \in \mathcal{B}(\mathbb{R})} |\mu(A) - \nu(A)|.$$

Further, if  $\mu$  is absolutely continuous with respect to  $\nu$ , then, the  $\chi^2$ -divergence  $\chi^2(\mu, \nu)$  between  $\mu$  and  $\nu$  is given by

$$\chi^2(\mu, \nu) = \int \left( \frac{d\mu}{d\nu} \right)^2 d\nu - 1.$$

It is a well-known fact that (see, e.g., Eq. (2.27) of Tsybakov (2008))

$$d_{\text{TV}}(\mu, \nu) \leq \sqrt{\chi^2(\mu, \nu)}. \quad (9)$$

In order to prove the lower bound, we need a bound on the total variation distance between  $\mathbb{P}_{B_0}^t$  and  $\mathbb{P}_{B_1}^t$ .

**Lemma A.1.** *Assume that  $\rho \leq \rho_0 < 1$ . Given any  $\epsilon \in (0, 1)$ , there exists  $c = c(\epsilon) > 0$ , such that if*

$$\|B_1 - B_0\|_F^2 \leq c(1 - \rho_0) \left( \frac{K}{n} \right)^3 \frac{d}{t},$$

then

$$d_{\text{TV}}(\mathbb{P}_{B_0}^t, \mathbb{E}_{B_1 \sim \nu_1} \mathbb{P}_{B_1}^t) \leq \epsilon.$$

The proof of Lemma A.1 is given in Section B.

*Proof of Theorem 4.1.* We will use the two point method of Le Cam. Fix  $\kappa \leq T/3$ . Consider the two distributions  $\mathbb{P}_{\kappa, \nu_1, \nu_0}^T$  and  $\mathbb{P}_{T-\kappa, \nu_0, \nu_1}^T$ . Here  $\nu_0 = \delta_{B_0}$  and  $\nu_1$  is the distribution of  $B_1$  as defined in (8), where  $\alpha$  is suitably chosen so that condition (b) in the definition of  $\mathcal{P}_{\gamma, \kappa}$  holds. Note that

$$|\tau(\mathbb{P}_{\kappa, \nu_1, \nu_0}^T) - \tau(\mathbb{P}_{T-\kappa, \nu_0, \nu_1}^T)| = T - 2\kappa.$$

We have by Le Cam's lemma, we have

$$\begin{aligned} \inf_{\hat{\tau}} \sup_{\mathbb{P}_0 \in \mathcal{P}_{\gamma, \kappa}} \mathbb{E}_0 |\hat{\tau} - \tau(\mathbb{P}_0)| &\geq (T - 2\kappa)(1 - d_{\text{TV}}(\mathbb{P}_{\kappa, \nu_1, \nu_0}^T, \mathbb{P}_{T-\kappa, \nu_0, \nu_1}^T)) \\ &\geq \frac{T}{3}(1 - d_{\text{TV}}(\mathbb{P}_{\kappa, \nu_1, \nu_0}^T, \mathbb{P}_{T-\kappa, \nu_0, \nu_1}^T)). \end{aligned}$$

Now using the fact that  $d_{\text{TV}}(\mu_1 \otimes \mu_2, \nu_1 \otimes \nu_2) \leq d_{\text{TV}}(\mu_1, \nu_1) + d_{\text{TV}}(\mu_2, \nu_2)$ , we get that (by breaking each product into three consecutive parts of size  $\kappa, (T - 2\kappa), \kappa$ )

$$d_{\text{TV}}(\mathbb{P}_{\kappa, \nu_1, \nu_0}^T, \mathbb{P}_{T-\kappa, \nu_0, \nu_1}^T) \leq 2d_{\text{TV}}(\mathbb{P}_{B_0}^\kappa, \mathbb{E}_{B_1 \sim \nu_1} \mathbb{P}_{B_1}^\kappa) \leq 2\epsilon,$$

where the last inequality follows from Lemma A.1, by suitably choosing  $\gamma = \gamma(\epsilon)$  depending on  $c = c(\epsilon)$ . All in all, we have shown that

$$\inf_{\hat{\tau}} \sup_{\mathbb{P}_0 \in \mathcal{P}_{\gamma, \kappa}} \mathbb{E}_0 |\hat{\tau} - \tau(\mathbb{P}_0)| \geq \frac{T}{3}(1 - 2\epsilon).$$

This completes the proof. ■

## A.2 Proof of the Upper Bound

In this section, we prove Theorem 4.4.

First, we briefly recall the oracle algorithm. Given the oracle parameter  $\Lambda$ , we divide the entire sequence  $[T]$  into multiple smaller windows of length  $\Lambda$ . The collection of smaller windows consists of intervals of the form  $\mathcal{I}_l = \mathcal{I}_l(\Lambda) := (T_l, T_l + \Lambda]$ , where

$$T_l = T_l(\Lambda) := (l-1) \left\lfloor \frac{\Lambda}{3} \right\rfloor \quad \text{for } l = 1, \dots, L = L(\Lambda) := \left\lceil \frac{3T}{\Lambda} \right\rceil - 2.$$

For each  $l \in [L]$ , we construct the cumulative sum (CUSUM) statistics corresponding to window  $\mathcal{I}_l$ , given by,

$$\begin{aligned} \hat{\mathbf{G}}_l^{(T_l+u)} = \hat{\mathbf{G}}_l^{(T_l+u)}(\Lambda) &:= \left( \frac{u}{\Lambda} \left( 1 - \frac{u}{\Lambda} \right) \right)^\delta \left( \frac{1}{u} \sum_{v=1}^u \hat{B}^{(T_l+v)} - \frac{1}{\Lambda-u} \sum_{v=u+1}^{\Lambda} \hat{B}^{(T_l+v)} \right), \\ \text{for } u \in S(\Lambda) &:= \left[ \left\lfloor \frac{\Lambda}{3} \wedge \frac{\kappa}{2} \right\rfloor + 1, \dots, \Lambda - \left\lfloor \frac{\Lambda}{3} \wedge \frac{\kappa}{2} \right\rfloor \right], \end{aligned}$$

where  $\delta \in [0, 1]$  is an input of our algorithm. In each window  $\mathcal{I}_l$ , we maximize the Frobenius norm of the corresponding CUSUM statistic w.r.t.  $u \in S(\Lambda)$  to obtain a candidate changepoint estimate  $\tau_l$ , given by

$$\tau_l = \tau_l(\Lambda) := T_l + \operatorname{argmax}_{u \in S(\Lambda)} \|\hat{\mathbf{G}}_l^{(T_l+u)}\|_F.$$

If the maximum of Frobenius norm of the CUSUM statistics evaluated at these candidates exceeds a threshold

$$\max_{l \in [L]} \|\hat{\mathbf{G}}_l^{(\tau_l)}\|_F > C_s \sqrt{\left( \frac{K}{n} \right)^3 \frac{\hat{d}}{\frac{\Lambda}{3} \wedge \frac{\kappa}{2}}},$$

we set the maximizing candidate as the changepoint estimate

$$l(\Lambda) = \operatorname{argmax}_{l \in [L]} \|\hat{\mathbf{G}}_l^{(\tau_l)}\|_F \quad \text{and} \quad \hat{\tau}(\Lambda) = \tau_{l(\Lambda)}.$$

In the above threshold,  $C_s$  is an input of our algorithm and  $\hat{d}$  is given by

$$\hat{d} = \frac{1}{nT} \sum_{t \in [T]} \sum_{i, j \in [n]} A_{ij}^{(t)}.$$

Subsequently, we shall omit the dependence of various quantities on  $\Lambda$  for ease of presentation.

We need a uniform concentration inequality for the deviation between the observed and population CUSUM statistics.

**Proposition A.2.** *Under Assumptions 3.1 and 4.2, there exist absolute constants  $\bar{C}_1, \bar{C}_2, \bar{C}_3, \bar{C}_4 > 0$  such that for all  $0 < \theta \leq (C_s \vee 1)K\rho$  and  $\epsilon_\theta = \bar{C}_4\theta/K^2$ ,*

$$\mathbb{P} \left( \max_{l \in [L]} \max_{u \in S(\Lambda)} \left\| \mathbf{G}_l^{(T_l+u)} - \hat{\mathbf{G}}_l^{(T_l+u)} \right\|_F > \theta \right) \leq \bar{C}_1 K^2 L \exp \left\{ -\frac{\bar{C}_2 \theta^2}{K \theta^2} \right\} + \bar{C}_3 \mathcal{P}(\epsilon_\theta, K, n, d, T).$$

The proof of Proposition A.2 is given in Section A.3.

We need a lower bound for the rate at which the population CUSUM statistic decays on both sides of  $\tau$ .

**Lemma A.3.** *Let  $l_0 \in [L]$  be such that  $\tau \in \mathcal{I}_{l_0}$ . Then for any  $\eta \in (0, 1)$  such that  $\tau \pm \eta\Lambda \in \mathcal{I}_{l_0}$ , we have*

$$\left\| \mathbf{G}_{l_0}^\tau \right\|_F - \left\| \mathbf{G}_{l_0}^{\tau \pm \eta\Lambda} \right\|_F \geq \frac{\eta}{1+\eta} \left( \frac{\tau - T_{l_0}}{\Lambda} \left( 1 - \frac{\tau - T_{l_0}}{\Lambda} \right) \right)^\delta \|B_0 - B_1\|_F.$$

We also need bounds on the rate of concentration of  $\hat{d}$  around  $d$ .

**Lemma A.4.** *Under Assumption 4.3,*

$$\mathbb{P}\left(\hat{d} \geq \frac{3c_0}{2}d\right) \leq \exp\left\{-\frac{c_0^2}{16}nTd\right\} \quad \text{and} \quad \mathbb{P}\left(\hat{d} \leq \frac{c_0}{2}d\right) \leq \exp\left\{-\frac{c_0^2}{16}nTd\right\}.$$

The proofs of Lemma A.3 and Lemma A.4 are given in Section B.

Finally define the collection  $\bar{\mathcal{I}} = \{\bar{\mathcal{I}}_l\}_{l=1}^L$  of intervals

$$\bar{\mathcal{I}}_l := \left[ T_l + \left\lfloor \frac{\Lambda}{3} \wedge \frac{\kappa}{2} \right\rfloor + 1, T_l + \Lambda - \left\lfloor \frac{\Lambda}{3} \wedge \frac{\kappa}{2} \right\rfloor \right].$$

*Proof of the type-I error bound of Theorem 4.4.* The algorithm produces an output only if

$$\max_{l \in [L]} \max_{u \in S(\Lambda)} \left\| \hat{\mathbf{G}}_l^{(T_l+u)} \right\|_F > C_s \sqrt{\left(\frac{K}{n}\right)^3 \frac{\hat{d}}{\frac{\Lambda}{3} \wedge \frac{\kappa}{2}}} > C_s \sqrt{\left(\frac{K}{n}\right)^3 \frac{\hat{d}}{\Lambda \wedge \kappa}}. \quad (10)$$

Let  $l_1$  be such that  $\hat{\tau} \in \bar{\mathcal{I}}_{l_1}$ . Observe that when  $B_0 = B_1$ ,  $\left\| \mathbf{G}_{l_1}^{(\hat{\tau})} \right\|_F = 0$  and (10) implies that

$$\left\| \hat{\mathbf{G}}_{l_1}^{(\hat{\tau})} \right\|_F > C_s \sqrt{\left(\frac{K}{n}\right)^3 \frac{\hat{d}}{\Lambda \wedge \kappa}}.$$

Thus the probability that the algorithm produces an output is at most

$$\begin{aligned} & \mathbb{P}\left(\left\| \hat{\mathbf{G}}_{l_1}^{(\hat{\tau})} \right\|_F - \left\| \mathbf{G}_{l_1}^{(\hat{\tau})} \right\|_F > C_s \sqrt{\left(\frac{K}{n}\right)^3 \frac{\hat{d}}{\Lambda \wedge \kappa}}\right) \\ & \leq \mathbb{P}\left(\max_{l \in [L]} \max_{u \in S(\Lambda)} \left\| \mathbf{G}_l^{(T_l+u)} - \hat{\mathbf{G}}_l^{(T_l+u)} \right\|_F \geq C_s \sqrt{\left(\frac{K}{n}\right)^3 \frac{\hat{d}}{\Lambda \wedge \kappa}}\right) \\ & \leq \mathbb{P}\left(\max_{l \in [L]} \max_{u \in S(\Lambda)} \left\| \mathbf{G}_l^{(T_l+u)} - \hat{\mathbf{G}}_l^{(T_l+u)} \right\|_F \geq \Delta_1 \theta_1\right) + \mathbb{P}\left(\hat{d} \leq \frac{c_0}{2}d\right). \end{aligned}$$

Then the desired result follows from Proposition A.2 and Lemma A.4. ■

*Proof of the type-II error bound of Theorem 4.4.* The algorithm gives no output only if

$$\max_{l \in [L]} \max_{u \in S(\Lambda)} \left\| \hat{\mathbf{G}}_l^{(T_l+u)} \right\|_F \leq C_s \sqrt{\left(\frac{K}{n}\right)^3 \frac{\hat{d}}{\frac{\Lambda}{3} \wedge \frac{\kappa}{2}}} \leq 3C_s \sqrt{\left(\frac{K}{n}\right)^3 \frac{\hat{d}}{\Lambda \wedge \kappa}} \quad (11)$$

Let  $l_0$  be such that  $\tau \in \bar{\mathcal{I}}_{l_0}$ . Observe that

$$\left\| \mathbf{G}_{l_0}^{(\tau)} \right\|_F = \left( \frac{\tau - T_{l_0}}{\Lambda} \left( 1 - \frac{\tau - T_{l_0}}{\Lambda} \right) \right)^\delta \|B_0 - B_1\|_F \geq \left( \frac{\Lambda \wedge \kappa}{3\Lambda} \left( 1 - \frac{\Lambda \wedge \kappa}{3\Lambda} \right) \right)^\delta \|B_0 - B_1\|_F,$$

and (11) implies that

$$\left\| \widehat{\mathbf{G}}_{l_0}^{(\tau)} \right\|_F \leq 3C_s \sqrt{\left(\frac{K}{n}\right)^3 \frac{\widehat{d}}{\Lambda \wedge \kappa}}.$$

Thus the probability that the algorithm gives no output is at most

$$\begin{aligned} & \mathbb{P}\left(\left\| \mathbf{G}_{l_0}^{(\tau)} \right\|_F - \left\| \widehat{\mathbf{G}}_{l_0}^{(\tau)} \right\|_F > \left(\frac{\Lambda \wedge \kappa}{3\Lambda} \left(1 - \frac{\Lambda \wedge \kappa}{3\Lambda}\right)\right)^\delta \|B_0 - B_1\|_F - 3C_s \sqrt{\left(\frac{K}{n}\right)^3 \frac{\widehat{d}}{\Lambda \wedge \kappa}}\right) \\ & \leq \mathbb{P}\left(\max_{l \in [L]} \max_{u \in S(\Lambda)} \left\| \mathbf{G}_l^{(T_l+u)} - \widehat{\mathbf{G}}_l^{(T_l+u)} \right\|_F \geq w\theta_1 - 3C_s \sqrt{\left(\frac{K}{n}\right)^3 \frac{\widehat{d}}{\Lambda \wedge \kappa}}\right) \\ & \leq \mathbb{P}\left(\max_{l \in [L]} \max_{u \in S(\Lambda)} \left\| \mathbf{G}_l^{(T_l+u)} - \widehat{\mathbf{G}}_l^{(T_l+u)} \right\|_F \geq \Delta_2 \theta_1\right) + \mathbb{P}\left(\widehat{d} \geq \frac{3c_0}{2} d\right). \end{aligned}$$

Then the desired result follows from Proposition A.2 and Lemma A.4.  $\blacksquare$

*Proof of the localization bound of Theorem 4.4.* Fix  $\eta \in (0, 1)$ . We first show that

$$\left\{ \frac{1}{\Lambda} |\widehat{\tau} - \tau| \geq \eta \right\} \subseteq \left\{ \max_{l \in [L]} \max_{u \in S(\Lambda)} \left\| \mathbf{G}_l^{(T_l+u)} - \widehat{\mathbf{G}}_l^{(T_l+u)} \right\|_F \geq \frac{\eta}{8} \left(\frac{\Lambda \wedge \kappa}{3\Lambda} \left(1 - \frac{\Lambda \wedge \kappa}{3\Lambda}\right)\right)^\delta \|B_0 - B_1\|_F \right\}. \quad (12)$$

So, assume that  $|\widehat{\tau} - \tau| \geq \eta\Lambda$ . Let  $l_0$  be such that  $\tau \in \bar{\mathcal{I}}_{l_0}$  and  $l_1$  be such that  $\widehat{\tau} \in \bar{\mathcal{I}}_{l_1}$ . We consider the following cases:

**Case 1:**  $\tau \in \bar{\mathcal{I}}_{l_1}$ . First consider the situation  $\widehat{\tau} \geq \tau + \eta\Lambda (> \tau)$ . Then  $\tau + \eta\Lambda \in \mathcal{I}_{l_1}$ . As  $\left\| \mathbf{G}_{l_1}^{(T_{l_1}+u)} \right\|_F$  is decreasing for  $u \in [\tau - T_{l_1}, \Lambda]$ , we have,  $\left\| \mathbf{G}_{l_1}^{(\widehat{\tau})} \right\|_F \leq \left\| \mathbf{G}_{l_1}^{(\tau + \eta\Lambda)} \right\|_F$ . Lemma A.3 then implies

$$\begin{aligned} \left\| \mathbf{G}_{l_1}^{(\tau)} \right\|_F - \left\| \mathbf{G}_{l_1}^{(\widehat{\tau})} \right\|_F & \geq \frac{\eta}{1 + \eta} \left(\frac{\tau - T_{l_1}}{\Lambda} \left(1 - \frac{\tau - T_{l_1}}{\Lambda}\right)\right)^\delta \|B_0 - B_1\|_F \\ & \geq \frac{\eta}{2} \left(\frac{\Lambda \wedge \kappa}{3\Lambda} \left(1 - \frac{\Lambda \wedge \kappa}{3\Lambda}\right)\right)^\delta \|B_0 - B_1\|_F. \end{aligned}$$

In the situation  $\widehat{\tau} \leq \tau - \eta\Lambda (< \tau)$ ,  $\tau - \eta\Lambda \in \mathcal{I}_{l_1}$ . As  $\left\| \mathbf{G}_{l_1}^{(T_{l_1}+u)} \right\|_F$  is increasing for  $u \in [1, \tau - T_{l_1}]$ , we have,  $\left\| \mathbf{G}_{l_1}^{(\widehat{\tau})} \right\|_F \leq \left\| \mathbf{G}_{l_1}^{(\tau - \eta\Lambda)} \right\|_F$ . Lemma A.3 then implies

$$\begin{aligned} \left\| \mathbf{G}_{l_1}^{(\tau)} \right\|_F - \left\| \mathbf{G}_{l_1}^{(\widehat{\tau})} \right\|_F & \geq \frac{\eta}{1 + \eta} \left(\frac{\tau - T_{l_1}}{\Lambda} \left(1 - \frac{\tau - T_{l_1}}{\Lambda}\right)\right)^\delta \|B_0 - B_1\|_F \\ & \geq \frac{\eta}{2} \left(\frac{\Lambda \wedge \kappa}{3\Lambda} \left(1 - \frac{\Lambda \wedge \kappa}{3\Lambda}\right)\right)^\delta \|B_0 - B_1\|_F. \end{aligned}$$

Thus in both situations of **Case I**, as  $\left\| \widehat{\mathbf{G}}_{l_1}^{(\tau)} \right\|_F \leq \left\| \widehat{\mathbf{G}}_{l_1}^{(\widehat{\tau})} \right\|_F$ ,

$$\begin{aligned} \left\| \mathbf{G}_{l_1}^{(\tau)} \right\|_F - \left\| \mathbf{G}_{l_1}^{(\widehat{\tau})} \right\|_F & = \left\| \mathbf{G}_{l_1}^{(\tau)} \right\|_F - \left\| \widehat{\mathbf{G}}_{l_1}^{(\tau)} \right\|_F + \left\| \widehat{\mathbf{G}}_{l_1}^{(\widehat{\tau})} \right\|_F - \left\| \mathbf{G}_{l_1}^{(\widehat{\tau})} \right\|_F + \left\| \widehat{\mathbf{G}}_{l_1}^{(\tau)} \right\|_F - \left\| \widehat{\mathbf{G}}_{l_1}^{(\widehat{\tau})} \right\|_F \\ & \leq 2 \max_{l \in [L]} \max_{u \in S(\Lambda)} \left\| \mathbf{G}_l^{(T_l+u)} - \widehat{\mathbf{G}}_l^{(T_l+u)} \right\|_F. \end{aligned}$$

**Case 2:**  $\tau \notin \bar{\mathcal{I}}_{l_1}$ . Then,  $\left\| \mathbf{G}_{l_1}^{(\widehat{\tau})} \right\|_F = 0$  and

$$\begin{aligned} \left\| \mathbf{G}_{l_0}^{(\tau)} \right\|_F - \left\| \mathbf{G}_{l_1}^{(\widehat{\tau})} \right\|_F & = \left(\frac{\tau - T_{l_0}}{\Lambda} \left(1 - \frac{\tau - T_{l_0}}{\Lambda}\right)\right)^\delta \|B_0 - B_1\|_F \\ & \geq \eta \left(\frac{\Lambda \wedge \kappa}{3\Lambda} \left(1 - \frac{\Lambda \wedge \kappa}{3\Lambda}\right)\right)^\delta \|B_0 - B_1\|_F. \end{aligned}$$

**Case 3:**  $\tau \in \mathcal{I}_{l_1} \setminus \bar{\mathcal{I}}_{l_1}$  and  $\left(\frac{\tau - T_{l_1}}{\Lambda} \left(1 - \frac{\tau - T_{l_1}}{\Lambda}\right)\right)^\delta < \frac{1}{2} \left(\frac{\Lambda \wedge \kappa}{3\Lambda} \left(1 - \frac{\Lambda \wedge \kappa}{3\Lambda}\right)\right)^\delta$ . Then,  $\|\mathbf{G}_{l_1}^{(\hat{\tau})}\|_F \leq \|\mathbf{G}_{l_1}^{(\tau)}\|_F$  and

$$\begin{aligned} \|\mathbf{G}_{l_0}^{(\tau)}\|_F - \|\mathbf{G}_{l_1}^{(\hat{\tau})}\|_F &\geq \left(\frac{\tau - T_{l_0}}{\Lambda} \left(1 - \frac{\tau - T_{l_0}}{\Lambda}\right)\right)^\delta \|B_0 - B_1\|_F - \|\mathbf{G}_{l_1}^{(\tau)}\|_F \\ &\geq \left(\frac{\Lambda \wedge \kappa}{3\Lambda} \left(1 - \frac{\Lambda \wedge \kappa}{3\Lambda}\right)\right)^\delta \|B_0 - B_1\|_F - \left(\frac{\tau - T_{l_1}}{\Lambda} \left(1 - \frac{\tau - T_{l_1}}{\Lambda}\right)\right)^\delta \|B_0 - B_1\|_F \\ &\geq \left(\frac{\Lambda \wedge \kappa}{3\Lambda} \left(1 - \frac{\Lambda \wedge \kappa}{3\Lambda}\right)\right)^\delta \|B_0 - B_1\|_F - \frac{1}{2} \left(\frac{\Lambda \wedge \kappa}{3\Lambda} \left(1 - \frac{\Lambda \wedge \kappa}{3\Lambda}\right)\right)^\delta \|B_0 - B_1\|_F \\ &\geq \frac{\eta}{2} \left(\frac{\Lambda \wedge \kappa}{3\Lambda} \left(1 - \frac{\Lambda \wedge \kappa}{3\Lambda}\right)\right)^\delta \|B_0 - B_1\|_F. \end{aligned}$$

**Case 4:**  $\tau \in \mathcal{I}_{l_1} \setminus \bar{\mathcal{I}}_{l_1}$  and  $\left(\frac{\tau - T_{l_1}}{\Lambda} \left(1 - \frac{\tau - T_{l_1}}{\Lambda}\right)\right)^\delta \geq \frac{1}{2} \left(\frac{\Lambda \wedge \kappa}{3\Lambda} \left(1 - \frac{\Lambda \wedge \kappa}{3\Lambda}\right)\right)^\delta$ . Note that in this case,  $\|\mathbf{G}_{l_0}^{(\tau)}\|_F \geq \|\mathbf{G}_{l_1}^{(\tau)}\|_F$  as  $\tau - T_{l_0} \geq \tau - T_{l_1}$ . Then, using a similar argument as in **Case 1** and Lemma A.3, we get

$$\begin{aligned} \|\mathbf{G}_{l_0}^{(\tau)}\|_F - \|\mathbf{G}_{l_1}^{(\hat{\tau})}\|_F &\geq \|\mathbf{G}_{l_1}^{(\tau)}\|_F - \|\mathbf{G}_{l_1}^{(\hat{\tau})}\|_F \\ &\geq \frac{\eta}{1 + \eta} \left(\frac{\tau - T_{l_1}}{\Lambda} \left(1 - \frac{\tau - T_{l_1}}{\Lambda}\right)\right)^\delta \|B_0 - B_1\|_F \\ &\geq \frac{\eta}{4} \left(\frac{\Lambda \wedge \kappa}{3\Lambda} \left(1 - \frac{\Lambda \wedge \kappa}{3\Lambda}\right)\right)^\delta \|B_0 - B_1\|_F. \end{aligned}$$

Thus in **Cases 2, 3, 4**, as  $\|\widehat{\mathbf{G}}_{l_0}^{(\tau)}\|_F \leq \|\widehat{\mathbf{G}}_{l_1}^{(\hat{\tau})}\|_F$ ,

$$\begin{aligned} \|\mathbf{G}_{l_0}^{(\tau)}\|_F - \|\mathbf{G}_{l_1}^{(\hat{\tau})}\|_F &= \|\mathbf{G}_{l_0}^{(\tau)}\|_F - \|\widehat{\mathbf{G}}_{l_0}^{(\tau)}\|_F + \|\widehat{\mathbf{G}}_{l_1}^{(\hat{\tau})}\|_F - \|\mathbf{G}_{l_1}^{(\hat{\tau})}\|_F + \|\widehat{\mathbf{G}}_{l_0}^{(\tau)}\|_F - \|\widehat{\mathbf{G}}_{l_1}^{(\hat{\tau})}\|_F \\ &\leq 2 \max_{l \in [L]} \max_{u \in S(\Lambda)} \left\| \mathbf{G}_l^{(T_l+u)} - \widehat{\mathbf{G}}_l^{(T_l+u)} \right\|_F. \end{aligned}$$

Hence (12) holds. Then the desired result follows from Proposition A.2. ■

### A.3 Proof of Proposition A.2

In this section, we give the proof of Proposition A.2.

Let  $C_m = \{k : Z_{km} = 1\}$  and  $n_m = |C_m|$  for  $m \in [K]$ . Let  $\{\tilde{B}^{(t)}\}_{t=1}^T$  be the sequence of  $K \times K$  matrices given by,

$$\tilde{B}_{k_1 k_2}^{(t)} = \frac{1}{n_{k_1 k_2}^2} \sum_{i \in C_{k_1}, j \in C_{k_2}, i < j} A_{ij}^{(t)}, \quad \text{where } n_{k_1 k_2}^2 = \begin{cases} \binom{n_{k_1}}{2}, & k_1 = k_2 \\ n_{k_1} n_{k_2}, & k_1 \neq k_2 \end{cases}, \quad k_1, k_2 \in [K].$$

For  $l \in [L]$ , define the collection of sets  $\chi_l^{(1)}$  and  $\chi_l^{(2)}$  as follows:

$$\chi_l^{(1)} = \{(T_l, T_l + u) : u \in S(\Lambda)\}, \quad \chi_l^{(2)} = \{(T_l + u, T_l + \Lambda) : u \in S(\Lambda)\}.$$

We need the following concentration inequalities to prove Proposition A.2.

**Lemma A.5.** *Under Assumption 4.2, there exists absolute constants  $\tilde{C}_1, \tilde{C}_2 > 0$  such that for all  $0 < \theta \leq (C_s \vee 1)K\rho$  and  $i = 1, 2$ ,*

$$\mathbb{P}\left(\max_{l \in [L]} \max_{J \in \chi_l^{(i)}} \left\| \frac{1}{|J|} \sum_{s \in J} (\tilde{B}^{(s)} - B^{(s)}) \right\|_F > \theta\right) \leq \tilde{C}_1 K^2 L \exp\left\{-\frac{\tilde{C}_2 \theta^2}{K \theta_1^2}\right\}.$$

**Lemma A.6.** Under Assumptions 3.1 and 4.2, for some absolute constant  $\tilde{C}_4 > 0$ , for all  $0 < \theta \leq (C_s \vee 1)K\rho$ ,  $i = 1, 2$  and  $\epsilon_\theta = \tilde{C}_4\theta/K^2$ ,

$$\mathbb{P}\left(\max_{l \in [L]} \max_{J \in \mathcal{X}_l^{(i)}} \left\| \frac{1}{|J|} \sum_{s \in J} (\hat{B}^{(s)} - \tilde{B}^{(s)}) \right\|_F > \theta\right) \leq \mathcal{P}\{\epsilon_\theta, k, n, T, d\}.$$

The proofs of Lemma A.5 and Lemma A.6 are given in Section B.

*Proof of Proposition A.2.* Observe that

$$\begin{aligned} & \max_{l \in [L]} \max_{u \in S(\Lambda)} \left\| \mathbf{G}_l^{(T_l+u)} - \hat{\mathbf{G}}_l^{(T_l+u)} \right\|_F \\ & \leq 2 \max_{i=1,2} \max_{l \in [L]} \max_{J \in \mathcal{X}_l^{(i)}} \left\| \frac{1}{|J|} \sum_{s \in J} (\hat{B}^{(s)} - B^{(s)}) \right\|_F \\ & \leq 2 \max_{i=1,2} \max_{l \in [L]} \max_{J \in \mathcal{X}_l^{(i)}} \left\| \frac{1}{|J|} \sum_{s \in J} (\hat{B}^{(s)} - \tilde{B}^{(s)}) \right\|_F + 2 \max_{i=1,2} \max_{l \in [L]} \max_{J \in \mathcal{X}_l^{(i)}} \left\| \frac{1}{|J|} \sum_{s \in J} (\tilde{B}^{(s)} - B^{(s)}) \right\|_F \end{aligned}$$

Hence,

$$\begin{aligned} & \mathbb{P}\left(\max_{l \in [L]} \max_{u \in S(\Lambda)} \left\| \mathbf{G}_l^{(T_l+u)} - \hat{\mathbf{G}}_l^{(T_l+u)} \right\|_F > \theta\right) \\ & \leq \sum_{i=1}^2 \left[ \mathbb{P}\left(\max_{l \in [L]} \max_{J \in \mathcal{X}_l^{(i)}} \left\| \frac{1}{|J|} \sum_{s \in J} (\hat{B}^{(s)} - \tilde{B}^{(s)}) \right\|_F > \frac{\theta}{2}\right) + \mathbb{P}\left(\max_{l \in [L]} \max_{J \in \mathcal{X}_l^{(i)}} \left\| \frac{1}{|J|} \sum_{s \in J} (\tilde{B}^{(s)} - B^{(s)}) \right\|_F > \frac{\theta}{2}\right) \right]. \end{aligned}$$

The desired result follows from Lemma A.5 and Lemma A.6. ■

## B PROOFS OF VARIOUS LEMMAS

In this section, we give the proofs of the required Lemmas.

*Proof of Lemma A.1.* Because of (9), it is enough to show that

$$\chi^2(\mathbb{P}_{B_0}^t, \mathbb{E}_{B_1 \sim \nu_1} \mathbb{P}_{B_1}^t) \leq \epsilon^2.$$

Using the same calculation as in Eq. (3.1) in the proof of Lemma 3.1 of [Bhattacharyya et al. \(2020\)](#), we get that

$$1 + \chi^2(\mathbb{P}_{B_0}^t, \mathbb{E}_{B_1 \sim \nu_1} \mathbb{P}_{B_1}^t) \leq \mathbb{E}_{U, \tilde{U}} \exp\left(\frac{t\alpha^2\rho}{1-\rho} \langle ZUU^\top Z^\top, Z\tilde{U}\tilde{U}^\top Z^\top \rangle\right)$$

Now note that

$$\begin{aligned} \langle ZUU^\top Z^\top, Z\tilde{U}\tilde{U}^\top Z^\top \rangle &= \text{tr}(ZUU^\top Z^\top Z\tilde{U}\tilde{U}^\top Z^\top) \\ &= \text{tr}(U^\top Z^\top Z\tilde{U}\tilde{U}^\top Z^\top ZU) \\ &= (U^\top Z^\top Z\tilde{U})^2 \\ &= \left(\frac{n}{K}\right)^2 (U^\top \tilde{U})^2. \end{aligned}$$

By Hoeffding's inequality,

$$\mathbb{P}(|U^\top \tilde{U}| \geq \sqrt{\ell K}) \leq 2 \exp\left(-\frac{2\ell K}{4K}\right) = 2 \exp(-\ell/2).$$

Therefore

$$\begin{aligned}
 \mathbb{E}_{U, \tilde{U}} \exp \left( \frac{t\alpha^2\rho}{1-\rho} \langle ZUU^\top Z^\top, Z\tilde{U}\tilde{U}^\top Z^\top \rangle \right) &= \mathbb{E}_{U, \tilde{U}} \exp \left( \frac{t\alpha^2\rho}{1-\rho} \left( \frac{n}{k} \right)^2 (U^\top \tilde{U})^2 \right) \\
 &= \int_0^\infty \mathbb{P} \left( \exp \left( \frac{t\alpha^2\rho}{1-\rho} \left( \frac{n}{k} \right)^2 (U^\top \tilde{U})^2 \right) > u \right) du \\
 &= \int_0^\infty \mathbb{P} \left( (U^\top \tilde{U})^2 > \log(u) \frac{K^2}{n^2} \frac{1-\rho}{t\alpha^2\rho} \right) du \\
 &= 1 + \int_1^\infty \mathbb{P} \left( (U^\top \tilde{U})^2 > \log(u) \frac{K^2}{n^2} \frac{1-\rho}{t\alpha^2\rho} \right) du \\
 &\leq 1 + \int_1^\infty 2 \exp \left( -\frac{1}{2} \log(u) \frac{K^2}{n^2} \frac{1-\rho}{t\alpha^2\rho} \right) du \\
 &= 1 + 2 \int_1^\infty \frac{du}{u^\zeta} \quad \left( \text{where } \zeta := \frac{1}{2} \frac{K^2}{n^2} \frac{1-\rho}{t\alpha^2\rho} \right) \\
 &= 1 + \frac{2}{\zeta - 1} \leq 1 + \epsilon^2,
 \end{aligned}$$

provided that  $\zeta \geq 1 + \frac{2}{\epsilon^2}$ , i.e.

$$t\alpha^2\rho \frac{n^2}{K} \leq \frac{1}{2} \left( 1 + \frac{2}{\epsilon^2} \right)^{-1} (1-\rho) = c(1-\rho),$$

where

$$c = c(\epsilon) = \frac{1}{2} \left( 1 + \frac{2}{\epsilon^2} \right)^{-1}.$$

It follows that given  $\epsilon > 0$ , there exists  $c(\epsilon) > 0$  such that

$$\chi^2(\mathbb{P}_{\Theta_0}^t, \mathbb{P}_{\Theta_1}^t) \leq \epsilon^2,$$

whenever

$$t\alpha^2\rho \frac{n^2}{K} \leq c(1-\rho).$$

We now note that

$$\|B_1 - B_0\|_F^2 = \alpha^2\rho^2 (U^\top U)^2 = \alpha^2\rho^2 K^2.$$

Thus we get that the condition

$$\frac{t\|B_1 - B_0\|_F^2 n^2}{\rho K^3} \leq c(1-\rho)$$

or, equivalently,

$$\|B_1 - B_0\|_F^2 \leq c(1-\rho) \left( \frac{K}{n} \right)^3 \frac{d}{t}$$

implies that

$$\chi^2(\mathbb{P}_{B_0}^t, \mathbb{E}_{B_1 \sim \nu_1} \mathbb{P}_{B_1}^t) \leq \epsilon^2.$$

In particular, assuming that  $\rho \leq \rho_0 < 1$  in the class of models to be considered, we get that the condition

$$\|B_1 - B_0\|_F^2 \leq c(1-\rho_0) \left( \frac{K}{n} \right)^3 \frac{d}{t}$$

implies that

$$\chi^2(\mathbb{P}_{B_0}^t, \mathbb{E}_{B_1 \sim \nu_1} \mathbb{P}_{B_1}^t) \leq \epsilon^2.$$

This completes the proof. ■

*Proof of Lemma A.3.* We show the proof for the case  $\|\mathbf{G}_{l_0}^{\tau+\eta\Lambda}\|_F$ . The proof for the case  $\|\mathbf{G}_{l_0}^{\tau-\eta\Lambda}\|_F$  is similar. Let  $\tau + \eta\Lambda \in \mathcal{I}_{l_0}$ . Then,

$$\begin{aligned} & \left\| \mathbf{G}_{l_0}^\tau \right\|_F - \left\| \mathbf{G}_{l_0}^{\tau+\eta\Lambda} \right\|_F \\ &= \left[ \left( \frac{\tau - T_{l_0}}{\Lambda} \left( 1 - \frac{\tau - T_{l_0}}{\Lambda} \right) \right)^\delta - \left( \frac{\tau - T_{l_0} + \eta\Lambda}{\Lambda} \left( 1 - \frac{\tau - T_{l_0} + \eta\Lambda}{\Lambda} \right) \right)^\delta \frac{\tau - T_{l_0}}{\tau - T_{l_0} + \eta\Lambda} \right] \|B_0 - B_1\|_F \\ &\geq \left( \frac{\tau - T_{l_0}}{\Lambda} \left( 1 - \frac{\tau - T_{l_0}}{\Lambda} \right) \right)^\delta \|B_0 - B_1\|_F \left[ 1 - \left( 1 - \frac{\eta\Lambda}{\tau - T_{l_0} + \eta\Lambda} \right)^{1-\delta} \left( 1 - \frac{\eta\Lambda}{\Lambda - \tau + T_{l_0} + \eta\Lambda} \right)^\delta \right] \end{aligned} \quad (13)$$

Using the weighted AM-GM inequality and the fact that  $0 \leq \tau - T_{l_0} \leq \Lambda$ , we get

$$\begin{aligned} & \left( 1 - \frac{\eta\Lambda}{\tau - T_{l_0} + \eta\Lambda} \right)^{1-\delta} \left( 1 - \frac{\eta\Lambda}{\Lambda - \tau + T_{l_0} + \eta\Lambda} \right)^\delta \leq (1-\delta) \left( 1 - \frac{\eta\Lambda}{\tau - T_{l_0} + \eta\Lambda} \right) + \delta \left( 1 - \frac{\eta\Lambda}{\Lambda - \tau + T_{l_0} + \eta\Lambda} \right) \\ &= 1 - \eta\Lambda \left[ \frac{1-\delta}{\tau - T_{l_0} + \eta\Lambda} + \frac{\delta}{\Lambda - \tau + T_{l_0} + \eta\Lambda} \right] \leq 1 - \eta\Lambda \left[ \frac{1-\delta}{\Lambda + \eta\Lambda} + \frac{\delta}{\Lambda + \eta\Lambda} \right] = 1 - \frac{\eta}{1+\eta}. \end{aligned} \quad (14)$$

Combining (13) and (14), we get the desired result.  $\blacksquare$

*Proof of Lemma A.5.* We give the proof for  $i = 1$ . The proof for  $i = 2$  is similar. Let  $M = \lceil \log_2 (\Lambda / (\frac{\Lambda}{3} \wedge \frac{\kappa}{2})) \rceil$ . For  $m \in [M]$ , define  $s_m = 2^m (\frac{\Lambda}{3} \wedge \frac{\kappa}{2})$ . Then, using the union bound, the fact that  $\text{Var}(A_{ij}^t) \leq \rho$  for all  $t \in [T]$  and  $i, j \in [n]$ , the maximal form of Bernstein's inequality and Assumption 4.2, we get that for some absolute constants  $C, C', \tilde{C}_2 > 0$ ,

$$\begin{aligned} & \mathbb{P} \left( \max_{u \in [s_m/2, s_m]} \left\| \frac{1}{u} \sum_{v=1}^u (\tilde{B}^{(T_l+v)} - B^{(T_l+v)}) \right\|_F > \theta \right) \\ &\leq \mathbb{P} \left( \max_{u \in [s_m/2, s_m]} \left\| \sum_{v=1}^u (\tilde{B}^{(T_l+v)} - B^{(T_l+v)}) \right\|_F > \frac{\theta s_m}{2} \right) \\ &\leq \mathbb{P} \left( \sum_{k_1, k_2 \in [K]} \max_{u \in [s_m/2, s_m]} \left( \sum_{v=1}^u (\tilde{B}_{k_1 k_2}^{(T_l+v)} - B_{k_1 k_2}^{(T_l+v)}) \right)^2 > \frac{\theta^2 s_m^2}{4} \right) \\ &\leq \sum_{k_1, k_2 \in [K]} \mathbb{P} \left( \max_{u \in [s_m/2, s_m]} \left| \sum_{v=1}^u (\tilde{B}_{k_1 k_2}^{(T_l+v)} - B_{k_1 k_2}^{(T_l+v)}) \right| > \frac{\theta s_m}{2K} \right) \\ &= \sum_{k_1, k_2 \in [K]} \mathbb{P} \left( \max_{u \in [s_m/2, s_m]} \left| \sum_{v=1}^u \frac{1}{n_{k_1 k_2}^2} \sum_{i \in C_{k_1}, j \in C_{k_2}, i < j} (A_{ij}^{(T_l+v)} - \mathbb{E}(A_{ij}^{(T_l+v)})) \right| > \frac{\theta s_m}{2K} \right) \\ &\leq \sum_{k_1, k_2 \in [K]} \mathbb{P} \left( \max_{u \in [s_m/2, s_m]} \left| \sum_{v=1}^u \sum_{i \in C_{k_1}, j \in C_{k_2}, i < j} (A_{ij}^{(T_l+v)} - \mathbb{E}(A_{ij}^{(T_l+v)})) \right| > \frac{C\theta s_m}{K} \left( \frac{n}{K} \right)^2 \right) \\ &\leq \sum_{k_1, k_2 \in [K]} 2 \exp \left\{ - \frac{C^2 \theta^2 s_m^2 n^4 / (2K^6)}{\sum_{v=1}^{s_m} \sum_{i \in C_{k_1}, j \in C_{k_2}, i < j} \text{Var}(A_{ij}^{(T_l+v)}) + C\theta s_m n^2 / (3K^3)} \right\} \\ &\leq 2K^2 \exp \left\{ - \frac{C' \theta^2 s_m}{K(K/n)^3 d} \right\} \\ &\leq 2K^2 \exp \left\{ - \frac{\tilde{C}_2 \theta^2 2^m}{K \theta_1^2} \right\}. \end{aligned}$$

Hence, using the union bound and the summability of a geometric series, for some absolute constant  $\tilde{C}_1$ , we get that

$$\mathbb{P} \left( \max_{m \in [M]} \max_{u \in [s_m/2, s_m]} \left\| \frac{1}{u} \sum_{v=1}^u (\tilde{B}^{(T_l+v)} - B^{(T_l+v)}) \right\|_F > \theta \right) \leq \tilde{C}_1 K^2 \exp \left\{ - \frac{\tilde{C}_2 \theta^2}{K \theta_1^2} \right\}.$$

Now note that

$$\max_{l \in [L]} \max_{J \in \chi_l^{(1)}} \left\| \frac{1}{|J|} \sum_{s \in J} (\tilde{B}^{(s)} - B^{(s)}) \right\|_F \leq \max_{l \in [L]} \max_{m \in [M]} \max_{u \in [s_m/2, s_m]} \left\| \frac{1}{u} \sum_{v=1}^u (\tilde{B}^{(T_l+v)} - B^{(T_l+v)}) \right\|_F.$$

Using the union bound again, we get the desired result.  $\blacksquare$

*Proof of Lemma A.6.* Fix  $i \in [2]$  and  $\epsilon > 0$ . Let  $\bar{Z}(\epsilon)$  be a (non-random) membership matrix with at most  $\epsilon n$  nodes being classified differently than  $Z$ . In other words,  $\bar{Z}(\epsilon)$  has at most  $\epsilon n$  rows that are different than the corresponding rows of  $Z$ . Let  $\bar{C}_m(\epsilon) = \{k : \bar{Z}(\epsilon)_{km} = 1\}$ ,  $\bar{n}_m(\epsilon) = |\bar{C}_m(\epsilon)|$  for  $m \in [K]$ . Define a sequence of matrices  $\{\bar{B}^{(t)}(\epsilon)\}_{t=1}^T$  in the following way: for  $k_1, k_2 \in [K]$ ,

$$\bar{B}^{(t)}(\epsilon)_{k_1 k_2} = \frac{1}{\bar{n}_{k_1 k_2}^2(\epsilon)} \sum_{i \in \bar{C}_{k_1}(\epsilon), j \in \bar{C}_{k_2}(\epsilon), i < j} A_{ij}^{(t)}, \quad \text{where } \bar{n}_{k_1 k_2}^2(\epsilon) = \begin{cases} \binom{\bar{n}_{k_1}(\epsilon)}{2}, & k_1 = k_2 \\ \bar{n}_{k_1}(\epsilon) \bar{n}_{k_2}(\epsilon), & k_1 \neq k_2 \end{cases}.$$

Note that for all  $m \in [K]$ ,  $n_m + \epsilon n \geq \bar{n}_m(\epsilon) \geq n_m - \epsilon n \geq C_2 n / K - \epsilon n$  and  $|C_m \setminus \bar{C}_m(\epsilon)| \leq \epsilon n$ ,  $|\bar{C}_m(\epsilon) \setminus C_m| \leq \epsilon n$ ,  $C_m \cap \bar{C}_m(\epsilon) \leq C_1 n / K$ . Therefore, for all  $t \in [T]$ ,

$$\begin{aligned} & \left| \tilde{B}_{k_1 k_2}^{(t)} - \bar{B}^{(t)}(\epsilon)_{k_1 k_2} \right| \\ &= \left| \frac{1}{\bar{n}_{k_1 k_2}^2(\epsilon)} \sum_{i \in \bar{C}_{k_1}(\epsilon), j \in \bar{C}_{k_2}(\epsilon), i < j} A_{ij}^{(t)} - \frac{1}{n_{k_1 k_2}^2} \sum_{i \in C_{k_1}, j \in C_{k_2}, i < j} A_{ij}^{(t)} \right| \\ &\leq \frac{1}{\bar{n}_{k_1 k_2}^2(\epsilon)} \left| \sum_{i \in \bar{C}_{k_1}(\epsilon), j \in \bar{C}_{k_2}(\epsilon), i < j} A_{ij}^{(t)} - \sum_{i \in C_{k_1}, j \in C_{k_2}, i < j} A_{ij}^{(t)} \right| + \left| \frac{n_{k_1 k_2}^2}{\bar{n}_{k_1 k_2}^2(\epsilon)} - 1 \right| \left( \frac{1}{n_{k_1 k_2}^2} \sum_{i \in C_{k_1}, j \in C_{k_2}, i < j} A_{ij}^{(t)} \right) \\ &\leq \frac{1}{\bar{n}_{k_1 k_2}^2(\epsilon)} \left( 2\epsilon^2 n^2 + 4\epsilon C_1 \frac{n^2}{K} \right) + \left| \frac{n_{k_1 k_2}^2}{\bar{n}_{k_1 k_2}^2(\epsilon)} - 1 \right|. \end{aligned}$$

Note that  $n_{k_1 k_2}^2 \leq C_1^2 (n/K)^2$  and  $\bar{n}_{k_1 k_2}^2(\epsilon) \geq (C_2 \frac{n}{K} - \epsilon n)^2 / 4$ . These and some algebra show that for some appropriately chosen constant  $\tilde{C}_4$ ,  $|\tilde{B}_{k_1 k_2}^{(t)} - \bar{B}^{(t)}(\epsilon)_{k_1 k_2}| \leq K\epsilon / \tilde{C}_4$ . Then,

$$\max_{l \in [L]} \max_{J \in \chi_l^{(i)}} \left\| \frac{1}{|J|} \sum_{s \in J} (\bar{B}^{(s)}(\epsilon) - \tilde{B}^{(s)}) \right\|_F \leq K^2 \epsilon / \tilde{C}_4.$$

Consequently,

$$\{\mathcal{M}(\hat{Z}) \leq \epsilon_\theta n\} \subseteq \left\{ \max_{l \in [L]} \max_{J \in \chi_l^{(i)}} \left\| \frac{1}{|J|} \sum_{s \in J} (\hat{B}^{(s)} - \tilde{B}^{(s)}) \right\|_F \leq \theta \right\}.$$

This implies that

$$\mathbb{P} \left( \max_{l \in [L]} \max_{J \in \chi_l^{(i)}} \left\| \frac{1}{|J|} \sum_{s \in J} (\hat{B}^{(s)} - \tilde{B}^{(s)}) \right\|_F > \theta \right) \leq \mathbb{P}(\mathcal{M}(\hat{Z}) \geq \epsilon_\theta n) \leq \mathcal{P}\{\epsilon_\theta, k, n, T, d\}.$$

This completes the proof.  $\blacksquare$

*Proof of Lemma A.4.* Note that  $\mathbb{E}(\hat{d}) = \bar{d} = c_0 d$ . Bernstein's inequality implies that for any  $t \in (0, 1)$ ,

$$\begin{aligned} \mathbb{P}(\hat{d} - \bar{d} \geq td) &= \mathbb{P} \left( \sum_{t \in [T]} \sum_{i, j \in [n]} (A_{ij}^{(t)} - \mathbb{E}(A_{ij}^{(t)})) \geq nTtd \right) \\ &\leq \exp \left\{ - \frac{\frac{1}{2} n^2 T^2 t^2 d^2}{\sum_{t \in [T]} \sum_{i, j \in [n]} \text{Var}(A_{ij}^{(t)}) + \frac{1}{3} nTtd} \right\} \\ &\leq \exp \left\{ - \frac{\frac{1}{2} n^2 T^2 t^2 d^2}{nTd + \frac{1}{3} nTtd} \right\} \\ &\leq \exp \left\{ - \frac{t^2}{4} nTd \right\}. \end{aligned}$$

Putting  $t = c_0/2$ , we get the upper-tail inequality. The lower-tail inequality can be proved similarly.  $\blacksquare$

## C ADDITIONAL EMPIRICAL RESULTS

### C.1 Simulation Studies

In Figures 4, 5 and 6, we illustrate the behavior of the CUSUM statistics, corresponding to both  $\hat{\tau}$  and  $\tilde{\tau}$ , in setups (a), (c) and (d) of Experiment I, respectively. In Tables 3, 4, 5 and 6, we report the values of the parameter  $Td$  and the corresponding mean (across 10 MC runs) relative errors for every method under comparison in setups (a), (b), (c) and (d) of Experiment II, respectively.

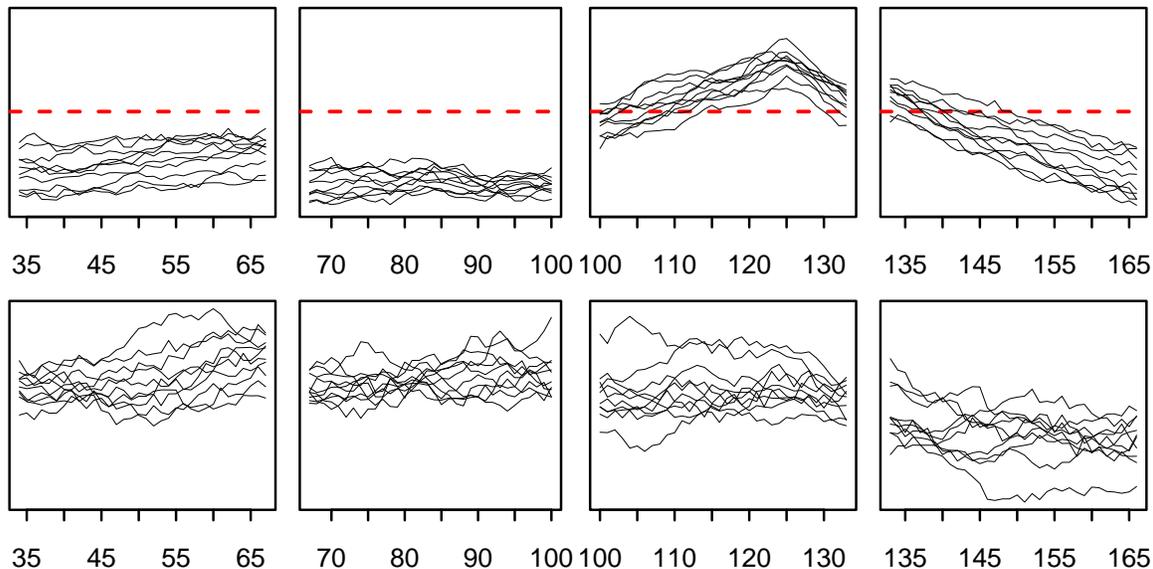


Figure 4: Experiment I-(a). Top row: Observed CUSUM statistics (4) (with  $l = 1, 2, 3, 4$  and  $\Lambda = 100$ ) for 10 MC runs. The dashed lines display the mean threshold value. Bottom row: Observed adjacency-based CUSUM statistics (5) (with  $l = 1, 2, 3, 4$  and  $\Lambda = 100$ ) for 10 MC runs.

Table 3: Experiment II-(a): Mean Relative Error

Method	$Td = 200$	$Td = 100$	$Td = 20$
Bhattacharjee et al. (2020)	0.001	0.004	0.114
Bhattacharyya et al. (2020)	0.318	0.3385	0.3245
Wang et al. (2021)	0.375	0.375	0.375
Zhao et al. (2019)	0.0825	0.1525	0.375
Our Method	0	0.0015	0.013

### C.2 A Real-World Example of Community Splitting

For estimating (multiple) changepoints in the US Senate roll call data, we couple our algorithm with wild binary segmentation (Fryzlewicz et al., 2014). The number and locations of the detected changepoints depend on the value of the threshold used in wild binary segmentation. We summarize them in Table 7. Each detected changepoint correspond to 49 consecutive roll calls within a certain time-period that we report in Table 8.

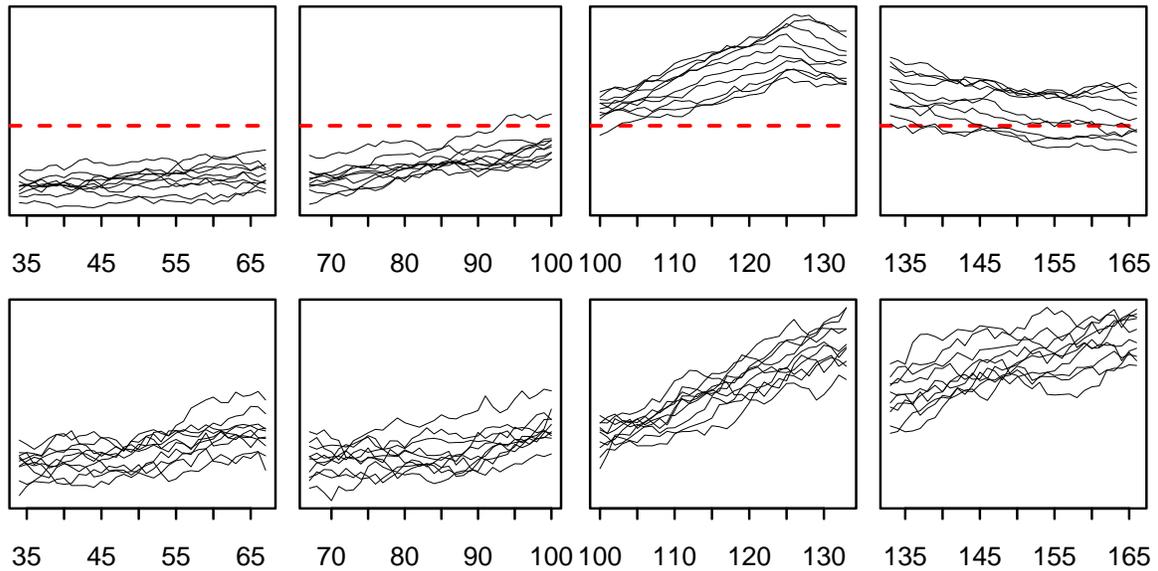


Figure 5: Experiment I-(c). Top row: Observed CUSUM statistics (4) (with  $l = 1, 2, 3, 4$  and  $\Lambda = 100$ ) for 10 MC runs. The dashed lines display the mean threshold value. Bottom row: Observed adjacency-based CUSUM statistics (5) (with  $l = 1, 2, 3, 4$  and  $\Lambda = 100$ ) for 10 MC runs.

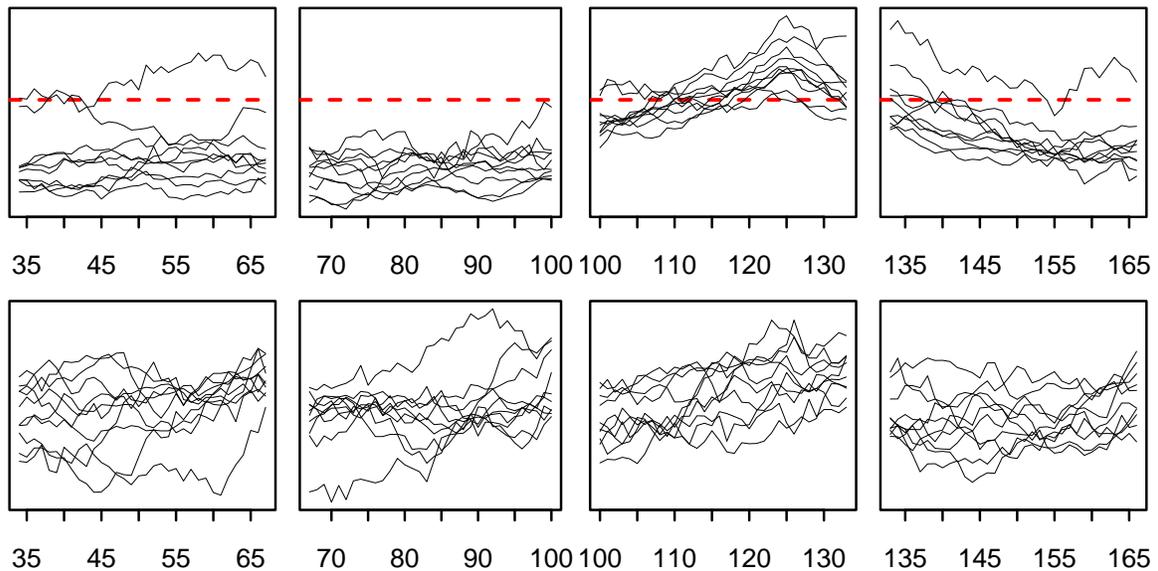


Figure 6: Experiment I-(d). Top row: Observed CUSUM statistics (4) (with  $l = 1, 2, 3, 4$  and  $\Lambda = 100$ ) for 10 MC runs. The dashed lines display the mean threshold value. Bottom row: Observed adjacency-based CUSUM statistics (5) (with  $l = 1, 2, 3, 4$  and  $\Lambda = 100$ ) for 10 MC runs.

Table 4: Experiment II-(b): Mean Relative Error

Method	$Td = 400$	$Td = 200$	$Td = 40$
Bhattacharjee et al. (2020)	0	0.0005	0.249
Bhattacharyya et al. (2020)	0.297	0.28	0.3085
Wang et al. (2021)	0.375	0.375	0.375
Zhao et al. (2019)	0.082	0.102	0.3825
Our Method	0	0	0.0125

Table 5: Experiment II-(c): Mean Relative Error

Method	$Td = 180$	$Td = 90$	$Td = 18$
Bhattacharjee et al. (2020)	0.0015	0.0015	0.0895
Bhattacharyya et al. (2020)	0.15	0.1705	0.1325
Wang et al. (2021)	0.375	0.375	0.375
Zhao et al. (2019)	0.156	0.1535	0.375
Our Method	0.0015	0.0005	0.0795

Table 6: Experiment II-(d): Mean Relative Error

Method	$Td = 200$	$Td = 100$	$Td = 20$
Bhattacharjee et al. (2020)	0.011	0.0165	0.2635
Bhattacharyya et al. (2020)	0.328	0.293	0.302
Wang et al. (2021)	0.375	0.375	0.375
Zhao et al. (2019)	0.1375	0.162	0.375
Our Method	0.0015	0.003	0.0705

Table 7: Wild Binary Segmentation

Threshold	Detected Changepoints
$60\sqrt{\left(\frac{K}{n}\right)^3 \frac{\hat{d}}{T}}$	62
$50\sqrt{\left(\frac{K}{n}\right)^3 \frac{\hat{d}}{T}}$	40, 62, 125
$40\sqrt{\left(\frac{K}{n}\right)^3 \frac{\hat{d}}{T}}$	40, 62, 71, 116, 125, 133, 158

Table 8: Time-Period Corresp. to the Detected Changepoints

Detected Changepoints	Time-Period
40	1986/09/16 - 1986/10/16
62	1992/09/10 - 1993/02/04
71	1994/07/13 - 1994/08/10
116	2004/05/04 - 2004/09/14
125	2006/07/12 - 2007/01/17
133	2008/05/14 - 2009/01/22
158	2014/12/01 - 2014/12/13