Journal of Machine Learning Research 21 (2020) 1-59 Submitted 12/18; Revised 4/20; Published 4/20

Change Point Estimation in a Dynamic Stochastic Block Model

Monika Bhattacharjee M.BHATTACHARJEE@UFL.EDU
Informatics Institute
University of Florida
Gainesville, USA

Moulinath Banerjee MOULIB@UMICH.EDU
Department of Statistics
University of Michigan
Ann Arbor, USA

George Michailidis GMICHAIL@UFL.EDU
Informatics Institute
Department of Statistics
University of Florida
Gainesville, USA

Editor: Jie Peng

Abstract

We consider the problem of estimating the location of a single change point in a network generated by a dynamic stochastic block model mechanism. This model produces community structure in the network that exhibits change at a single time epoch. We propose two methods of estimating the change point, together with the model parameters, before and after its occurrence. The first employs a least-squares criterion function and takes into consideration the full structure of the stochastic block model and is evaluated at each point in time. Hence, as an intermediate step, it requires estimating the community structure based on a clustering algorithm at every time point. The second method comprises the following two steps: in the first one, a least-squares function is used and evaluated at each time point, but *ignoring the community structure* and only considering a random graph generating mechanism exhibiting a change point. Once the change point is identified, in the second step, all network data before and after it are used together with a clustering algorithm to obtain the corresponding community structures and subsequently estimate the generating stochastic block model parameters. The first method, since it requires knowledge of the community structure and hence clustering at every point in time, is significantly more computationally expensive than the second one. On the other hand, it requires a significantly less stringent identifiability condition for consistent estimation of the change point and the model parameters than the second method; however, it also requires a condition on the misclassification rate of misallocating network nodes to their respective communities that may fail to hold in many realistic settings. Despite the apparent stringency of the identifiability condition for the second method, we show that networks generated by a stochastic block mechanism exhibiting a change in their structure can easily satisfy this condition under a multitude of scenarios, including merging/splitting communities, nodes joining another community, etc. Further, for both methods under their respective identifiability and certain additional regularity conditions, we establish rates of convergence and derive the asymptotic distributions of the change point estimators. The results are illustrated on synthetic data. In summary, this work provides an in-depth investigation of the novel problem of change point analysis for networks generated by stochastic block models, identifies key conditions for the consistent estimation of the change point, and proposes a computationally fast algorithm that solves the problem in many settings that occur in applications. Finally, it discusses challenges posed by employing clustering algorithms in this problem, that require additional investigation for their full resolution.

Keywords: stochastic block model, Erdős-Rényi random graph, change point, edge probability matrix, community detection, estimation, clustering algorithm, convergence rate

1. Introduction

The modeling and analysis of network data have attracted the attention of multiple scientific communities, due to their ubiquitous presence in many application domains; see Newman et al. (2006), Kolaczyk and Csárdi (2014), Crane (2018) and references therein. A popular and widely used statistical model for network data is the Stochastic Block Model (SBM) introduced in Holland et al. (1983). It is a special case of a random graph model, where the nodes are partitioned into K disjoint groups (communities) and the edges between them are drawn independently with probabilities that only depend on the community membership of the nodes. This leads to a significant reduction in the dimension of the parameter space, from $\mathcal{O}(m^2)$ for the random graph model, with m being the number of nodes in the network, to $\mathcal{O}(K^2)$ ($K << m$).

There has been a lot of technical work on the SBM, including (i) estimation of the underlying community structure and the corresponding community connection probabilities, for examples Choi and Wolfe (2014); Jin (2015); Joseph and Yu (2016); Lei and Rinaldo (2015); Rohe et al. (2011); Sarkar and Bickel (2015); Zhao et al. (2012); (ii) establishing the minimax rate for estimating the SBM parameters—for examples Gao et al. (2015); Klopp et al. (2017)—and the community structure—for examples Zhang and Zhou (2016); Gao et al. (2017)—under the assumption that the assignment problem of nodes to communities can be solved *exactly*. However, the latter problem is computationally NP-hard and hence estimates of the community structure and connection probabilities based on easy to compute procedures compromise the minimax rate—see Zhang et al. (2015).

There has also been some recent work on understanding the evolution of community structure over time, based on observing a sequence of network adjacency matrices—for examples Durante et al. (2017); Durante and Dunson (2016); Han et al. (2015); Kolar et al. (2010); Matias and Miele (2017); Minhas et al. (2015); Xing et al. (2010); Xu (2015); Yang et al. (2011); Bao and Michailidis (2018). Various modeling formalism is employed including Markov random field models, low rank plus sparse decompositions and dynamic versions of SBM (DSBM). These studies focus primarily on fast and scalable computational procedures for identifying the evolution of community structure over time. Some work that investigated theoretical properties of the DSBM and more generally graphon models assuming that the node assignment problem can be solved exactly includes Pensky (2019), while the theoretical performance of spectral clustering for the DSBM was examined in Bhattacharyya and Chatterjee (2017) and Pensky and Zhang (2019). The last two studies estimate the edge probability matrices by either directly averaging adjacency matrices observed at different time points, or by employing a kernel-type smoothing procedure and extract the group memberships of the nodes by using spectral clustering.

The objective of this paper is to examine the *offline estimation* of a single change point under a DSBM network generating mechanism. Specifically, a sequence of networks is generated independently across time through the SBM mechanism, whose community connection probabilities exhibit a change at some time epoch. Then, the problem becomes one of identifying the change point epoch based on the observed sequence of network adjacency matrices, detecting the community structures before and after the change point, and estimating the corresponding SBM model parameters.

The existence of change points and their estimation has been well-studied for many univariate statistical models evolving independently over time and with shifts in mean and in variance structures. A broad overview of the corresponding literature can be found in Brodsky and Darkhovsky (2013) and Csörgö and Horváth (1997). However, in many applications, multivariate (even high dimensional) signals are involved, while also exhibiting dependence across time—see the review article by Alexander and Horváth (2013) and references therein.

The emergence of data with network structure has accentuated the need to study change point problems in that context. For example, in political science, the study of political polarization has received increased attention (Moody and Mucha (2013)) and especially the detection of regime changes (Peel and Clauset (2015)), including network-based approaches (Bao and Michailidis (2018)). Recent work on change-point detection for network structured data includes Wang et al. (2017) and Wang et al. (2018). Specifically, Wang et al. (2017) considers a generalized hierarchical random graph model, whereas Wang et al. (2018) focuses on change-point detection in sparse networks. The latter study deals with the Erdős-Rényi random graph model. However, to the best of our knowledge, a detailed and rigorous investigation of change-point detection for the dynamic SBM is largely lacking.

Therefore, the key contributions of this paper are threefold: first, the development of a computational strategy for solving the problem and establishing its theoretical properties under suitable regularity conditions, including (i) establishing the rate of convergence for the least-squares estimate of the change-point and (ii) the DSBM parameters, as well as (iii) deriving the asymptotic distribution of the change-point. An important step in the strategy for obtaining an estimate of the change-point involves clustering the nodes to communities, for which we employ a spectral clustering algorithm that exhibits cubic computational cost in the number of edges in the adjacency matrix. However, the theoretical analysis of the first method which involves clustering *at every time point* requires imposing a rather stringent assumption on the rate of misclassifying nodes to communities. For these reasons, the second key contribution of this work is the introduction of a two-step computational strategy, wherein the first step, the change-point is estimated based on a procedure that *ignores* the community structure, while in the second step the pre- and post-change-point model parameters are estimated using a spectral clustering algorithm, but at *a single time point*. It is established that this strategy yields consistent estimates for the change-point and the community connection probabilities, at a linear computational cost in the number of edges. However, the procedure requires a stronger identifiability condition compared to the first strategy. Naturally, no additional condition of controlling the rate of misclassifying nodes to communities during the first step is required. The third contribution of the paper is to show that the more stringent identifiability condition under the second strategy is easily satisfied in a number of scenarios by the DSBM, including splitting/merging communities and reallocating nodes to other communities before and after the change-point. Overall, this work provides valuable insights into the technical challenges of change-point analysis for DSBM and also an efficient computational strategy that delivers consistent estimates of all the model parameters. Nevertheless, the challenges identified require further investigation for their complete resolution, as discussed in Section 7.5.

The remainder of the paper is organized as follows. In Section 2, we introduce the DSBM model together with the necessary definitions and notation for technical development. Subsequently, we present the strategy to detect the change-point that involves a *community detection step at each time point*, followed by estimation of the DSBM model parameters together with the asymptotic properties of the estimators. In Section 3, we introduce a 2-step computational strategy for the DSBM change-point detection problem, which is computationally significantly less expensive and discuss the consistency of these estimators. Section 4 involves a comparative study between the two change-point detection strategies previously presented and also provides many realistic settings where the computationally fast 2-step algorithm is provably consistent. The numerical performance of the two strategies based on synthetic data is illustrated in Section 4.1. We briefly discuss other community detection methods for DSBM involving a single change-point in Section 7.5. Finally, the asymptotic distribution of the change-point estimates along with a data-driven procedure for identifying the correct limiting regime is presented in Section 5. Some concluding remarks are drawn in Section 6. All proofs and additional technical material are presented in Section 7.

2. The dynamic stochastic block model (DSBM)

The structure of the SBM is determined by the following parameters: (i) m the number of nodes or vertices in the network, (ii) a symmetric $K \times K$ matrix $\Lambda = ((\lambda_{ij}))_{K\times K}$ containing the community connection probabilities and (iii) a partition of the node set $\{1, 2, \ldots, m\}$ into K communities, which is represented by a many-to-one onto map $z : \{1, 2, \ldots, m\} \to \{1, 2, \ldots, K\}$ for some $K \leq m$. Hence, for each $1 \leq i \leq m$, the community of node i is determined by $z(i)$, or equivalently

$$l\text{-th community} = \mathcal{C}_l = \{i \in \{1, 2, \ldots, m\} : z(i) = l\} \ \forall 1 \leq l \leq K.$$

The map z determines the community structure under the SBM. The observed edge set of the network is obtained as follows: any two nodes $i \in \mathcal{C}_l$ and $j \in \mathcal{C}_{l'}$ are connected by an edge with probability $\lambda_{ll'} I(i \neq j)$, independent of any other pair of nodes. Self-loops are not considered and hence the probability of having an edge between nodes i and j is 0 whenever $i = j$. Henceforth, we use $\text{SBM}(z, \Lambda)$ for denoting an SBM with community structure z and community connection probability matrix Λ. Next, let

$$\text{Ed}_z(\Lambda) = ((\lambda_{z(i)z(j)} I(i \neq j)))_{m \times m},$$

which is the edge probability matrix whose (i, j)-th entry represents the probability of having an edge between nodes i and j. Note that we are dealing with undirected networks and thus Λ and $\text{Ed}_z(\Lambda)$ are symmetric matrices.

The data come in the form of an observed square symmetric matrix $A = ((A_{ij}))_{m \times m}$ with entries

$$A_{ij} = \begin{cases} 1, & \text{if an edge between nodes } i \text{ and } j \text{ is observed} \\ 0, & \text{otherwise.} \end{cases}$$

An adjacency matrix A is said to be generated according to $\text{SBM}(z, \Lambda)$, if $A_{ij} \sim$ $\text{Bernoulli}(\Lambda_{z(i)z(j)} I(i \neq j))$ independently, and is denoted by $A \sim \text{SBM}(z, \Lambda)$. It is easy to see that all diagonal entries of A are 0.

In a DSBM, we consider a sequence of stochastic block models evolving *independently* over time, with $A_{t,n} = ((A_{ij,(t,n)}))_{m \times m}$ denoting the adjacency matrix at time point t. Hence, $A_{t,n} \sim \text{SBM}(z_t, \Lambda_t)$ independently, with n being the total number of time points available, and note $\{A_{t,n} : 1 \leq t \leq n, n \geq 1\}$ forms a triangular sequence of adjacency matrices. Further, in the technical analysis we assume that both the total number of time points and the number of nodes are growing to infinity, that is $n, m \to \infty$. Moreover, the number of communities depends on both the number of nodes and time, and grows to infinity, that is, $K = K_{m,n} \to \infty$ as $m, n \to \infty$.

We are interested in a DBSM exhibiting a single change-point and its estimation in an offline manner. For presenting the results, we embed all the time points in the $[0, 1]$ interval, by dividing them by n. Suppose $\tau_{m,n} \in (c^*, 1 - c^*)$ corresponds to the change-point epoch in the DSBM. Hence,

$$A_{t,n} \sim \begin{cases} \text{SBM}(z, \Lambda), & \text{if } 1 \leq t \leq \lfloor n\tau_{m,n} \rfloor \\ \text{SBM}(w, \Delta), & \text{if } \lfloor n\tau_{m,n} \rfloor < t < n, \end{cases} \tag{2.1}$$

and $z \neq w$ and/or $\Lambda \neq \Delta$. Note that z and w correspond to the pre- and post-change-point community structures, respectively. Similarly Λ and Δ are the pre- and post-change-point community connection probability matrices, respectively. Further, note that z, w, Λ and Δ may depend on m and n.

Our objective is to estimate the model parameters $\tau_{m,n}, z, w, \Lambda$ and Δ. Throughout this paper, we assume that $0 < c^* < \tau_{m,n} < 1 - c^* < 1$ with c^* being known to avoid unnecessary technical complications if the true change-point is located arbitrarily close to boundary time points. We also assume that the total number of communities before and after the change-point are equal. *Even if they are different, our results continue to hold after replacing K by the maximum of the number of communities to the left and the right.*

To identify the change-point, we employ the following least-squares criterion function

$$\tilde{L}(b,z,w,\Lambda,\Delta) \;=\; \frac{1}{n}\sum_{i,j=1}^{m}\left[\sum_{t=1}^{nb}(A_{ij,(t,n)}-\lambda_{z(i)z(j)})^2+\sum_{t=nb+1}^{n}(A_{ij,(t,n)}-\delta_{w(i)w(j)})^2\right]. \quad (2.2)$$

Remark 1 *Note that the Bernoulli likelihood function criterion could also be used that will yield similar results for the change-point estimators, but it will need stronger assumptions and will involve more technicalities compared to the least-squares criterion function adopted. More details on the likelihood criterion function for detecting a change-point in a random graph model can be found in Bhattacharjee et al. (2017). Results on the maximum likelihood estimator of the change-point in a random graph model are consequences of the results in Bhattacharjee et al. (2017). However, in DSBM one also needs to address the problem of assigning nodes to their respective communities, which makes the problem technically more involved as shown next. However, the main message of this paper will remain the same, irrespective of employing a likelihood or a least-squares criterion.*

Let

$$\begin{aligned} S_{u,z} &= \{i\in\{1,2\ldots,m\}: z(i)=u\}, \quad s_{u,z}=|S_{u,z}|, \\ S_{u,w} &= \{i\in\{1,2\ldots,m\}: w(i)=u\}, \quad s_{u,w}=|S_{u,w}| \end{aligned} \quad (2.3)$$

denote the u-th block and block size under the community structures z and w. Also define,

$$\begin{aligned} \tilde{\Lambda}_{z,(b,n),m} &= ((\tilde{\lambda}_{uv,z,(b,n),m}))_{m\times m}, \quad \tilde{\lambda}_{uv,z,(b,n),m}=\frac{1}{nb}\frac{1}{s_{u,z}s_{v,z}}\sum_{t=1}^{nb}\sum_{\substack{i:z(i)=u\\ j:z(j)=v}}A_{ij,(t,n)}, \\ \tilde{\Delta}_{w,(b,n),m} &= ((\tilde{\delta}_{uv,w,(b,n),m}))_{m\times m}, \quad \tilde{\delta}_{uv,w,(b,n),m}=\frac{(n(1-b))^{-1}}{s_{u,w}s_{v,w}}\sum_{t=nb+1}^{n}\sum_{\substack{i:w(i)=u\\ j:w(j)=v}}A_{ij,(t,n)}. \end{aligned} \quad (2.4)$$

Define the *sparsity parameter* $\rho_{m,n}=\max_{1\le u\le v\le K}\{\Lambda_{u,v},\Delta_{u,v}\}$ which may depend on both m and n. The DSBM in (2.1) is dense if $\inf_{m,n}\rho_{m,n}>C>0$ and sparse if $\rho_{m,n}\to 0$ as $m,n\to\infty$.

We start our analysis by assuming that the **community structures** z and w are **known**. In that case, an estimate of the change-point can be obtained by solving

$$\tilde{\tau}_{m,n} \;=\; \arg\min_{b\in(c^*,1-c^*)}\tilde{L}(b,z,w,\tilde{\Lambda}_{z,(b,n),m},\tilde{\Delta}_{w,(b,n),m}). \quad (2.5)$$

The following signal-to-noise condition guarantees that the change-point is identifiable under a known community structure.

SNR-DSBM: $\frac{n}{K^2\rho_{m,n}}||\mathrm{Ed}_z(\Lambda)-\mathrm{Ed}_w(\Delta)||_F^2\to\infty$

Intuitively, it implies that the signal per connection probability parameter, after scaling by the sparsity parameter, needs to grow faster than $1/\sqrt{n}$, which is in accordance with identifiability conditions for other change-point problems (for example see Kosorok (2008) Section 14.5.1).

The following assumption on the sparsity parameter restricts its rate of decay and ensures consistency of $\tilde{\tau}_{m,n}$ scaled by the sparsity parameter with details given in Remark 3.

Sparse-DSBM: $\rho_{m,n} > CK^{-2}$ for some $C > 0$.

It is easy to see that Sparse-DSBM is always satisfied by a dense DSBM.

The following Theorem establishes asymptotic properties for $\tilde{\tau}_{m,n}$. Its proof is similar (albeit much simpler in structure) to the proof of Theorem 11, where we deal with unknown community structures. Hence, it is omitted.

Theorem 2 ***(Convergence rate of*** $\tilde{\tau}_{m,n}$***)*** *Suppose SNR-DSBM and Sparse-DSBM hold. Then,*

$$n\rho_{m,n}^{-1}||Ed_z(\Lambda) - Ed_w(\Delta)||_F^2(\tilde{\tau}_{m,n} - \tau_{m,n}) = O_P(1).$$

Remark 3 *As stated in Theorem 2, the convergence rate of the appropriately centered and scaled change-point estimator* $\rho_{m,n}^{-1}(\tilde{\tau}_{m,n}-\tau_{m,n})$ *is* $n^{-1}||Ed_z(\Lambda)-Ed_w(\Delta)||_F^{-2}$*. In the presence of SNR-DSBM, Sparse-DSBM ensures that the convergence rate* $n^{-1}||Ed_z(\Lambda) - Ed_w(\Delta)||_F^{-2}$ *decays to* 0*. Sparse-DSBM can be weakened further under stronger signal-to-noise condition, for example see Remark 21.*

Remark 4 *In the ensuing Theorem 11, we will establish that the DSBM change-point estimator* $\tilde{\tau}_{m,n}$ *with an* unknown community structure *(that needs to be estimated from the available data) has exactly the same convergence rate as the one posited in Theorem 2. However, a much stronger identifiability condition than SNR-DSBM and Sparse-DSBM is needed, since more parameters are involved.*

Recall the estimates in (2.4) given by $\tilde{\Lambda} = ((\tilde{\lambda}_{ab,z,(\tilde{\tau}_{m,n},n),m}))_{K\times K}$ and $\tilde{\Delta} = ((\tilde{\delta}_{ab,w,(\tilde{\tau}_{m,n},n),m}))_{K\times K}$. The edge probability matrices $\mathrm{Ed}_z(\Lambda)$ and $\mathrm{Ed}_w(\Delta)$ can also be estimated by $\mathrm{Ed}_z(\tilde{\Lambda})$ and $\mathrm{Ed}_w(\tilde{\Delta})$, respectively. The following Theorem provides the convergence rate of the corresponding estimators. Its proof is similar (and structurally simpler) to the proof of Theorem 13 where we deal with unknown community structures and hence omitted.

Theorem 5 ***(Convergence rate of edge probabilities when*** z ***and*** w ***are known)*** *Suppose SNR-DSBM holds. Let* $\mathcal{S}_{m,n} = \min(\min_u s_{u,z}, \min_u s_{u,w})$*. Then*

$$\frac{1}{K^2}||\tilde{\Lambda} - \Lambda||_F^2, \quad \frac{1}{K^2}||\tilde{\Delta} - \Delta||_F^2 = O_P\left(\frac{I(n>1)\rho_{m,n}^2}{n^2||Ed_z(\Lambda) - Ed_w(\Delta)||_F^4} + \frac{\rho_{m,n}\log K}{n\mathcal{S}_{m,n}^2}\right),$$

$$\frac{1}{m^2}||Ed_z(\tilde{\Lambda}) - Ed_z(\Lambda)||_F^2, \quad \frac{1}{m^2}||Ed_w(\tilde{\Delta}) - Ed_w(\Delta)||_F^2$$

$$= O_P\left(\frac{I(n>1)\rho_{m,n}^2}{n^2||Ed_z(\Lambda)-Ed_w(\Delta)||_F^4}+\frac{\rho_{m,n}\log m}{n\mathcal{S}_{m,n}^2}\right). \tag{2.6}$$

Note that $\tilde{\Lambda} = ((\tilde{\lambda}_{ab,z,(\tilde{\tau}_{m,n},n),m}))_{K\times K}$. To compute the rate for $\frac{1}{K^2}||\tilde{\Lambda}-\Lambda||_F^2$, we have $(\tilde{\lambda}_{ab,z,(\tilde{\tau}_{m,n},n),m}-\lambda_{ab})^2 \leq 2(\tilde{\lambda}_{ab,z,(\tilde{\tau}_{m,n},n),m}-\tilde{\lambda}_{ab,z,(\tau_{m,n},n),m})^2+2(\tilde{\lambda}_{ab,z,(\tau_{m,n},n),m}-\lambda_{ab})^2 \equiv T_1+T_2$. It is easy to see that the first term T_1 is dominated by $(\tilde{\tau}_{m,n}-\tau_{m,n})^2$ and thus by Theorem 2, $(\tilde{\tau}_{m,n}-\tau_{m,n})^2 = O_{\mathrm{P}}\left(\frac{I(n>1)\rho_{m,n}^2}{n^2||\mathrm{E}d_z(\Lambda)-\mathrm{E}d_w(\Delta)||_F^4}\right)$. Moreover, the rate of T_2 is $\frac{\rho_{m,n}\log K}{n\mathcal{S}_{m,n}^2}$. Details are given in Section 7.4. Similar arguments work for the other matrices presented in Theorem 5.

Remark 6 ***(Rate for*** $n=1$***).*** *If* $n=1$*, then there is no change-point and* T_1 *does not appear. In this case, we have only one community structure (say)* z *and one community connection matrix (say)* Λ*. Moreover, the number of communities* $K=K_m$*, sparsity parameter* $\rho_m=\max_{1\leq u,v\leq K}\lambda_{uv}$ *and the minimum block size* $\mathcal{S}_m=\min_u s_{u,z}$ *depend only on* m*. Estimation of* Λ *for* $n=1$ *is studied in Zhang et al. (2015) when the community structure* z *is unknown. In this remark, we assume that* z *is known. We estimate* Λ *by* $\tilde{\Lambda}=((\tilde{\lambda}_{ab,z,(1/n,n),m}))_{K\times K}$*. Then,*

$$\frac{1}{K^2}||\tilde{\Lambda}-\Lambda||_F^2 = O_P\left(\frac{\rho_m\log K}{\mathcal{S}_m^2}\right),\quad \frac{1}{m^2}||Ed_z(\tilde{\Lambda})-Ed_z(\Lambda)||_F^2 = O_P\left(\frac{\rho_m\log m}{\mathcal{S}_m^2}\right).$$

Similar results for the case of unknown communities are discussed in Remark 15 and Section 7.5.

In Theorem 13, we establish the results for the same quantities in the case of unknown community structures. It will be seen that the convergence rate of $\tilde{\Lambda}$ and $\tilde{\Delta}$ given above, is much sharper compared to the case of unknown community structures, despite using repeated observations in the latter one; see also discussion in Remark 14.

The real problem of interest is when the community structure is **unkown** and needs to be estimated from the observed sequence of adjacency matrices along with the change-point. A standard strategy in the change-point analysis is to optimize the least-squares criterion function $\tilde{L}(b,z,w,\Lambda,\Delta)$ posited above with respect to *all* the model parameters. This becomes challenging both computationally since one needs to find a good assignment of nodes to communities, and technically, since for any time point away from the true change-point the node assignment problem needs to be solved under a misspecified model; namely, the available adjacency matrices are generated according to both the pre- and post-change-point community connection probability matrices.

A natural estimator of $\tau_{m,n}$ can be obtained by solving

$$\tilde{\tilde{\tau}}_{m,n} \quad = \quad \arg\min_{b\in(c^*,1-c^*)} \tilde{L}(b,\tilde{z}_{b,n,m},\tilde{w}_{b,n,m},\tilde{\Lambda}_{\tilde{z}_{b,n,m},(b,n),m},\tilde{\Delta}_{\tilde{w}_{b,n,m},(b,n),m}), \tag{2.7}$$

where $\tilde{z}_{b,n,m}$ and $\tilde{w}_{b,n,m}$ are obtained using the clustering algorithm from Bhattacharyya and Chatterjee (2017) (details below). While other clustering algorithms can also be employed, and are discussed in Section 7.5, all clustering algorithms incur some degree of misclassification (while assigning nodes to communities) which must be suitably controlled by an appropriate assumption. The employed clustering algorithm requires a simpler and somewhat easier assumption on the misclassification rate, compared to other available clustering methods.

Clustering Algorithm I: (proposed in Bhattacharyya and Chatterjee (2017))

1. Obtain sums of the adjacency matrices before and after b as $B_1 = \sum_{t=1}^{nb} A^{(t)}$ and $B_2 = \sum_{t=nb+1}^{n} A^{(t)}$ respectively.

2. Obtain $\hat{U}_{m\times K}$ and $\hat{V}_{m\times K}$ consisting of the leading K eigenvectors of B_1 and B_2, respectively, corresponding to its largest absolute eigenvalues.

3. Use an $(1+\epsilon)$ approximate K-means clustering algorithm on the row vectors of $\hat{U}$ and $\hat{V}$ to obtain $\tilde{z}_{b,n,m}$ and $\tilde{w}_{b,n,m}$ respectively.

Note that in Step 3 above, an $(1+\epsilon)$ approximate K-means clustering procedure is employed, instead of the K-means. It is known that finding a global minimizer for the K-means clustering problem is NP-hard (for example see Aloise et al. (2009)). However, efficient algorithms such as $(1+\epsilon)$ approximate K-means clustering provide an approximate solution, with the value of the objective function being minimized to within a constant fraction of the optimal value (see Kumar et al. (2004) for more details).

Computational complexity of the procedure: Note that for each $b \in (c^*, 1-c^*)$, the complexity of computing $\tilde{z}_{b,n,m}$ (or $\tilde{w}_{b,n,m}$) and $\tilde{\Lambda}_{\tilde{z}_{b,n,m},(b,n),m}$ (or $\tilde{\Delta}_{\tilde{w}_{b,n,m},(b,n),m}$) is $O(m^3)$ and $O(m^2 n)$, respectively. Hence, $\tilde{L}(b, \tilde{z}_{b,n,m}, \tilde{w}_{b,n,m}, \tilde{\Lambda}_{\tilde{z}_{b,n,m},(b,n),m}, \tilde{\Delta}_{\tilde{w}_{b,n,m},(b,n),m})$ at b has computational complexity $O(m^3+m^2 n)$. Some calculations show that only finitely many binary operations are needed to update $\tilde{\Lambda}_{\tilde{z}_{b,n,m},(b,n),m}$ and $\tilde{\Delta}_{\tilde{w}_{b,n,m},(b,n),m}$ for the next available time point. However, computing $\tilde{z}_{b,n,m}$ and $\tilde{w}_{b,n,m}$ requires $O(m^3)$ operations for each time point. Therefore, the computational complexity for obtaining $\tilde{L}(b, \tilde{z}_{b,n,m}, \tilde{w}_{b,n,m}, \tilde{\Lambda}_{\tilde{z}_{b,n,m},(b,n),m}, \tilde{\Delta}_{\tilde{w}_{b,n,m},(b,n),m})$ for n-many time points is $O(m^3 n + m^2 n) = O(m^3 n)$.

To establish consistency results for $\tilde{\tilde{\tau}}_{m,n}$, an additional assumption on the misclassification rate of $\tilde{z}_{b,n,m}$ and $\tilde{w}_{b,n,m}$ is needed, given next. We start with the following definition.

Definition 7 ***(Misclassification rate)*** *A node is considered as misclassified, if it is not allocated to its true community. The misclassification rate corresponds to the fraction of misclassified nodes. Let $\mathcal{M}_{(z,\tilde{z}_{b,n,m})}$ and $\mathcal{M}_{(w,\tilde{w}_{b,n,m})}$ be the misclassification rates of estimating z and w by $\tilde{z}_{b,n,m}$ and $\tilde{w}_{b,n,m}$, respectively. Then,*

$$\begin{aligned} \mathcal{M}_{(z,\tilde{z}_{b,n,m})} &= \min_{\pi\in S_k} \sum_{i=1}^{m} \frac{I(\tilde{z}_{b,n,m}(i) \neq \pi(z(i)))}{s_{z,\pi(z(i))}}, \\ \mathcal{M}_{(w,\tilde{w}_{b,n,m})} &= \min_{\pi\in S_k} \sum_{i=1}^{m} \frac{I(\tilde{w}_{b,n,m}(i) \neq \pi(w(i)))}{s_{w,\pi(w(i))}} \end{aligned} \tag{2.8}$$

where S_K denotes the set of all permutations of $\{1, 2, \ldots, K\}$.

Define

$$\mathcal{M}_{b,n,m} = \max(\mathcal{M}_{(z,\tilde{z}_{b,n,m})}, \mathcal{M}_{(w,\tilde{w}_{b,n,m})}).$$

Consider the following assumption.

(NS) Λ and Δ are non-singular.

(NS) implies that there are exactly K non-empty communities in DSBM and hence we can use an $(1+\epsilon)$ approximate K-clustering algorithm. If (NS) does not hold, then we have K' ($< K$) non-empty communities and an $(1+\epsilon)$ approximate K'-clustering algorithm performs better.

The following theorem provides the convergence rate of $\mathcal{M}_{b,n}$. Its proof is given in Section 7.1. Let $\nu_{m,n}$ denote the minimum between the smallest non-zero singular values of $\mathrm{Ed}_z(\Lambda)$ and $\mathrm{Ed}_w(\Delta)$.

Theorem 8 *Suppose (NS) holds. Then, for all $b \in (c^*, 1-c^*)$, we have*

$$\mathcal{M}_{b,n,m} = O_P\left(\frac{K}{n\nu_{m,n}^2}(\tau_{m,n}m + |\tau_{m,n} - b|\ ||Ed_z(\Lambda) - Ed_w(\Delta)||_F^2)\right).$$

Remark 9 *$\nu_{m,n}$ becomes closer to 0 (that is we have less signal) as the regime becomes more sparse. Thus sparse regime has high misclassification rate for community detection.*

Remark 10 *To establish consistency of $\tilde{\tilde{\tau}}_{m,n}$, we require that the misclassification rate $\mathcal{M}_{b,n,m}$ decays faster than $\rho_{m,n}^{-1}n^{-1}||Ed_z(\Lambda) - Ed_w(\Delta)||_F$; see the proof of Theorem 11 and Remark 35 for technical details. By the identifiability condition SNR-DSBM and Theorem 8, we have*

$$\mathcal{M}_{b,n,m}\rho_{m,n}n||Ed_z(\Lambda) - Ed_w(\Delta)||_F^{-1} \leq C\frac{K\rho_{m,n}}{\nu_{m,n}^2}(\frac{m\sqrt{n}}{K}o(1) + m) \leq C\rho_{m,n}(\frac{m\sqrt{n}}{\nu_{m,n}^2}o(1) + \frac{Km}{\nu_{m,n}^2})$$

holds with probability tending to 1. Consistency of $\tilde{\tilde{\tau}}_{m,n}$ can be achieved under the SNR-DSBM condition and the following assumption (A1).

(A1) $\frac{Km}{\nu_{m,n}^2}\rho_{m,n} \to 0$ and $\frac{m\sqrt{n}}{\nu_{m,n}^2}\rho_{m,n} = O(1)$

We note that (A1) is compatible with the clustering algorithm employed in our procedure. Other clustering algorithms may also be used which would lead to modifications of (A1), as discussed in Section 7.5.

2.1. Theoretical properties of $\tilde{\tilde{\tau}}_{m,n}$

Our first result establishes the convergence rate of the proposed estimate of the change-point.

Theorem 11 ***(Convergence rate of $\tilde{\tilde{\tau}}_{m,n}$)***
Suppose SNR-DSBM, Sparse-DSBM, (NS) and (A1) hold. Then,

$$n\rho_{m,n}^{-1}||Ed_z(\Lambda) - Ed_w(\Delta)||_F^2(\tilde{\tilde{\tau}}_{m,n} - \tau_{m,n}) = O_P(1).$$

The proof of the theorem is given in Section 7.3.

The next result focuses on the misclassification rate for $\tilde{\tilde{z}} = \tilde{z}_{\tilde{\tilde{\tau}}_{m,n},n,m}$ and $\tilde{\tilde{w}} = \tilde{w}_{\tilde{\tilde{\tau}}_{m,n},n,m}$, respectively.

Theorem 12 ***(Rate of misclassification)***
Suppose SNR-DSBM, Sparse-DSBM, (NS) and (A1) hold. Then,

$$\mathcal{M}_{(z,\tilde{\tilde{z}})}, \mathcal{M}_{(w,\tilde{\tilde{w}})} = O_P\left(\frac{Km}{n\nu_{m,n}^2}\right).$$

The proof of the Theorem is immediate from Theorems 8 and 11.

Let $\tilde{\tilde{\Lambda}} = ((\tilde{\lambda}_{ab,\tilde{\tilde{z}},(\tilde{\tilde{\tau}}_{m,n},n),m}))_{K\times K}$ and $\tilde{\tilde{\Delta}} = ((\tilde{\delta}_{ab,\tilde{\tilde{w}},(\tilde{\tilde{\tau}}_{m,n},n),m}))_{K\times K}$. The final result obtained is on the convergence rate of the community connection probability matrices $\tilde{\tilde{\Lambda}}$ and $\tilde{\tilde{\Delta}}$, respectively. Let $\mathcal{S}_{m,n,\tilde{\tilde{z}}} = \min_u s_{u,\tilde{\tilde{z}}}$, $\mathcal{S}_{m,n,\tilde{\tilde{w}}} = \min_u s_{u,\tilde{\tilde{w}}}$ and $\tilde{\mathcal{S}}_{m,n} = \min(\mathcal{S}_{n,\tilde{\tilde{z}}}, \mathcal{S}_{n,\tilde{\tilde{w}}})$.

Theorem 13 ***(Convergence rate of edge probabilities $\tilde{\tilde{\Lambda}}$ and $\tilde{\tilde{\Delta}}$)***
Suppose SNR-DSBM, Sparse-DSBM, (NS) and (A1) hold. Further, for some positive sequence $\{\tilde{C}_{m,n}\}$, we have that $\tilde{\mathcal{S}}_{m,n} \geq \tilde{C}_{m,n}$ $\forall m,n$ with probability 1. Then,

$$\begin{aligned}
\frac{1}{m^2}||Ed_{\tilde{\tilde{z}}}(\tilde{\tilde{\Lambda}}) - Ed_z(\Lambda)||_F^2 &= O_P\left(\left(\frac{Km}{n\nu_{m,n}^2}\right)^2 + \frac{I(n>1)\rho_{m,n}^2}{n^2||Ed_z(\Lambda) - Ed_w(\Delta)||_F^4} + \frac{\log m}{n\tilde{C}_{m,n}^2}\rho_{m,n}\right), \\
\frac{1}{m^2}||Ed_{\tilde{\tilde{w}}}(\tilde{\tilde{\Delta}}) - Ed_w(\Delta)||_F^2 &= O_P\left(\left(\frac{Km}{n\nu_{m,n}^2}\right)^2 + \frac{I(n>1)\rho_{m,n}^2}{n^2||Ed_z(\Lambda) - Ed_w(\Delta)||_F^4} + \frac{\log m}{n\tilde{C}_{m,n}^2}\rho_{m,n}\right).
\end{aligned}$$

The proof of the Theorem is given in Section 7.4.

Remark 14 *Note that the first term in the convergence rate of $Ed_{\tilde{\tilde{z}}}(\tilde{\tilde{\Lambda}})$, which is the square of the misclassification rate obtained in Theorem 12, measures the closeness of $Ed_{\tilde{\tilde{z}}}(\tilde{\tilde{\Lambda}})$ to $Ed_z(\tilde{\tilde{\Lambda}})$. On the other hand, the second term is the convergence rate of $Ed_z(\tilde{\tilde{\Lambda}})$ for $Ed_z(\Lambda)$ and coincides with the convergence rate of the edge probability matrix estimator when the communities are known—see Theorem 5 for details.*

As expected, the convergence rate of $\tilde{\tilde{\Lambda}}$ and $\tilde{\tilde{\Delta}}$, given in Theorem 13, is slower than the rate of $\tilde{\Lambda}$ and $\tilde{\Delta}$ when the communities are known. The reason is that the former estimates involve the misclassification rate of estimating z and w by $\tilde{\tilde{z}}$ and $\tilde{\tilde{w}}$, respectively.

Remark 15 ***(Rate for*** $n = 1$***).*** *For* $n = 1$*, we go back to the setup of Remark 6. Suppose* z *is unknown. We estimate* z *and* Λ *respectively by* $\tilde{\tilde{z}}$ *and* $\tilde{\tilde{\Lambda}} = ((\tilde{\lambda}_{ab,\tilde{\tilde{z}},(1/n,n),m}))_{K\times K}$*. Further, for some positive sequence* $\{\tilde{C}_m\}$*, suppose we have that* $\tilde{\mathcal{S}}_m \geq \tilde{C}_m\ \forall m$ *with probability* 1*. Then*

$$\frac{1}{m^2}||Ed_{\tilde{\tilde{z}}}(\tilde{\tilde{\Lambda}}) - Ed_z(\Lambda)||_F^2 \quad = \quad O_P\left(\left(\frac{Km}{\nu_{m,n}^2}\right)^2 + \frac{\log m}{\tilde{C}_m^2}\rho_{m,n}\right).$$

The above rate of convergence is slower than the rate obtained in Remark 6 where communities are known. This rate of convergence varies with different clustering methods employed for estimating z*. Zhang et al. (2015) used a clustering algorithm for dense SBM (that is* $\inf_{m,n}\rho_{m,n} > C > 0$*) so that the square of misclassification rate is* $\sqrt{\frac{\log m}{m}}$ *and* $\tilde{C}_m^2 = \sqrt{m\log m}$*. A detailed discussion on the impact of the clustering algorithm is provided in Section 7.5.*

2.2. On the condition (A1)

As seen from the results in Section 2.1, condition (A1) plays a critical role. Next, we discuss examples where it holds—Examples 1 and 3—and where it fails to do so—Example 2.

Example 1 *Suppose we have* K *balanced communities of size* m/K*. Let* $\Lambda = (p_1 - q_1)I_K + q_1 J_K$ *and* $\Delta = (p_2 - q_2)I_K + q_2 J_K$*, where* p_1, p_2, q_1, q_2 *may depend on both* m *and* n*,* I_K *is the identity matrix of order* K *and* J_K *is the* $K \times K$ *matrix whose entries are all equal to* 1*. Also assume* $|p_1 - q_1|, |p_2 - q_2| > \epsilon$ *and* $0 < C < p_1, q_1, p_2, q_2 < 1 - C < 1$ *for some* $C, \epsilon > 0$*, which implies dense regime. Then,* $\inf_{m,n}\rho_{m,n} > C$ *for some* $C > 0$ *and the smallest non-zero singular value of* $Ed_z(\Lambda)$ *and* $Ed_w(\Delta)$ *are* $\frac{m}{K}|p_1 - q_1|$ *and* $\frac{m}{K}|p_2 - q_2|$*, respectively. Therefore,* $\nu_{m,n} = O(\frac{m}{K})$ *and (A1) reduces to*

$$\frac{K^3}{m} \to 0 \quad \textit{and} \quad \frac{K^2\sqrt{n}}{m} = O(1). \tag{2.9}$$

If K *is finite, then we need* $n = O(m^2)$*, which is a rather stringent requirement for most real applications.*

If $K = \sqrt{m}$*, the condition does not hold as* $m, n \to \infty$*. If* $K = Cm^{0.5-\delta}$ *for some* $C, \delta > 0$*, then (2.9) holds if* $m^{0.5-3\delta} \to 0$ *and* $n = O(m^{4\delta})$*. In summary, if* $K = Cm^{0.5-\delta}$*,* $n = O(m^{4\delta})$ *for some* $C > 0$ *and* $\delta > 1/6$*, then (A1) holds.*

Next, define

$$\begin{aligned} m_{\max,z}, m_{\max,w} &= \textit{largest community size of } z \textit{ and } w \textit{ respectively} \\ m_{\min,z}, m_{\min,w} &= \textit{smallest community size of } z \textit{ and } w \textit{ respectively.} \end{aligned}$$

The above conclusion also holds if we have $\lim \frac{m_{\max,z}}{m_{\min,z}} = \lim \frac{m_{\max,w}}{m_{\min,w}} = 1$ *instead of having balanced communities.*

In the sparse regime, the assumption $|p_1 - q_1|, |p_2 - q_2| > \epsilon$ *and* $0 < C < p_1, q_1, p_2, q_2 < 1 - C < 1$ *for some* $C, \epsilon > 0$ *do not hold. In this case, (2.9) needs to hold after multiplying by* $\sup_{m,n}\{(\min\{|p_1 - q_1|, |p_2 - q_2|\})^{-2} \max\{p_1, q_1, p_2, q_2\}\}$.

Example 2 *Consider the same model as in Example 1 with* $|p_1 - q_1| = |p_2 - q_2| = n^{-\delta}$ *for some* $\delta > 0$. *Suppose* K *is finite. Then, (A1) reduces to*

$$\frac{n^{2\delta}}{m} \to 0 \quad \textit{and} \quad \frac{n^{1/2+2\delta}}{m} = O(1). \tag{2.10}$$

In this case, (A1) does not hold if $m = C\sqrt{n}$ *for some* $C > 0$.

We can also take $|p_1 - q_1| = |p_2 - q_2| = m^{-\delta}$ *(instead of* $n^{-\delta}$*) for some* $\delta > 0$. *Then also (A1) does not hold for* $m = n$ *and* $\delta > 1/4$.

Example 3 *Let* $\Lambda = p_1 I_K$ *and* $\Delta = p_2 I_K$, *where* p_1 *and* p_2 *may depend on both* m *and* n. *Assume that there is* $\epsilon > 0$ *such that* $p_1, p_2 > \epsilon$, *that is dense regime. Then, the smallest non-zero singular values of* $Ed_z(\Lambda)$ *and* $Ed_w(\Delta)$ *are* $m_{\min,z}p_1$ *and* $m_{\min,w}p_2$ *respectively. Let* $m_{\min} = \min\{m_{\min,z}, m_{\min,w}\}$. *Therefore,* $\nu_{m,n} = O(m_{\min})$. *Let* $\tilde{\rho}_m = \frac{m_{\min}}{m}$. *Thus, (A1) reduces to*

$$\frac{K}{m\tilde{\rho}_m^2} \to 0 \quad \textit{and} \quad \frac{\sqrt{n}}{m\tilde{\rho}_m^2} = O(1). \tag{2.11}$$

Suppose $K = Cm^\lambda$ *and* $m_{\min} = Cm^\delta$ *for some* $\lambda, \delta \in [0, 1]$. *Then,* $\tilde{\rho}_m = m^{\delta-1}$ *and (2.11) reduces to*

$$\frac{1}{m^{2\delta-\lambda-1}} \to 0 \quad \textit{and} \quad \frac{\sqrt{n}}{m^{2\delta-1}} = O(1).$$

Thus, (A1) holds if $K = Cm^\lambda$, $m_{\min} = Cm^\delta$, $n = m^{4\delta-2}$ *for some* $\lambda, \delta \in [0, 1]$ *and* $2\delta - \lambda - 1 > 0$.

In the sparse regime, the assumption $p_1, p_2 > \epsilon > 0$ *does not hold. In this case, (2.11) needs to hold after multiplying by* $\sup_{m,n}\{(\min\{p_1, p_2\})^{-2} \max\{p_1, p_2\}\}$.

Remark 16 *Examples 2.1 and 2.3 provide sufficient conditions for Assumption (A1) in specific cases. These conditions require a shorter time horizon (that means smaller value of* n*) in comparison with network sizes* m *that are of interest in many real applications. Hence, these sufficient conditions are easily applicable to many real-world networks. Nevertheless,* $\hat{\tau}_{m,n}$ *(see Section 3) proves useful in many practical settings for estimating the change-point and avoids the above trade-off between* m *and* n.

Remark 17 ***A variant of the every point clustering algorithm with a weaker assumption on the misclassification rate:*** *Note that computation of* $\tilde{\tilde{\tau}}_{m,n}$ *involves estimation of communities at every time point. The necessity of clustering at every time-point leads us to consider condition (A1). One may note though that since for theoretical* *considerations the change-point needs to be contained in the interval* $(c^*, 1-c^*)$*, the following alternative approach can be employed: use Clustering Algorithm I for the first* $[nc^*]$ *and the last* $[nc^*]$ *time points for estimating* z *and* w*, respectively. Denote the corresponding estimators by* z^* *and* w^**. Then, the corresponding change-point estimator* $\tau^*_{m,n}$ *can be obtained by*

$$\tau^*_{m,n} \quad = \quad \arg\min_{b\in(c^*,1-c^*)} \tilde{L}(b, z^*, w^*, \tilde{\Lambda}_{z^*,(b,n),m}, \tilde{\Delta}_{w^*,(b,n),m}).$$

Since we are using order n*-many time points in the clustering step and also for estimating the true change-point* $\tau_{m,n} \in (c^*, 1-c^*)$*, the misclassification rates for* z^* *and* w^* *are similar to those of* $\tilde{\tilde{z}}$ *and* $\tilde{\tilde{w}}$ *obtained in Theorem 12. As pointed out in the discussion preceding the statement of assumption (A1), clustering at every time point requires the misclassification rate* $\mathcal{M}_{b,n,m}$ *to decay faster than* $n^{-1}\rho^{-1}_{m,n}||Ed_z(\Lambda) - Ed_w(\Delta)||_F$*. However, when computing* $\tau^*_{m,n}$*, we use the same estimates* z^* *and* w^* *for all time points. As will be seen later in Remark 35, if we use the same clustering solution (assignment of nodes to communities) through all the time points, we only require the misclassification rate to decay faster than* $\rho^{-1}_{m,n}||Ed_z(\Lambda) - Ed_w(\Delta)||_F$ *(instead of* $n^{-1}\rho^{-1}_{m,n}||Ed_z(\Lambda) - Ed_w(\Delta)||_F$*) for consistency of the change-point estimator. As a consequence, a weaker assumption on the misclassification rate*

(A1*) $\frac{m}{\sqrt{n}\nu^2_{m,n}}\rho_{m,n} = O(1)$

together with the SNR-DSBM condition are needed to establish the consistency of $\tau^*_{m,n}$*. The upshot is that if node assignments* z^* *and* w^* *are employed, assumption (A1) becomes weaker.*

To further illustrate the latter point, note that in Example 1, (A1) reduces to* $K^2 = O(m\sqrt{n})$*. Therefore, if* $K = Cm^\delta$ *and* $n = Cm^\lambda$ *for some* $\delta \in (0,1)$ *and* $\lambda > 0$*, then (A1*) reduces to* $1 - 2\delta + \lambda/2 \geq 0$*. For Example 2, (A1*) boils down to* $n^{2\delta} = O(m)$*. Finally, in Example 3, (A1*) holds if* $m_{\min} = Cm^\delta$*,* $n = m^\lambda$ *for some* $\lambda > 0$*,* $\delta \in [0,1]$ *and* $2\delta + \lambda/2 - 1 \geq 0$*.*

Though the method in Remark 17 needs a weaker assumption on the misclassification rate compared to (A1), this paper focuses on the "every time-point clustering algorithm" due to the following two considerations.

Remark 18 ***Considerations for*** $\tau^*_{m,n}$***:*** *Note that in practice the strategy in Remark 17 requires that* c^* *be known, which may not be the case in most applications. If* c^* *is not known, a reasonable practical alternative is to use only the first and last time points to estimate* z *and* w*, respectively. Further,* $\frac{m\rho_{m,n}}{\nu^2_{m,n}} = O(1)$ *is required for consistency of the change-point estimator. This is stronger than (A1*) but weaker than (A1). One can argue that, in principle, the value of* c^* *is needed to compute the change-point, since for establishing the theoretical results the search to identify it is restricted in the interval* $(c^*, 1-c^*)$*. However, in practice, one always searches throughout the entire interval, and hence the practical alternative of using the first and last time points to estimate* z *and* w *is compatible with it.*

Finally, note that this alternative, that is known stretches of points that belong to only a single regime, is viable for estimating a single change-point, but no longer so when multiple change-points are involved. In the latter case, one would still assume that the first and last change-points are separated away from the boundary by some fixed amount, but no such restrictions on the locations of the intermediate change-points can be imposed, and hence a full search strategy (see for example the algorithm proposed in Auger and Lawrence (1989)) combined with clustering is unavoidable. This is the reason that our analysis focuses on the "every time-point clustering algorithm" since it provides insights on where challenges will arise for the case of multiple change-points, appropriate treatment of which is nevertheless beyond the scope of this paper.

Remark 19 *In this section, we used a specific clustering procedure proposed in Bhattacharyya and Chatterjee (2017) to identify the communities and to locate the change-point. Nevertheless, other clustering algorithms proposed in the literature [Pensky and Zhang (2019); Rohe et al. (2011)] could be employed. For any given clustering algorithms the following statements hold.*

(a) The conclusions of Theorems 11, 12 and 13 hold once we replace (NS) and (A1) by

(A9) $n^2\mathcal{M}^2_{b,n,m}\rho^2_{m,n}||Ed_z(\Lambda) - Ed_w(\Delta)||_F^{-2} \to 0, \ \forall \ b \in (c^*, 1-c^*)$,

where $\mathcal{M}_{b,n,m}$ *is the maximum misclassification error that* $\tilde{z}_{b,n,m}$ *and* $\tilde{w}_{b,n,m}$ *in estimating* z *and* w*, respectively, given in (2.8).*

(b) Suppose (A9) and SNR-DSBM hold and in addition $\mathcal{M}^2_{\tilde{\tilde{\tau}}_{m,n},n,m} = O_P(E_{m,n})$ *for some sequence* $E_{m,n} \to 0$*. Moreover, assume that for some positive sequence* $\{\tilde{C}_{m,n}\}$, $\hat{\mathcal{S}}_{m,n} \geq \tilde{C}_{m,n} \ \forall \ n$ *with probability* 1 *and* $n\tilde{C}^{-2}_{m,n}\log m||Ed_z(\Lambda) - Ed_w(\Delta)||_F^4 \geq I(n > 1)$*. Then*

$$\frac{1}{m^2}||Ed_{\tilde{\tilde{z}}}(\tilde{\tilde{\Lambda}}) - Ed_z(\Lambda)||_F^2, \quad \frac{1}{m^2}||Ed_{\tilde{\tilde{w}}}(\tilde{\tilde{\Delta}}) - Ed_w(\Delta)||_F^2$$
$$= O_P\left(E_{m,n} + \frac{I(n>1)\rho^2_{m,n}}{n^2|Ed_{\tilde{\tilde{z}}}(\tilde{\tilde{\Lambda}}) - Ed_z(\Lambda)||_F^4} + \frac{\log m}{n\tilde{C}^2_{m,n}}\rho_{m,n}\right) \quad (2.12)$$

The proofs of statements (a) and (b) follow immediately from those of Theorems 11–13 and Remark 35.

The analogue of (A9) for existing clustering algorithms in the literature and their comparison with (A1) are discussed in Section 7.5 in more detail.

3. A fast 2-step procedure for change-point estimation in the DSBM

The starting point of our exposition is the fact that the SBM is a special form of the Erdős-Rényi random graph model. The latter is characterized by the following edge generating mechanism. Let p_{ij} be the probability of having an edge between nodes i and j and let P be the $m \times m$ corresponding connectivity probability matrix. We denote this model by ER(P). An adjacency matrix A is said to be generated according to ER(P), if $A_{ij} \sim$

Bernoulli(p_{ij}) *independently* and we denote it by $A \sim \text{ER}(P)$. Clearly $A \sim \text{SBM}(z, \Lambda)$ implies $A \sim \text{ER}(\text{Ed}_z(\Lambda))$.

The DSBM with single change-point in (2.1) can be represented as a random graph model as follows: there is a sequence $\tau_{m,n} \in (0,1)$ such that for all $n \geq 1$,

$$A_{t,n} \sim \begin{cases} \text{ER}(\text{Ed}_z(\Lambda)), & \text{if } 1 \leq t \leq \lfloor n\tau_{m,n} \rfloor \\ \text{ER}(\text{Ed}_w(\Delta)), & \text{if } \lfloor n\tau_{m,n} \rfloor < t < n \end{cases} \tag{3.1}$$

and $\Lambda \neq \Delta$ and/or $z \neq w$.

In general, without any structural assumptions, a dynamic Erdős-Rényi model with a single change-point has $m(m+1)+1$ many unknown parameters, the $0.5m(m+1)$ pre- and post- change-point edge probabilities and 1 change-point. An estimate of $\tau_{m,n}$ can be obtained by optimizing the following least-squares criterion function.

$$\begin{aligned} \hat{\tau}_{m,n} &= \arg\min_{b \in (c^*, 1-c^*)} L(b) \quad \text{where} \\ L(b) &= \frac{1}{n} \sum_{i,j=1}^{m} \left[\sum_{t=1}^{nb} (A_{ij,(t,n)} - \hat{p}_{ij,(b,n),m})^2 + \sum_{t=nb+1}^{n} (A_{ij,(t,n)} - \hat{q}_{ij,(b,n),m})^2 \right], \\ \hat{p}_{ij,(b,n),m} &= \frac{1}{nb} \sum_{t=1}^{nb} A_{ij,(t,n)} \quad \text{and} \quad \hat{q}_{ij,(b,n),m} = \frac{1}{n(1-b)} \sum_{t=nb+1}^{n} A_{ij,(t,n)}. \end{aligned} \tag{3.2}$$

Next, we present our 2-step algorithm.

2-Step Algorithm:

Step 1: In this step, we ignore the community structures and assume $z(i) = w(i) = i$ for all $1 \leq i \leq m$. We compute the least-squares criterion function $L(\cdot)$ given in (3.2) and obtain the estimate $\hat{\tau}_{m,n} = \arg\min_{b \in (c^*, 1-c^*)} L(b)$.

Step 2: This step involves estimation of other parameters in DSBM. We estimate z and w by $\hat{z} = \tilde{z}_{\hat{\tau}_{m,n},n,m}$ and $\hat{w} = \tilde{w}_{\hat{\tau}_{m,n},n,m}$, respectively, and subsequently Λ and Δ by $\hat{\hat{\Lambda}} = \tilde{\Lambda}_{\hat{z},(\hat{\tau}_{m,n},n),m}$ and $\hat{\hat{\Delta}} = \tilde{\Delta}_{\hat{w},(\hat{\tau}_{m,n},n),m}$, respectively.

Computational complexity of the 2-Step Algorithm

It can easily be seen that Step 1 requires $O(m^2 n)$ operations, while Step 2 due to performing clustering requires $O(m^3)$ operations. Thus, the total computational complexity of the entire algorithm is $O(m^3 + m^2 n) \sim O(m^2 \max(m,n))$, which is significantly smaller than that of obtaining $\tilde{\tilde{\tau}}_{m,n}$ in (2.7).

3.1. Theoretical Results for $\hat{\tau}_{m,n}$

The following identifiability condition and the restriction on the sparse parameter are required.

SNR-ER: $\frac{n}{m^2\rho_{m,n}}||\mathrm{E}d_z(\Lambda) - \mathrm{E}d_w(\Delta)||_F^2 \to \infty$.

Sparse-ER: $\rho_{m,n} > Cm^{-2}$ for some $C > 0$.

SNR-ER requires that the signal per edge parameter and scaled by the sparsity parameter grows faster than $1/\sqrt{n}$. Clearly, SNR-ER is stronger than SNR-DSBM, as expected since the ER model involves m^2 parameters, as opposed to K^2 parameters for the DSBM.

Similar to the discussed in Remark 3, in the presence of SNR-ER, Sparse-ER ensures consistency of the change-point estimator $\hat{\tau}_{m,n}$ after scaling by the sparsity parameter. Also, Sparse-ER is weaker than Sparse-DSBM.

The following Theorem provides asymptotic properties for the estimates of the DSBM parameters obtained from the 2-Step Algorithm. Its proof is given in Section 7.6.

Theorem 20 *Suppose that SNR-ER and Sparse-ER hold. Then, the conclusion of Theorem 11 holds for $\hat{\tau}_{m,n}$. Similarly, under SNR-ER, Sparse-ER and (NS), the conclusions of Theorems 12 and 13 continue to hold for $\hat{z}$, $\hat{w}$, $\hat{\hat{\Lambda}}$ and $\hat{\hat{\Delta}}$.*

Remark 21 *As we can observe above, there is a trade-off between the signal-to-noise condition and the decay rate of the sparsity parameter. Though the convergence rate of $\tilde{\tau}_{m,n}$ and $\hat{\tau}_{m,n}$ are the same, the former one needs stronger Sparse-DSBM under weaker SNR-DSBM compared to weaker Sparse-ER under stronger SNR-ER required for the later estimator.*

Remark 22 *It is easy to see that the average signal per edge $(m^2\rho_{m,n})^{-1}||Ed_z(\Lambda) - Ed_w(\Delta)||_F^2 \leq \rho_{m,n} \to 0$ for sparse graphs. On the other side, $\rho_{m,n}$ is bounded away from 0 for dense graphs and $(m^2\rho_{m,n})^{-1}||Ed_z(\Lambda) - Ed_w(\Delta)||_F^2 \geq C$ for some $C > 0$ can be often satisfied. Therefore dense regime generates more signal compared to the sparse regime and consequently the later one needs larger sample size (number of time points) to satisfy the signal-to-noise conditions SNR-DSBM and SNR-ER for detecting the change-point consistently and has slower convergence rate of the change-point estimator.*

Remark 23 *One may wonder regarding dense settings (similar discussion is true for the sparse graphs as well) where SNR-DSBM holds, but neither (A1) nor SNR-ER do. Examples 5 and 6 introduce such settings in the context of changes in the connection probabilities and in the community structures, respectively. The methods discussed in Sections 2 and 3 fail to detect the change-point under the above-presented settings. Therefore, alternative strategies need to be investigated. One possibility for the case of a single change-point being present was discussed in Remark 17 and more details are given in Example 7. Another setting that does not require clustering is presented in Example 8 and builds on the model discussed in Gao et al. (2015). However, the setting in Example 8 is very specific involving two parameters only. Nevertheless, a generally applicable strategy is currently lacking for the regime where SNR-DSBM holds, but neither SNR-ER or (A1) does. This constitutes an interesting direction for future research.*

4. Comparison of the "Every time point clustering algorithm" vs the 2-step algorithm

Our analysis up to this point has highlighted the following key findings. If the total signal is strong enough (that means if SNR-ER holds), then it is beneficial to use the 2-step algorithm that provides consistent estimates of *all* DSBM parameters at *reduced computational cost*. On the other hand, if the signal is not adequately strong (that means if SNR-ER fails to hold, but SNR-DSBM holds) then the only option available is to use the computationally expensive "every time point algorithm", provided that (A1) also holds. Our discussion in Section 2.2 indicates that (A1) (and (A1*)) is not an innocuous condition and may fail to hold in real application settings.

For example, consider a DSBM with $m = 60$ nodes, $K = 2$ communities and $n = 60$ time points. Suppose that there is a break at $n\tau_n = 30$, due to a change in community connection probabilities. Further, assume that the community connection probabilities before and after the change-point are given by $\Lambda = \begin{pmatrix} 0.6 & 0.3 \\ 0.3 & 0.6 \end{pmatrix}$ and $\Delta = \Lambda + \frac{1}{n^{1/4}} J_2$ (or $= \Lambda + \frac{1}{m^{1/4}} J_2$), respectively. Finally, suppose that there is no change in community structures and $z(i) = w(i) = I(1 \leq i \leq m/2) + 2I(m/2 + 1 \leq i \leq m)$. In this case, one can check that $\inf_{m,n} \rho_{m,n} > C > 0$, $\frac{n}{m^2} ||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_2^2 = 7.75$, $\frac{Km}{\nu_{m,n}^2} = 1.48$, $\frac{m\sqrt{n}}{\nu_{m,n}^2} = 5.7$ and hence SNR-ER holds but (A1) fails. Figure 1 plots the least-squares criterion function against time scaled by $1/n$, corresponding to the 2-step, known communities, and "every time point" algorithms, respectively. The plots show that the trajectory of the least-squares criterion function is much smoother and the change-point is easily detectable when known community structures are assumed. It is also the case for the 2-step algorithm, albeit with more variability. However, since (A1) fails to hold, the objective function depicted in Figure 1 (bottom middle panel) clearly illustrates that the change-point is not detectable for the "every time point" algorithm.

The next question to address is *"How stringent is SNR-ER"* under the DSBM model. As the following discussion shows, the reallocation of nodes to new communities generates strong enough signal, and therefore SNR-ER may be easier to satisfy in practice than one might suppose.

Sufficient conditions for SNR-ER under the DSBM model.

We examine a number of settings where SNR-ER holds under the DSBM network generating mechanisms and hence the 2-step algorithm can be employed. Specifically, the following proposition provides sufficient conditions for SNR-ER to hold. Let $\mathbb{N}$ be the set of all natural numbers. Define the classes of functions $\mathcal{F} := \{f \big| f : \mathbb{N} \times \mathbb{N} \to (0,1)\}$ and $\mathcal{G}_m := \{g = ((m^2 f) \vee 1) \wedge 0.5m(m-1) \big| f \in \mathcal{F}\}$ for all $m \geq 1$. For any $f \in \mathcal{F}$, let

$$\mathcal{A}(f) = \{(i,j) : \; |\lambda_{z(i)z(j)} - \delta_{w(i)w(j)}| > f(m,n)\}.$$

Hence, $\mathcal{A}(f)$ corresponds to the set of all edges for which the connection probability changes at least by an $f(m,n)$ amount.

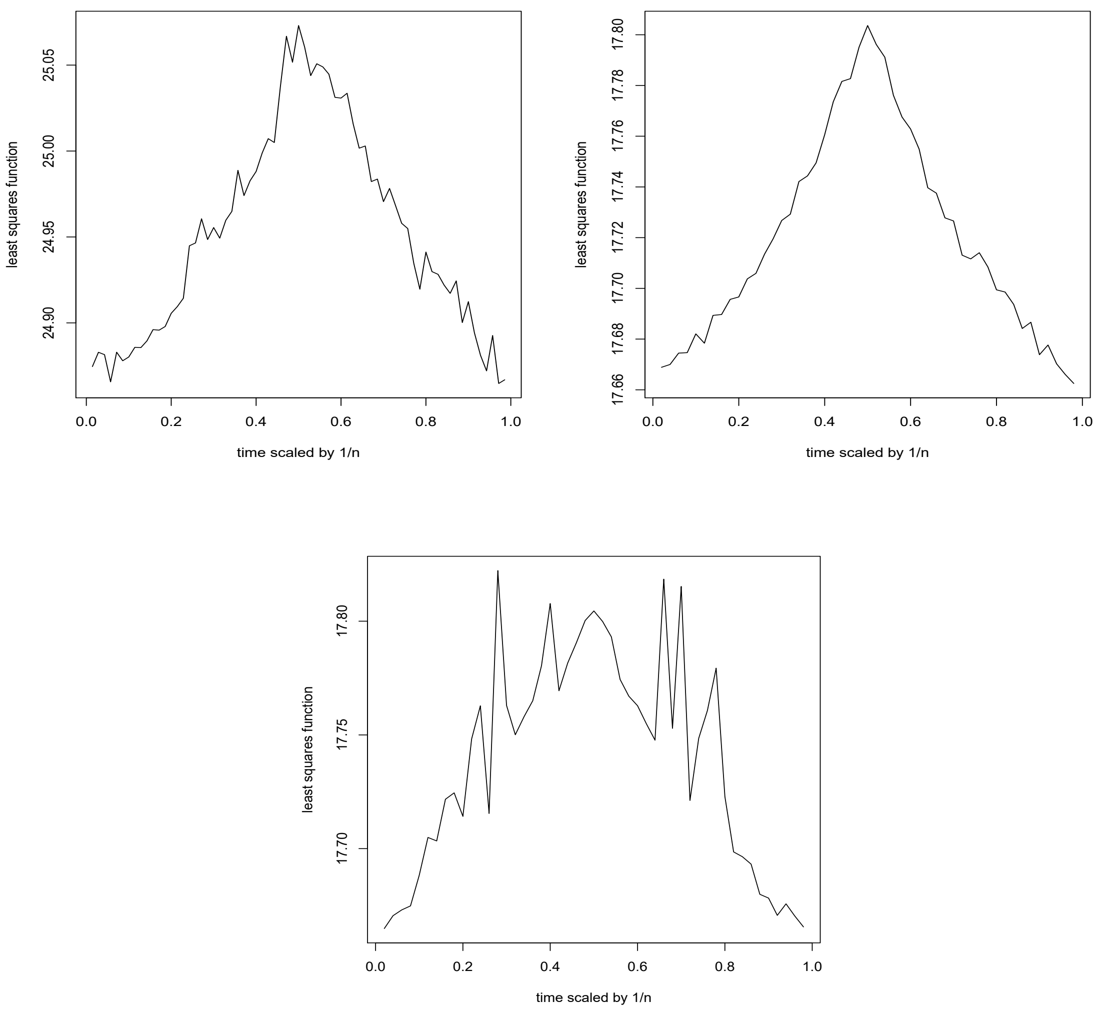

Figure 1: A plot of the least-squares criterion function against time scaled by $1/n$. Top left and right panels correspond to the 2-step and known communities algorithm, while the bottom middle depicts the "every time point" algorithm, respectively.

Proposition 24 *Suppose $|A(f)| \geq Cg(m,n)$ for some $f \in \mathcal{F}, g \in \mathcal{G}_m$ such that $nm^{-2}g(m,n)(f(m,n))^2 \to \infty$. Then, SNR-ER holds.*

The above proposition follows from the fact that

$$\frac{n}{m^2 \rho_{m,n}} ||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2 \geq nm^{-2}g(m,n)(f(m,n))^2 \to \infty.$$

This implies that at least $g(m,n)$-many edges need to change their connection probability by at least $f(m,n)$ amount for SNR-ER to be satisfied. This leads us to the following scenarios (A)-(D) that often arise in practice.

The following example provides a choice for f and g, respectively.

Example 4 *Let $A(\epsilon, \delta_1) = \{(i,j) : \ |\lambda_{z(i)z(j)} - \delta_{w(i)w(j)}| > \epsilon m^{-\lambda_1/2} n^{-\delta_1/2}\}$, $|A(\epsilon,\delta_1)| \geq \max\{Cm^{2-\lambda_2}n^{-\delta_2}, 1\}$ and $m^{\lambda_1+\lambda_2} = o(n^{(1-\delta_1-\delta_2)})$ for some $C, \epsilon > 0$ and $0 \leq \delta_1 + \delta_2 < 1, \lambda_1 \geq 0, 0 \leq \lambda_2 \leq 2$. This implies that at least $m^{2-\lambda_2}n^{-\delta_2}$-many edges need to change their connection probability by at least $\epsilon m^{-\lambda_1/2}n^{-\delta_1/2}$ amount. Then Proposition 24, by setting $f = \epsilon m^{-\lambda_1/2}n^{-\delta_1/2}$ and $g = \max\{Cm^{2-\lambda_2}n^{-\delta_2}, 1\}$, establishes that SNR-ER holds. As $\lambda_1, \lambda_2, \delta_1, \delta_2$ increase, the signal $||Ed_z(\Lambda) - Ed_w(\Delta)||_F^2$ due to the change decreases and therefore a large number of time points n is required to accumulate adequate signal (that is to satisfy the SNR condition) for detecting the change-point.*

Next, we discuss settings motivated by real-world applications, wherein the SNR-ER condition holds for DSBM.

(A) **Reallocation of nodes**: Suppose that the pre- and post-community connection probabilities are the same; that is $\Lambda = \Delta$. This also implies that the total number of communities before and after the change-point are equal. Suppose that some of the nodes are reallocated to new communities after the change-point epoch.
A motivating example for this scenario comes from voting patterns of legislative bodies as analyzed in Bao and Michailidis (2018). In this setting, one is interested in identifying when voting patterns of legislators change significantly. By considering the legislators as the nodes of the network, an edge between two of them indicates voting similarly on a legislative measure (for examples bill, resolution), while the communities reflect their political affiliations, it can be seen that after an election the composition of the communities may be altered—reassignment of nodes.

In this situation, SNR-ER holds if the entries of Λ (or Δ) are adequately separated and enough nodes are reallocated. Specifically, for some $\epsilon, C > 0$ and $0 \leq \delta_1 + \delta_2 < 1, \lambda_1 \geq 0, 0 \leq \lambda_2 \leq 2$, suppose we have $|\Lambda_{ij} - \Lambda_{i'j'}| > \epsilon m^{-\lambda_1/2}n^{-\delta_1/2}$ $\forall (i,j) \neq (i',j')$ and $\max\{Cm^{2-\lambda_2}n^{-\delta_2}, 1\}$-many nodes change their community after time $n\tau_n$. Then, by Proposition 24 and Example 4, SNR-ER holds.

(B) **Change in connectivity**: Suppose that the community structures remain the same before and after the change-point (that implies $z = w$), but their community connection probabilities change (therefore $\Lambda \neq \Delta$). This scenario is motivated by the following examples: in transportation networks, when service is reduced or even halted between two service locations, in social media platforms (for example Facebook) when a new online game launches, or in collaboration networks when a large scale project is completed.

Then, SNR-ER holds if entries of Λ are adequately separated from those of Δ. Specifically, for some $\epsilon > 0$ and $0 \le \delta < 1, \lambda \ge 0$, suppose we have $|\lambda_{ij} - \delta_{ij}| > \epsilon m^{-\lambda/2} n^{-\delta/2}\ \forall i, j = 1, 2, \ldots, K$ and $m^{2+\lambda} = o(n^{1-\delta})$. Then, by Proposition 24 and Examples 4, SNR-ER holds.

(C) **Merging Communities**: Sometimes, when two user communities cover the same subject matter and share similar contributors, they may wish to merge their communities to push their efforts forward in a desired direction. Suppose that the 1st and the Kth communities in z merge into the 1st community in w. In this situation, SNR-ER holds if the pre-connection probability between the 1st and the K-th communities and the post-connection probability within the 1st community are adequately separated and if the sizes of the 1st and the K-th communities are large before the change. Precisely, suppose $|\lambda_{1K} - \delta_{11}| > Cm^{-\lambda_1/2} n^{-\delta_1/2}$, $s_{1,z} s_{K,z} \ge Cm^{2-\lambda_2} n^{-\delta_2}$ and $m^{\lambda_1+\lambda_2} = o(n^{1-\delta_1-\delta_2})$ for some $C > 0$ and $0 \le \delta_1 + \delta_2 < 1, \lambda_1 \ge 0, 0 \le \lambda_2 \le 2$. Then, by Proposition 24 and Example 4, SNR-ER holds.

(D) **Splitting communities**: One community often splits into two communities when conflicts and disagreements arise among its members. Suppose that the 1st community in z splits into the 1st and Kth communities in w. In this case, SNR-ER holds if the pre-connection probability within the 1st community and the post-connection probability between the 1st and the Kth communities are adequately separated and the size of the 1st and the Kth communities are large after the change. Suppose $|\lambda_{11} - \delta_{1K}| > Cm^{-\lambda_1/2} n^{-\delta_1/2}$, $s_{1,w} s_{K,w} \ge Cm^{2-\lambda_2} n^{-\delta_2}$ and $m^{\lambda_1+\lambda_2} = o(n^{1-\delta_1-\delta_2})$ for some $C > 0$ and $0 \le \delta_1 + \delta_2 < 1, \lambda_1 \ge 0, 0 \le \lambda_2 \le 2$. Then, by Proposition 24 and Example 4, SNR-ER holds.

Remark 25 *Examples (A)-(D) above and Proposition 24 hold for both dense and sparse networks. However, as discussed in Remark 22, for sparse networks, a large enough number of time points n is required compared to the total number of nodes m. This is because that in a sparse network, there are relatively few edges to contribute to the total signal in Proposition 24.*

Next, we discuss two examples for a dense network regime wherein the SNR-ER condition fails to hold, but SNR-DSBM does.

(E) If most edges change their connection probabilities by an amount of $C_1/\sqrt{n}$ for some $C_1 > 0$, then SNR-ER does not hold, but SNR-DSBM does. Specifically, let $A(C_1) = \{(i, j) : |\lambda_{z(i)z(j)} - \delta_{w(i)w(j)}| = C_1/\sqrt{n}\}$. Suppose $|A(C_1)| = C_2 m^2$ for some $C_1, C_2 > 0$, $|\lambda_{z(i)z(j)} - \delta_{w(i)w(j)}| = 0\ \forall (i, j) \in A^c$ and $\min(\min\limits_u s_{u,z}, \min\limits_u s_{u,w}) \to \infty$. Then $\frac{n}{m^2} ||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2 = C_1^2 C_2 \not\to \infty$ but $\frac{n}{K^2} ||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2 = C_1^2 C_2 \frac{m^2}{K^2} \to \infty$.

(F) If the connection probabilities between the smallest community and the remaining ones change by $C/\sqrt{n}$ for some $C > 0$, then for an appropriate choice of K and smallest community size, SNR-ER does not hold, but SNR-DSBM does. Specifically, suppose $z = w$, $K = C_1 m^{\delta_1/2}$, $\min\limits_u s_{u,z} = s_{1,z} = C_2 m^{\delta_2/2}$, $|\lambda_{1j} - \delta_{1j}| = C_3/\sqrt{n}\ \forall j$ and $|\lambda_{ij} - \delta_{ij}| = 0\ \forall i \neq$

$1, j \neq 1$ for some $C_1, C_2, C_3 > 0$ and $0 < \delta_1 + \delta_2 \leq 2$, $\delta_1 < \delta_2$. Then $\frac{n}{m^2}||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2 = C_3^2 C_2 m^{-(2-\delta_2)} \to 0$ but $\frac{n}{K^2}||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2 = C_3^2 C_2 C_1^{-2} m^{\delta_2 - \delta_1} \to \infty$.

Note that examples (E)-(F) only deal with the SNR-DBSM condition and do not address the equally important (A1) condition for the "every time point clustering algorithm" to work. The next example provides a dense setting where SNR-ER does not hold, but both SNR-DSBM and (A1) hold.

(G) Consider the model and assumptions in Example 3. Suppose $p_2 = p_1 + \frac{1}{\sqrt{n}}$. Then, $mn^{-1} \leq ||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2 \leq m^2 n^{-1}$. Hence, SNR-ER does not hold. Further, if $K^2 = o(m)$, then SNR-DSBM holds. Thus, SNR-DSBM and (A1) hold if $K = Cm^{\lambda}$, $m_{\min} = Cm^{\delta}$ and $n = m^{4\delta-2}$ for some $\lambda \in [0, 0.5), \delta \in [0, 1]$ and $2\delta - \lambda - 1 > 0$.

The upshot of the above examples is that due to the structure of the DSBM, there are many instances arising in real settings where SNR-ER holds. On the other hand, as an example (G) illustrates, some rather special settings are required for SNR-ER to fail, while both SNR-DSBM and (A1) hold. Thus, it is relatively safe to assume that the 2-step algorithm is applicable across a wide range of network settings, making it a very attractive option to practitioners.

4.1. Numerical Illustration

Next, we discuss the performance of the three change-point estimates $\tilde{\tau}_{m,n}$, $\hat{\tau}_{m,n}$ and $\tilde{\tilde{\tau}}_{m,n}$ based on synthetic data generated according to the following mechanism, focusing on the impact of the parameters m, n and small community connection probabilities on their performance.

Effect of m and n: We simulate from the following DSBMs (1), (2), (3) for three choices of $(m, n, n\tau_n) = (60, 60, 30), (500, 20, 10), (500, 100, 50)$ and two choices of $\lambda = 0, 1/20$. These results are presented in Tables 1-6. Although the following DSBMs satisfy the assumptions in Proposition 24, SNR-ER may be small for dealing with finite samples.[1]

(1) **Reallocation of nodes**: $K = 2$, $z(i) = I(1 \leq i \leq m/2) + 2I(m/2 + 1 \leq i \leq m)$, $w(2i-1) = 1$, $w(2i) = 2$ $\forall 1 \leq i \leq m/2$. $\Lambda = \Delta = \begin{pmatrix} 0.6 & 0.6 - \frac{1}{n^{\delta} m^{\lambda}} \\ 0.6 - \frac{1}{n^{\delta} m^{\lambda}} & 0.6 \end{pmatrix}$ for $\delta = 1/20, 1/10, 1/4$.

(2) **Change in connectivity**: $K = 2$, $z(i) = w(i) = I(1 \leq i \leq m/2) + 2I(m/2+1 \leq i \leq m)$, $\Lambda = \begin{pmatrix} 0.6 & 0.3 \\ 0.3 & 0.6 \end{pmatrix}$, $\Delta = \Lambda + \frac{1}{n^{1/4} m^{\lambda}} J_2$.

1. Note that by Proposition 24 and Example 4, for finite number of communities (K is finite), $\frac{n}{m^2}||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2 = O(m^{-\lambda_1-\lambda_2} n^{1-\delta_1-\delta_2})$ and $\frac{n}{K^2}||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2 = O(m^{2-\lambda_1-\lambda_2} n^{1-\delta_1-\delta_2})$ and for balanced community with $\min\{|\lambda_{ij} - \lambda_{i'j'}|, |\delta_{ij} - \delta_{i'j'}| : (i,j) \neq (i',j'), (i,j) \neq (j',i')\} \geq Cn^{-\delta} m^{-\lambda}$ ($C > 0$, $\delta \geq 0$, $\lambda \geq 0$), we have $\nu_{m,n} = O(\frac{m^{1-\lambda} n^{-\delta}}{K})$, we have $\frac{Km}{\nu_{m,n}^2} = O(n^{2\delta}/m^{1-2\lambda})$ and $\frac{m\sqrt{n}}{\nu_{m,n}^2} = O(\frac{n^{0.5+2\delta}}{m^{1-2\lambda}})$. Thus, for small n and large m, SNR-ER becomes small, but SNR-DSBM and (A1) hold. Moreover, (A1) is not satisfied for large δ and λ and small m.

(3) **Merging communities**: $K = 3$, $z(i) = I(1 \le i \le 20) + 2I(21 \le i \le 40) + 3I(41 \le i \le 60)$, $w(i) = I(1 \le i \le 20,\ 41 \le i \le 60) + 2I(21 \le i \le 40)$, $\Lambda = \begin{pmatrix} 0.6 & 0.3 & 0.6 - \frac{1}{n^{1/20}m^{\lambda}} \\ 0.3 & 0.6 & 0.3 \\ 0.6 - \frac{1}{n^{1/20}m^{\lambda}} & 0.3 & 0.6 \end{pmatrix}$, $\Delta = \begin{pmatrix} 0.6 & 0.3 & 0 \\ 0.3 & 0.6 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.
Splitting communities and merging communities are similar once we interchange z, w, and Λ, Δ.

Table 1: Illustrating the performance of the change-point estimators with $m = 60, n = 60, n\tau_{m,n} = 30$, $\lambda = 0$ based on 100 replicates and DSBMs (1), (2) and (3). Figures in brackets are frequencies of the number of change-points the corresponding change-point is observed. Further, $F_n := ||\mathrm{E}d_z(\Lambda) - \mathrm{E}d_w(\Delta)||_F^2$.

	Reallocation of nodes,in (1)			Change in connectivity, in (2)	Merging communities, in (3)
	$\delta = 1/20$	$\delta = 1/10$	$\delta = 1/4$		
F_n	1195.246	793.6742	232.379	2390.49	531.2205
$\frac{n}{m^2}F_n$	19.92077	13.2279	3.873	39.84	8.8537
$\frac{n}{K^2}F_n$	17928.69	11905.11	3485.685	35837.39	3541.47
$\frac{Km}{\nu_{m,n}^2}$	0.198	0.3	1.2	1.03	3.11
$\frac{m\sqrt{n}}{\nu_{m,n}^2}$	0.7777	1.1712	4	5.738	8.03
$\tilde{\tau}_{m,n}$	30(90), 29(4), 28(2), 31(4)	30(85), 29(6), 28(4), 31(5)	30(88), 29(7), 28(5)	30(85), 29(5), 28(6), 31(4)	30(88), 29(3), 28(4), 31(5)
$\hat{\tau}_{m,n}$	30(88), 28(5), 31(3), 34(4)	30(83), 29(3), 28(7), 31(7)	30(80), 29(9), 28(7), 31(4)	30(83), 28(10), 31(7)	30(88), 29(8), 28(4)
$\tilde{\tilde{\tau}}_{m,n}$	30(85), 28(5), 31(6), 32(4)	30(80), 28(8), 31(6), 32(4), 33(2)	30(34), 22(42), 25(10), 33(14)	30(40), 21(30), 28(18), 26(12)	30(21), 19(10), 23(48), 26(14), 38(7)

The following conclusions are in accordance with the results presented in Tables 1 through 6.

Table 2: Illustrating the performance of the change-point estimators with $m = 500, n = 20, n\tau_{m,n} = 10$, $\lambda = 0$ based on 100 replicates and DSBMs (1), (2) and (3). Figures in brackets are frequencies of the number of times the corresponding change-point is observed. Further, $F_n := ||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2$.

	Reallocation of nodes, in (1)			Change in connectivity, in (2)	Merging communities, in (3)
	$\delta = 1/20$	$\delta = 1/10$	$\delta = 1/4$		
F_n	92641.81	68660.03	27950.85	185283.6	41174.14
$\frac{n}{m^2}F_n$	7.411	5.49	2.24	14.82	3.294
$\frac{n}{K^2}F_n$	463209	343300.2	139754.2	926418.1	91498.08
$\frac{Km}{\nu_{m,n}^2}$	0.0216	0.0291	0.0716	0.072	0.3732
$\frac{m\sqrt{n}}{\nu_{m,n}^2}$	0.0483	0.0651	0.16	0.16	0.576
$\tilde{\tau}_{m,n}$	10(88), 9(6), 11(6)	10 (85), 9(5), 8(3), 11(6), 12(1)	10(90), 9(5), 8(1), 12(4)	10(89), 9(5), 11(4), 12(2)	10(88), 9(6), 8(4), 12(2)
$\hat{\tau}_{m,n}$	10(90), 8(6), 11(4)	10(88), 8(5), 11(7)	10(39), 3(23), 7(30), 13(8)	10(85), 9(7), 8(8)	10(83), 9(7), 8(4) 11(4), 12(2)
$\tilde{\tilde{\tau}}_{m,n}$	10(85), 9(7), 11(5), 12(3)	10(82), 8(6), 11(5), 12(7)	10(77), 8(11), 9(4), 11(8)	10(83), 9(9), 8(4), 11(4)	10(80), 8(9), 9(7) 11(4)

(a) SNR-ER holds for large n, small δ, λ and large signal $||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2$. We observe large SNR-ER and consequently good performance of $\hat{\tau}_{m,n}$, throughout Tables 1-6 except Column 3 in Table 2 and Column 3 and 5 of Table 5, which involve a small n and large δ, λ, leading to poor performance of $\hat{\tau}_{m,n}$.

(b) SNR-ER implies SNR-DSBM and thus a large SNR-DSBM is observed throughout Tables 1-6. Moreover, if $\nu_{m,n} = O(\frac{m^{1-\lambda}n^{-\delta}}{K})$ for some $\delta > 0$, then (A1) holds for small δ, λ, n, small K and large m. Thus, (A1) holds and $\tilde{\tilde{\tau}}_{m,n}$ exhibits good performance throughout Tables 1-6 except Columns $3-5$ in Table 1 and Columns $2-5$ in Table 4 where δ and λ are large and m small.

(c) Throughout Tables 1-6, SNR-DSBM holds and $\tilde{\tau}_{m,n}$ exhibits good performance, as expected. The estimator $\tilde{\tau}_{m,n}$ performs equally well to $\hat{\tau}_{m,n}$ and $\tilde{\tilde{\tau}}_{m,n}$, whenever SNR-ER and (A1) are satisfied. In all other settings, $\tilde{\tau}_{m,n}$ clearly outperforms them. For example, $\tilde{\tau}_{m,n}$ performs better than $\tilde{\tilde{\tau}}_{m,n}$ in Columns 3-5 of Table 1 and Columns 2-5 of Table 4 and better than $\hat{\tau}_{m,n}$ in Column 3 of Table 2 and Columns 3 and 5 of Table 5.

The above numerical results amply demonstrate the competitive nature of the computationally inexpensive 2-step algorithm under the settings posited. However, note that the

Table 3: Illustrating the performance of the change-point estimators with $m = 500, n = 100, n\tau_{m,n} = 50$, $\lambda = 0$, based on 100 replicates and DSBMs (1), (2) and (3). Figures in brackets are frequencies of the number of times the corresponding change-point is observed. Further, $F_n := ||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2$.

	Reallocation of nodes, in (1)			Change in connectivity, in (2)	Merging communities, in (3)
	$\delta = 1/20$	$\delta = 1/10$	$\delta = 1/4$		
F_n	78869.67	49763.4	12500	157739.3	35053.19
$\frac{n}{m^2}F_n$	31.548	19.905	5	63.1	14.0213
$\frac{n}{K^2}F_n$	1971742	1244085	312500	3943483	1389479.8
$\frac{Km}{\nu_{m,n}^2}$	0.02536	0.04019	0.16	0.1778	0.3732
$\frac{m\sqrt{n}}{\nu_{m,n}^2}$	0.1268	0.201	0.8	0.889	1.244
$\tilde{\tau}_{m,n}$	50(90), 49(4), 48(4), 51(2)	50(89), 49(8), 51(3)	50(93), 49(5), 48(1), 51(1)	50(91), 49(6), 48(3)	50(89), 49(7), 48(1), 51(3)
$\hat{\tau}_{m,n}$	50(92), 48(3), 51(5)	50(88), 49(7), 48(3), 47(1), 51(1)	50(82), 49(5), 47(3), 52(7), 53(3)	50(84), 49(4), 47(2), 51(6), 52(4)	50(87), 49(6), 51(3), 52(4)
$\tilde{\tilde{\tau}}_{m,n}$	50(87), 49(7), 48(5), 91(1)	50(88), 48(6), 47(2), 51(4)	50(82), 49(7), 48(5), 51(4), 52(2)	50(82), 49(7), 48(5), 52(6)	50(87), 49(4), 47(3), 51(6)

connection probabilities assumed are in general strong that leads to a large F_n signal. Next, we illustrate the performance for the case of excessively small connection probabilities.

Effect of excessively small connection probabilities: In this paper, we assume that the entries of Λ and Δ are bounded away from 0 and 1, to establish results on the asymptotic distribution of the change-point estimators (see Section 5). This assumption is, however, not needed for establishing consistency and the convergence rate of the estimators. Here we consider DSBMs with small entries in Λ and Δ and illustrate their effect on the performance of the change-point estimators based on simulated results. For DSBMs (4) and (5), we consider $(m, n, n\tau_{m,n}) = (60, 60, 30)$.

(4) **Reallocation of nodes**: Let $K = 2$, $z(i) = I(1 \leq i \leq m/2) + 2I(m/2 + 1 \leq i \leq m)$, $w(2i-1) = 1,\ w(2i) = 2\ \forall 1 \leq i \leq m/2$.
Further, $\Lambda = \Delta = \begin{pmatrix} \frac{1}{n^{\lambda_1}m^{\lambda_2}} & \frac{1}{n^{\lambda_1}m^{\lambda_2}} - \frac{1}{n^{\delta_1}m^{\delta_2}} \\ \frac{1}{n^{\lambda_1}m^{\lambda_2}} - \frac{1}{n^{\delta_1}m^{\delta_2}} & \frac{1}{n^{\lambda_1}m^{\lambda_2}} \end{pmatrix}$
for $(\delta_1, \delta_2, \lambda_1, \lambda_2) = (3/8, 3/8, 1/4, 1/4), (5/8, 1/4, 1/4, 3/8)$.

Table 4: Illustrating the performance of the change-point estimators with $m = 60, n = 60, n\tau_{m,n} = 30$, $\lambda = 1/20$ based on 100 replicates and DSBMs (1), (2) and (3). Figures in brackets are frequencies of the number of change-points the corresponding change-point is observed. Further, $F_n := ||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2$.

	Reallocation of nodes,in (1)			Change in connectivity, in (2)	Merging communities, in (3)
	$\delta = 1/20$	$\delta = 1/10$	$\delta = 1/4$		
F_n	793.67	527.82	154.31	187.35	231.02
$\frac{n}{m^2} F_n$	13.23	8.784	2.572	26.456	3.85
$\frac{n}{K^2} F_n$	11905.11	7905.3	2314.58	23810.26	1540.12
$\frac{Km}{\nu_{m,n}^2}$	0.3	0.455	1.56	6.024	1.021
$\frac{m\sqrt{n}}{\nu_{m,n}^2}$	1.17	1.764	6.024	5.738	2.64
$\tilde{\tau}_{m,n}$	30(88), 29(7), 28(3), 32(2)	30(91), 29(5), 28(1), 31(3)	30(92), 28(2), 31(6)	30(88), 29(8), 28(2), 31(2)	30(89), 29(3), 31(7), 32(1)
$\hat{\tau}_{m,n}$	30(84), 29(7), 32(5), 28(4)	30(87), 29(6), 31(7)	30(78), 29(11), 28(8), 32(3)	30(90), 28(2), 29(8)	30(88), 31(7), 28(5)
$\tilde{\tilde{\tau}}_{m,n}$	30(83), 29(5), 28(7), 31(5)	30(23), 26(23), 41(34), 22(14), 23(6)	30(29), 22(45), 19(10), 41(16)	30(35), 22(21), 24(6), 19(18), 42(20)	30(21), 22(20), 17(48), 37(11)

(5) **Change in connectivity**: Let $K = 2$, $z(i) = w(i) = I(1 \leq i \leq m/2) + 2I(m/2 + 1 \leq i \leq m)$, $\Lambda = \begin{pmatrix} \frac{2}{n^{\lambda_1} m^{\lambda_2}} & \frac{1}{n^{\lambda_1} m^{\lambda_2}} \\ \frac{1}{n^{\lambda_1} m^{\lambda_2}} & \frac{2}{n^{\lambda_1} m^{\lambda_2}} \end{pmatrix}$, $\Delta = \Lambda + \frac{1}{n^{1/8} m^{1/8}} J_2$, $(\lambda_1, \lambda_2) = (1/4, 1/4), (1/4, 3/8)$.

The results are presented in Table 7. For models (4) and (5), SNR-ER is proportional to $n^{-2\delta_1} m^{-2\delta_2}$ and $n^{-1/4} m^{-1/4}$ respectively. The choices of δ_1 and δ_2 taken in (4) suffice to make the connection probabilities in Λ and Δ small enough, so that the resulting SNR-ER is small. Consequently, $\hat{\tau}_{m,n}$ does not perform well in Columns 1 and 2 of Table 7. On the other hand, $\delta_1 = 1/8$ and $\delta_2 = 1/8$ in (5) are adequate to induce a large SNR-ER, as reflected in the improved performance of $\hat{\tau}_{m,n}$ for $\tau_{m,n}$ in Columns 3 and 4 of Table 7. On the other hand, while m/K in Table 7 is large enough to satisfy SNR-DSBM, $\nu_{m,n}$ is proportional to $n^{-\lambda_1} m^{-\lambda_2}$ and by the choices of λ_1, λ_2 in models (4) and (5), quite small, as a consequence of which (A1) does not hold for the settings depicted in Table 7, and the performance of $\tilde{\tilde{\tau}}_n$ suffers. The estimator $\tilde{\tau}_{m,n}$ performs very well throughout Table 7, since SNR-DSBM holds.

Table 5: Illustrating the performance of the change-point estimators with $m = 500, n = 20, n\tau_{m,n} = 10$, $\lambda = 1/20$ based on 100 replicates and DSBMs (1), (2) and (3). Figures in brackets are frequencies of the number of times the corresponding change-point is observed. Further, $F_n := ||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2$.

	Reallocation of nodes, in (1)			Change in connectivity, in (2)	Merging communities, in (3)
	$\delta = 1/20$	$\delta = 1/10$	$\delta = 1/4$		
F_n	49763.4	36881.4	15014.7	99526.8	14501.2
$\frac{n}{m^2} F_n$	3.98	2.95	1.2	7.96	1.13
$\frac{n}{K^2} F_n$	24881.7	184406.8	75070.2	497634	32224.9
$\frac{Km}{\nu_{m,n}^2}$	0.04	0.054	0.13	0.13	0.13
$\frac{m\sqrt{n}}{\nu_{m,n}^2}$	0.089	0.121	0.298	0.298	0.1997
$\tilde{\tau}_{m,n}$	10(89), 9(8), 8(3)	10(87), 9(5), 8(3), 11(5)	10(88), 9(3), 8(2), 11(6), 12(1)	10(90), 9(8), 8(2)	10(89), 9(4), 8(1), 11(6)
$\hat{\tau}_{m,n}$	10(87), 8(4), 11(9)	10(85), 8(8), 11(7)	10(29), 3(23), 6(35), 15(13)	10(85), 9(9), 8(6)	10(28), 7(33), 4(14), 15(25)
$\tilde{\tilde{\tau}}_{m,n}$	10(88), 9(8), 11(4)	10(89), 8(6), 11(5)	10(79), 8(9), 9(5), 11(7)	10(83), 9(10), 8(7)	10(85), 8(7), 9(5) 11(3)

Simulation on setting (G): Consider the setup in setting (G) previously presented and let $n = 20$, $n\tau_{m,n} = 10$, $m = 20$, $K = 2$, $z(i) = w(i) = I(1 \leq i \leq 9) + 2I(10 \leq i \leq 20)$, $p_1^2 = 0.8$, $p_2 = p_1 + 1/\sqrt{n}$, $\Lambda = p_1 I_2$, $\Delta = p_2 I_2$. Simulation results are given in Table 8. In this case, both SNR-DSBM and (A1) hold. Hence, both $\tilde{\tau}_{m,n}$ and $\tilde{\tilde{\tau}}_{m,n}$ perform well as expected. However, due to the failure of SNR-ER to hold, the performance of $\hat{\tau}_{m,n}$ suffers.

5. Asymptotic distribution of change-point estimators and adaptive inference

Up to this point, the analysis focused on establishing consistency results for the derived change-point estimators and the corresponding convergence rates. Nevertheless, it is also of interest to provide confidence intervals, primarily for the change-point estimates. This issue is addressed next for $\tilde{\tau}_{m,n}$, $\tilde{\tilde{\tau}}_{m,n}$, $\tau_{m,n}^*$ and $\hat{\tau}_{m,n}$, and as will be shortly seen the distributions are different depending on the behavior of the norm difference of the parameters before and after the change-point. Since this norm difference is not usually known a priori, we solve this problem through a data-based adaptive procedure to determine the quantiles of the asymptotic distribution, irrespective of the specific regime pertaining to the data at hand.

Table 6: Illustrating the performance of the change-point estimators with $m = 500, n = 100, n\tau_{m,n} = 50$, $\lambda = 1/20$, based on 100 replicates and DSBMs (1), (2) and (3). Figures in brackets are frequencies of the number of times the corresponding change-point is observed. Further, $F_n := ||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2$.

	Reallocation of nodes, in (1)			Change in connectivity, in (2)	Merging communities, in (3)
	$\delta = 1/20$	$\delta = 1/10$	$\delta = 1/4$		
F_n	42365.37	26730.8	67145	84731.13	11943.6
$\frac{n}{m^2} F_n$	16.94	10.69	2.686	33.89	4.65
$\frac{n}{K^2} F_n$	1059139	668271.6	167862.2	2118278	132706.8
$\frac{Km}{\nu_{m,n}^2}$	0.047	0.0748	0.298	0.298	0.157
$\frac{m\sqrt{n}}{\nu_{m,n}^2}$	0.236	0.374	0.49	1.49	0.524
$\tilde{\tau}_{m,n}$	50(88), 49(6), 48(3), 51(3)	50(90), 49(3), 51(5), 52(2)	50(88), 49(4), 48(2), 51(4), 52(2)	50(92), 49(3), 51(5)	50(87), 49(7), 48(2), 51(4)
$\hat{\tau}_{m,n}$	50(88), 48(5), 49(7)	50(91), 49(6), 48(3)	50(78), 49(7), 48(5), 52(8), 53(2)	50(90), 49(6), 48(2), 51(2)	50(87), 49(8), 51(5)
$\tilde{\tilde{\tau}}_{m,n}$	50(83), 49(8), 48(7), 53(2)	50(88), 48(9), 47(2), 51(1)	50(85), 49(5), 48(5), 51(5)	50(82), 49(5), 48(8), 51(2), 52(3)	50(85), 49(8), 48(2), 51(5)

5.1. Form of asymptotic distribution

For ease of presentation, we focus on $\hat{\tau}_{m,n}$, but analogous results hold for $\tau^*_{m,n}$, $\tilde{\tau}_{m,n}$ and $\tilde{\tilde{\tau}}_{m,n}$. As previously mentioned, there are three different regimes for its asymptotic distribution depending on:—(I) $||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2 \to \infty$, (II) $||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2 \to 0$ and (III) $||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F \to c > 0$.

We need additional regularity assumptions (A2)-(A7) for the other regimes. Assumption (A2) stated below ensures that the connection probabilities are bounded away from 0 and 1, which gives rise to a dense graph and ensures the positive asymptotic variance of the change-point estimators.

(A2) For some $c > 0$, $0 < c < \inf_{u,v} \lambda_{uv}, \inf_{u,v} \delta_{uv} \leq \sup_{u,v} \lambda_{uv}, \sup_{u,v} \delta_{uv} < 1 - c < 1$.

The precise statements of (A3)-(A7) are given in Section 7.10, but a brief discussion of their roles is presented below.

Assumption (A3) is required in Regime II and guarantees the existence of the asymptotic variance of the change-point estimator. In Theorem 26(b), this variance is denoted by γ^2.

Table 7: Illustrating the performance of the change-point estimators with $m = 60, n = 60, n\tau_{m,n} = 30$, based on 100 replicates and DSBMs (4) and (5). Figures in brackets are frequencies of the number of times the corresponding change-point is observed. Further, $F_n := ||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2$.

	Reallocation of nodes, in (4)		Change in connectivity, in (5)	
	$(\delta_1, \delta_2, \lambda_1, \lambda_2) = (3/8, 3/8, 1/4, 1/4)$	$(\delta_1, \delta_2, \lambda_1, \lambda_2) = (5/8, 1/4, 1/4, 3/8)$	$(\lambda_1, \lambda_2) = (1/4, 1/4)$	$(\lambda_1, \lambda_2) = (1/4, 3/8)$
F_n	3.873	1.3915	464.758	464.758
$\frac{n}{m^2}F_n$	0.0645	0.0232	7.746	7.746
$\frac{n}{K^2}F_n$	58.095	20.874	6971.37	6971.37
$\frac{Km}{\nu_{m,n}^2}$	61.968	172.466	8	22.265
$\frac{m\sqrt{n}}{\nu_{m,n}^2}$	240	667.96	30.984	86.232
$\tilde{\tau}_{m,n}$	30(85), 29(6), 28(3), 31(5), 32(1)	30(88), 29(4), 31(8)	30(91), 29(8), 28(1)	30(89), 29(5), 28(2), 31(4)
$\hat{\tau}_{m,n}$	30(15), 27(18), 24(38), 22(19), 41(10)	30(21), 28(4), 18(31), 39(14), 47(19), 49(11)	30(88), 29(9), 32(3)	30(88), 28(6), 31(6)
$\tilde{\tilde{\tau}}_{m,n}$	30(7), 28(23), 22(13), 18(18), 38(20),43(19)	26(15), 22(23), 32(27), 39(25), 43(10)	30(14), 25(8), 23(28), 38(34), 43(16)	30(15), 26(10), 21(25), 21(28), 39(14), 44(8)

Table 8: Illustrating the performance of change-point estimators with $m = 20, n = 20, n\tau_{m,n} = 10$ based on 100 replicates for DSBMs in setting (G). Figures in brackets are frequencies of the number of times the corresponding change-point is observed. Further, $F_n := ||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2$.

	F_n	$\frac{n}{m^2}F_n$	$\frac{n}{K^2}F_n$	$\frac{Km}{\nu_{m,n}^2}$	$\frac{m\sqrt{n}}{\nu_{m,n}^2}$	$\tilde{\tau}_{m,n}$	$\hat{\tau}_{m,n}$	$\tilde{\tilde{\tau}}_{m,n}$
On (G)	10.1	0.51	50.5	0.62	1.38	10(90), 9(5), 8(2), 11(4)	4(42), 5(33), 8(12), 10(5), 14(8)	10(78), 9(15), 8(5), 12(2)

In Regime III, we consider the following set of edges

$$\mathcal{K}_n = \{(i,j) : 1 \leq i,j \leq m, \ \ |\lambda_{z(i)z(j)} - \delta_{w(i)w(j)}| \to 0\} \tag{5.1}$$

and treat edges in $\mathcal{K}_n$ and $\mathcal{K}_0 = \mathcal{K}_n^c$ separately. Note that in Regime II, $\mathcal{K}_n = \{(i,j) : 1 \leq i,j \leq m\}$ is the set of all edges. Hence, we can treat $\mathcal{K}_n$ in a similar way as in Regime II. The role of (A4) in Regime III is analogous to that of (A3) in Regime II. In the limit, $\mathcal{K}_n$ contributes a Gaussian process with a triangular drift term. (A4) ensures the existence of the asymptotic variance $\tilde{\gamma}^2$ of the limiting Gaussian process as well as the drift c_1^2. (A5) is a technical assumption and is required for establishing asymptotic normality on $\mathcal{K}_n$. Moreover, $\mathcal{K}_0$ is a finite set. (A6) guarantees that $\mathcal{K}_0$ does not vary with n. (A7) guarantees that $\tau_{m,n} \to \tau^*$ for some $\tau^* \in (c^*, 1-c^*)$, $\lambda_{z(i)z(j)} \to a^*_{ij,1}$ and $\delta_{w(i)w(j)} \to a^*_{ij,2}$ for all $(i,j) \in \mathcal{K}_0$. Consider the collection of independent Bernoulli random variables $\{A^*_{ij,l} : (i,j) \in \mathcal{K}_0, l = 1,2\}$ with $E(A^*_{ij,l}) = a^*_{ij,l}$. Then, (A7) implies $A_{ij,(\lfloor nf \rfloor, n)} \xrightarrow{\mathcal{D}} A^*_{ij,1} I(f < \tau^*) + A^*_{ij,2} I(f > \tau^*)\ \forall (i,j) \in \mathcal{K}_0$.

The following Theorem summarizes the asymptotic distribution results.

Theorem 26 *Suppose SNR-ER holds for $\hat{\tau}_{m,n}$, SNR-DSBM, (NS) and (A1) hold for $\tilde{\tilde{\tau}}_{m,n}$, SNR-DSBM, (NS) and (A1*) hold for $\tau^*_{m,n}$ and SNR-DSBM holds for $\tilde{\tau}_{m,n}$. Then, the following statements are true. Let $\hat{\eta}_{m,n}$ denote generically any of the following change-point estimators: $\hat{\tau}_{m,n}$, $\tilde{\tau}_{m,n}$, $\tilde{\tilde{\tau}}_{m,n}$ and $\tau^*_{m,n}$.*

(a) *If $||Ed_z(\Lambda) - Ed_w(\Delta)||_F^2 \to \infty$, then* $\lim_{n\to\infty} P(\hat{\eta}_{m,n} = \tau_{m,n}) = 1$.

(b) *If (A2) and (A3) hold and $||Ed_z(\Lambda) - Ed_w(\Delta)||_F^2 \to 0$, then*

$$n||Ed_z(\Lambda) - Ed_w(\Delta)||_F^2 (\hat{\eta}_{m,n} - \tau_{m,n}) \xrightarrow{\mathcal{D}} \gamma^2 \arg\max_{h\in\mathbb{R}} (-0.5|h| + B_h), \tag{5.2}$$

where B_h denotes the standard Brownian motion.

(c) *Suppose (A2), (A4)-(A7) hold and $||Ed_z(\Lambda) - Ed_w(\Delta)||_F \to c > 0$, then*

$$n(\hat{\eta}_{m,n} - \tau_{m,n}) \quad \xrightarrow{\mathcal{D}} \quad \arg\max_{h\in\mathbb{Z}} (D(h) + C(h) + A(h))$$

where for each $h \in \mathbb{Z}$,

$$\begin{aligned} D(h+1) - D(h) &= 0.5\text{Sign}(-h)c_1^2, && (5.3)\\ C(h+1) - C(h) &= \tilde{\gamma} W_h, \quad W_h \overset{i.i.d.}{\sim} \mathcal{N}(0,1), && (5.4)\\ A(h+1) - A(h) &= \sum_{k\in\mathcal{K}_0} \Big[(Z_{ij,h} - a^*_{ij,1})^2 - (Z_{ij,h} - a^*_{ij,2})^2 \Big], && (5.5) \end{aligned}$$

$\{Z_{ij,h}\}$ are independently distributed with $Z_{ij,h} \overset{d}{=} A^*_{ij,1} I(h < 0) + A^*_{ij,2} I(h \geq 0)$ for all $(i,j) \in \mathcal{K}_0$.

Remark 27 *As we have already noted, consistency of the change-point estimators holds for both dense and sparse graphs. The same conclusion holds for the asymptotic distribution under Regime I. However, (A2) is a crucial assumption for establishing the asymptotic distribution of the change-point estimator under Regimes II and III. (A2) implies that the random graph is dense. The different statistical and probabilistic aspects of sparse random graphs constitute a growing area of research in the recent literature. Most of the results in the sparse setting do not follow from the dense case and different tools and techniques are needed for their analysis; see Remark 36 for examples. Though the convergence rate results established in Sections 2 and 3 hold for the sparse setting, deriving the asymptotic distribution of the change-point estimator under Regimes II and III in sparse random graphs requires further investigation.*

5.2. Adaptive Inference

Next, we present a data adaptive procedure that does *not* require a priori knowledge of the limiting regime. Recall the estimators $\hat{\tau}_{m,n}$, $\hat{\hat{\Lambda}}$, $\hat{\hat{\Delta}}$, $\hat{z}$ and $\hat{w}$ of the parameters in the DSBM model given in (2.1). We generate independent $m \times m$ adjacency matrices $A_{t,n,\text{DSBM}}$, $1 \le t \le n$, where

$$A_{t,n,\text{DSBM}} = ((A_{ij,(t,n),\text{DSBM}})) \sim \begin{cases} \text{SBM}(\hat{z}, \hat{\hat{\Lambda}}), & \text{if } 1 \le t \le \lfloor n\hat{\tau}_{m,n} \rfloor \\ \text{SBM}(\hat{w}, \hat{\hat{\Delta}}), & \text{if } \lfloor n\hat{\tau}_{m,n} \rfloor < t < n. \end{cases} \tag{5.6}$$

Obtain

$$\hat{h}_{\text{DSBM}} = \arg \max_{h \in (n(c^* - \hat{\tau}_{m,n}), n(1-c^*-\hat{\tau}_{m,n}))} \tilde{L}^*(\hat{\tau}_{m,n} + h/n, \hat{z}, \hat{w}, \hat{\hat{\Lambda}}, \hat{\hat{\Delta}}) \tag{5.7}$$

where

$$\begin{aligned} \tilde{L}^*(\hat{\tau}_{m,n} + h/n, \hat{z}, \hat{w}, \hat{\hat{\Lambda}}, \hat{\hat{\Delta}}) \quad &= \quad \frac{1}{n} \sum_{i,j=1}^{m} \Bigg[\sum_{t=1}^{n\hat{\tau}_{m,n}+h} (A_{ij,(t,n),\text{DSBM}} - \hat{\hat{\lambda}}_{\hat{z}(i),\hat{z}(j)})^2 \\ &\quad + \sum_{t=n\hat{\tau}_{m,n}+h+1}^{n} (A_{ij,(t,n),\text{DSBM}} - \hat{\hat{\delta}}_{\hat{w}(i),\hat{w}(j)})^2 \Bigg]. \end{aligned} \tag{5.8}$$

Theorem 28 states the asymptotic distribution of $\hat{h}_{\text{DSBM}}$ under a stronger identifiability condition. Specifically,

SNR-ER-ADAP: $\frac{\sqrt{n}}{m^2\sqrt{\log m}}||\text{Ed}_z(\Lambda) - \text{Ed}_w(\Delta)||_F^2 \to \infty$

It is easy to show that SNR-ER-ADAP holds if all assumptions in Proposition 24 hold and

(AD) $m = e^{n^{\delta_3}}$ for some $\delta_1, \delta_2, \delta_3 > 0$ and $0 < \delta_1 + \delta_2 + \delta_3/2 < 1/2$ (δ_1, δ_2 are as in Proposition 24)
is satisfied.

Specifically, Examples (A)-(D) in Section 4 satisfy SNR-ER-ADAP in the presence of condition (AD).

We also need the following condition to ensure that $\hat{z}$ and $\hat{w}$ are consistent estimates for z and w, respectively.

(A1-ADAP) $\frac{Km}{n\nu_{m,n}^2}||\text{Ed}_z(\Lambda) - \text{Ed}_w(\Delta)||_2^{-1} \to 0.$

Under SNR-ER-ADAP, one can reduce A1-ADAP to $\frac{K}{(n^3 \log m)^{1/4}\nu_{m,n}^2} \to O(1)$. This holds whenever within and between community connection probabilities are equal (that is $\lambda_{ij} = q_1, \delta_{ij} = q_2 \ \forall \ i \neq j$ and $\lambda_{ii} = p_1, \delta_{ii} = p_2 \ \forall i$), balanced communities of size $O(m/K)$ are present, and their number is $K = O(m^{2/3})$. This is because the first two conditions implies $\nu_{m,n} = O(m/k)$ (for example see Example 1).

We also require $\log m = o(\sqrt{n})$, so that the entries of $\text{Ed}_{\hat{z}}(\hat{\hat{\Lambda}})$ and $\text{Ed}_{\hat{w}}(\hat{\hat{\Delta}})$ are bounded away from 0 and 1. Note that this assumption implies $0 < \delta_3 < 1/2$ in (AD).

Theorem 28 ***(Asymptotic distribution of*** $\hat{h}_{DSBM}$***)*** *Suppose (A2), SNR-ER-ADAP and A1-ADAP hold and* $\log m = o(\sqrt{n})$. *Then, the following results are true.*

(a) If $||Ed_z(\Lambda) - Ed_w(\Delta)||_F \to \infty$, *then* $\lim_{n\to\infty} P(\hat{h}_{DSBM} = 0) = 1$

(b) If (A3) holds and $||Ed_z(\Lambda) - Ed_w(\Delta)||_F \to 0$, *then*

$$||Ed_z(\Lambda) - Ed_w(\Delta)||_F^2 \hat{h}_{DSBM} \xrightarrow{\mathcal{D}} \gamma^{-2} \arg\max_{h\in\mathbb{R}}(-0.5|h| + B_h) \tag{5.9}$$

where B_h *corresponds to a standard Brownian motion.*

(c) If (A4)-(A7) hold and $||Ed_z(\Lambda) - Ed_w(\Delta)||_F \to c > 0$, *then*

$$\hat{h}_{DSBM} \quad \xrightarrow{\mathcal{D}} \quad \arg\max_{h\in\mathbb{Z}}(D(h) + C(h) + A(h)),$$

where $D(\cdot)$, $C(\cdot)$ *and* $A(\cdot)$ *are same as (5.3)-(5.5).*

The proof of the Theorem is given in Section 7.9.

Remark 29 *Though SNR-ER-ADAP is stronger than SNR-ER, we have not been able to relax it. Data-driven methods usually require stronger assumptions even in simple models like the Erdős-Rényi random graph one, with a single change-point—see for example Bhattacharjee et al. (2017). To establish the form of the asymptotic distribution of* $\hat{h}_{DSBM}$, *we need to establish convergence in probability as stated in Lemma 34 which do not occur without SNR-ER-ADAP.*

Remark 30 *Similar conclusions as in Theorem 28 hold for* $\tilde{\tilde{\tau}}_{m,n}$ *and* $\tau^*_{m,n}$, *but under the stronger assumptions (A1) and (A1*), respectively, instead of A1-ADAP. Details are omitted to avoid repetition. Moreover, as discussed in Sections 2 and 4, (A1) and (A1*) are difficult to satisfy whereas SNR-ER-ADAP may often hold (see discussion after stating the assumption). Also computation of* $\tilde{\tilde{\tau}}_{m,n}$ *is itself expensive and performing adaptive inference for it will make it even more so. For these reasons, we have mainly focused on adaptive inference for* $\hat{\tau}_{m,n}$.

Remark 31 *Note that the asymptotic distribution of* $\hat{h}_{DSBM}$ *is identical to the asymptotic distribution of* $\hat{\tau}_{m,n}$. *Therefore, in practice we can simulate* $\hat{h}_{DSBM}$ *for a large number of replicates and use their empirical quantiles as estimates of the quantiles of the limiting distribution under the (unknown) true regime. Moreover, the adaptive inference is a computationally expensive procedure and comes at a certain cost, namely the stronger assumption SNR-ER-ADAP.*

6. Concluding Remarks

In this paper, we have addressed the change-point problem in the context of DSBM. We establish consistency of the change-point estimator under a suitable identifiability condition and a second condition that controls the misclassification rate arising from using clustering for assigning nodes to communities and discuss the stringency of the latter condition. Further, we propose a fast computational strategy that ignores the underlying community structure but provides a consistent estimate of the change-point. Further, for both methods under their respective identifiability and certain additional regularity conditions, we establish rates of convergence and derive the asymptotic distributions of the change-point estimators.

In addition, this work identifies an interesting issue that requires further research; namely, a range of models where the SNR-DSBM identifiability condition holds, but the misclassification rate condition (A1) needed for the "every time point clustering algorithm" and the identifiability condition (SNR-ER) of the alternative strategy fails to hold. In that range, no general strategy for solving the change-point problem for DSBM seems to be currently available.

Note that even adopting a very simple (almost toy-like) structure for the DSBM (for example the framework in Gao et al. (2017)), detecting the change-point remains a hard problem if neither (A1) or (SNR-ER) hold. The modified algorithm proposed in Remark 17 is a good possibility, but the limitations discussed in Remark 18 remain a concern. At present, we are not aware of any detection algorithm that works by imposing only SNR-DSBM type assumptions for the change-point problem in DSBM.

Acknowledgments. We are thankful to Dr. Daniel Sussman for valuable comments and suggestions. We also thank the two anonymous referees for many helpful comments and suggestions that led to improvements in the structure of the paper and presentation of the material.

The work of Monika Bhattacharjee was entirely supported by a fellowship of Informatics Institute, University of Florida, USA. Currently, Monika Bhattacharjee is at the Department of Mathematics, Indian Institute of Technology Bombay.

The work of Moulinath Banerjee was supported by NSF DMS-1712962.

The work of George Michailidis was supported in part by NSF grants DMS 1830175, IIS 1632730 and DMS 1821220.

7. Proofs and Other Technical Material

Throughout this section, C is a generic positive constant. Often we write $\tau_{m,n}$ as τ.

7.1. Proof of Theorem 8

Without loss of generality, assume $\tau < b$. By Lemmas 5.1 and 5.3 of Lei and Rinaldo (2015), with probability tending to 1, we have

$$\mathcal{M}_{b,n,m} \quad \leq \quad C\frac{K}{n\nu_{m,n}^2}\Big[||\frac{1}{n}\sum_{t=1}^{n\tau}(A_{t,n} - \mathrm{Ed}_z(\Lambda))||_F^2 + ||\frac{1}{n}\sum_{t=n\tau+1}^{nb}(A_{t,n} - \mathrm{Ed}_w(\Delta))||_F^2$$

$$+||\frac{1}{n}\sum_{t=n\tau+1}^{nb}(\mathrm{E}d_z(\Lambda)-\mathrm{E}d_w(\Delta))||_F^2\Big]$$
$$= C\frac{K}{n\nu_{m,n}^2}(A_1+A_2+|\tau-b|\ ||\mathrm{E}d_z(\Lambda)-\mathrm{E}d_w(\Delta)||_F^2),\ \text{say}.$$

Now by Theorem 5.2 of Lei and Rinaldo (2015), $A_1, A_2 = O_{\mathrm{P}}(m)$. Thus,

$$\mathcal{M}_{b,n,m} = O_{\mathrm{P}}\left(\frac{K}{n\nu_{m,n}^2}(m+|\tau-b|||\mathrm{E}d_z(\Lambda)-\mathrm{E}d_w(\Delta)||_F^2)\right).$$

This completes the proof of Theorem 8. ■

7.2. Selected useful lemmas

The following two lemmas directly quoted from van der Vaart and Wellner (1996) are needed to establish Theorems 11 and 26.

Lemma 32 *For each n, let $\mathbb{M}_n$ and $\tilde{\mathbb{M}}_n$ be stochastic processes indexed by a set $\mathcal{T}$. Let τ_n (possibly random) $\in \mathcal{T}_n \subset \mathcal{T}$ and $d_n(b,\tau_n)$ be a map (possibly random) from $\mathcal{T}$ to $[0,\infty)$. Suppose that for every large n and $\delta \in (0,\infty)$*

$$\sup_{\delta/2<d_n(b,\tau_n)<\delta,\ b\in\mathcal{T}} (\tilde{\mathbb{M}}_n(b)-\tilde{\mathbb{M}}_n(\tau_n)) \le -C\delta^2, \tag{7.1}$$
$$E\sup_{\delta/2<d_n(b,\tau_n)<\delta,\ b\in\mathcal{T}} \sqrt{n}|\mathbb{M}_n(b)-\mathbb{M}_n(\tau_n)-(\tilde{\mathbb{M}}_n(b)-\tilde{\mathbb{M}}_n(\tau_n))| \le C\phi_n(\delta), \tag{7.2}$$

for some $C>0$ and for function ϕ_n such that $\delta^{-\alpha}\phi_n(\delta)$ is decreasing in δ on $(0,\infty)$ for some $\alpha<2$. Let r_n satisfy

$$r_n^2\phi(r_n^{-1}) \le \sqrt{n} \quad \textit{for every } n. \tag{7.3}$$

Further, suppose that the sequence $\{\hat{\tau}_{m,n}\}$ takes its values in $\mathcal{T}_n$ and satisfies $\mathbb{M}_n(\hat{\tau}_n) \ge \mathbb{M}_n(\tau_n) - O_P(r_n^{-2})$ for large enough n. Then, $r_n d_n(\hat{\tau}_n,\tau_n) = O_P(1)$.

Lemma 33 *Let $\mathbb{M}_n$ and $\mathbb{M}$ be two stochastic processes indexed by a metric space $\mathcal{T}$, such that $\mathbb{M}_n \Rightarrow \mathbb{M}$ in $l^\infty(\mathcal{C})$ for every compact set $\mathcal{C}\subset\mathcal{T}$, that is,*

$$\sup_{h\in\mathcal{C}}|\mathbb{M}_n(h)-\mathbb{M}(h)| \xrightarrow{P} 0. \tag{7.4}$$

Suppose that almost all sample paths $h\to\mathbb{M}(h)$ are upper semi-continuous and possess a unique maximum at a (random) point $\hat{h}$, which as a random map in $\mathcal{T}$ is tight. If the sequence $\hat{h}_n$ is uniformly tight and satisfies $\mathbb{M}_n(\hat{h}_n) \ge \sup_n \mathbb{M}_n(h) - o_P(1)$, then $\hat{h}_n \xrightarrow{\mathcal{D}} \hat{h}$ in $\mathcal{T}$.

The following lemma is needed in the proof of Theorem 28.

Lemma 34 *Suppose SNR-ER-ADAP, A1-ADAP holds and* $\log m = o(\sqrt{n})$*. Then, the following statements hold.*

(a) $\frac{||Ed_{\hat{z}}(\hat{\hat{\Lambda}})-Ed_{\hat{w}}(\hat{\hat{\Delta}})||_F^2}{||Ed_z(\Lambda)-Ed_w(\Delta)||_F^2} \xrightarrow{P} 1.$

(b) If $||Ed_z(\Lambda) - Ed_w(\Delta)||_F^2 \to 0$*, then*

$$\frac{\sum_{i,j=1}^m (\lambda_{z(i)z(j)} - \delta_{w(i)w(j)})^2 \lambda_{z(i)z(j)}(1-\lambda_{z(i)z(j)})}{||Ed_z(\Lambda) - Ed_w(\Delta)||_F^2} \xrightarrow{P} \gamma^2,$$
$$\frac{\sum_{i,j=1}^m (\lambda_{z(i)z(j)} - \delta_{w(i)w(j)})^2 \delta_{w(i)w(j)}(1-\delta_{w(i)w(j)})}{||Ed_z(\Lambda) - Ed_w(\Delta)||_F^2} \xrightarrow{P} \gamma^2.$$

(c) If $||Ed_z(\Lambda) - Ed_w(\Delta)||_F^2 \to c^2 > 0$*, then*

$$\sum_{i,j \in \mathcal{K}_n} (\lambda_{z(i)z(j)} - \delta_{w(i)w(j)})^2 \lambda_{z(i)z(j)}(1-\lambda_{z(i)z(j)}) \xrightarrow{P} \tilde{\gamma}^2,$$
$$\sum_{i,j \in \mathcal{K}_n} (\lambda_{z(i)z(j)} - \delta_{w(i)w(j)})^2 \delta_{w(i)w(j)}(1-\delta_{w(i)w(j)}) \xrightarrow{P} \tilde{\gamma}^2.$$

Proof We only show the proof of part (a), since parts (b) and (c) follow employing similar arguments.

$$\begin{aligned}
\left|\frac{||\mathrm{Ed}_{\hat{z}}(\hat{\hat{\Lambda}}) - \mathrm{Ed}_{\hat{w}}(\hat{\hat{\Delta}})||_F^2}{||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2} - 1\right| &= \frac{|||\mathrm{Ed}_{\hat{z}}(\hat{\hat{\Lambda}}) - \mathrm{Ed}_{\hat{w}}(\hat{\hat{\Delta}})||_F^2 - ||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2|}{||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2} \\
&\le \frac{||\mathrm{Ed}_{\hat{z}}(\hat{\hat{\Lambda}}) - \mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_{\hat{w}}(\hat{\hat{\Delta}}) + \mathrm{Ed}_w(\Delta)||_F^2}{||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2} \\
&\le \frac{||\mathrm{Ed}_{\hat{z}}(\hat{\hat{\Lambda}}) - \mathrm{Ed}_z(\Lambda)||_F^2 + ||\mathrm{Ed}_{\hat{w}}(\hat{\hat{\Delta}}) - \mathrm{Ed}_w(\Delta)||_F^2}{||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2}
\end{aligned}$$

Therefore, part (a) follows from Theorem 13, SNR-ER-ADAP, A1-ADAP and $\log m = o(\sqrt{n})$. ■

7.3. Proof of Theorem 11

Throughout this proof, we use the following simplified notation for ease of exposition: $A_{ijt} = A_{ij,(t,n)}, z_1 = \tilde{z}_{b,n,m}, z_2 = \tilde{z}_{\tau_{m,n},n,m}, w_1 = \tilde{w}_{b,n,m}, w_2 = \tilde{w}_{\tau_{m,n},n,m},\ \Lambda_1 = \tilde{\Lambda}_{\tilde{z}_{b,n,m},(b,n),m},\ \Lambda_2 = \tilde{\Lambda}_{\tilde{z}_{\tau_{m,n},n,m},(\tau_{m,n},n),m},\ \Lambda_3 = \tilde{\Lambda}_{\tilde{z}_{\tau_{m,n},n,(b,n),m}},\ \Delta_1 = \tilde{\Delta}_{\tilde{w}_{b,n,m},(b,n),m},\ \Delta_2 = \tilde{\Delta}_{\tilde{w}_{\tau_{m,n},n,m},(\tau_n,n),m},$ $\Delta_w = \tilde{\Delta}_{w,(b,n),m},\ \lambda_{uv,1} = \tilde{\lambda}_{uv,\tilde{z}_{b,n,m},(b,n),m},\ \lambda_{uv,2} = \tilde{\lambda}_{uv,\tilde{z}_{\tau_{m,n},n,m},(\tau_{m,n},n),m},$ $\lambda_{uv,3} = \tilde{\lambda}_{uv,\tilde{z}_{\tau_{m,n},n,(b,n),m}},\ \delta_{uv,1} = \tilde{\delta}_{uv,\tilde{w}_{b,n,m},(b,n),m},\ \delta_{uv,2} = \tilde{\delta}_{uv,\tilde{w}_{\tau_{m,n},n,m},(\tau_{m,n},n),m},\ \delta_{uv,w} = \tilde{\delta}_{uv,w,(b,n),m}$. Suppose $b < \tau_{m,n}$. Similar arguments work when $b > \tau_{m,n}$. Note that

$$\tilde{\tilde{\tau}}_{m,n} = \arg\min_{b \in (c^*, 1-c^*)} \tilde{L}(b, z_1, w_1, \Lambda_1, \Delta_1)$$

where

$$\tilde{L}(b, z_1, w_1, \Lambda_1, \Delta_1) = \frac{1}{n} \sum_{i,j=1}^{m} \left[\sum_{t=1}^{nb} (A_{ijt} - \lambda_{z_1(i) z_1(j),1})^2 + \sum_{t=nb+1}^{n} (A_{ijt} - \hat{\delta}_{w_1(i) w_1(i),1})^2 \right]. \quad (7.5)$$

To prove Theorem 11, we need Lemma 32 quoted from van der Vaart and Wellner (1996). For our purpose, we make use of the above lemma with $\mathbb{M}_n(\cdot) = \tilde{L}(\cdot, \tilde{z}_{\cdot,n,m}, \tilde{w}_{\cdot,n,m}, \tilde{\Lambda}_{\tilde{z}_{\cdot,n,m},(\cdot,n),m}, \tilde{\Delta}_{\tilde{w}_{\cdot,n,m},(\cdot,n),m})$, $\widetilde{\mathbb{M}}_n(\cdot) = E\tilde{L}(\cdot, \tilde{z}_{\cdot,n,m}, \tilde{w}_{\cdot,n,m}, \tilde{\Lambda}_{\tilde{z}_{\cdot,n,m},(\cdot,n),m}, \tilde{\Delta}_{\tilde{w}_{\cdot,n,m},(\cdot,n),m})$, $\mathcal{T} = [0,1]$, $\mathcal{T}_n = \{1/n, 2/n, \ldots, (n-1)/n, 1\} \cap [c^*, 1-c^*]$, $d_n(b, \tau_{m,n}) = ||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F \sqrt{|b - \tau_{m,n}|}$, $\phi_n(\delta) = \delta$, $\alpha = 1.5$, $r_n = \sqrt{n}(\rho_{m,n})^{-1/2}$ and $\hat{\tau}_n = \tilde{\tilde{\tau}}_{m,n}$. Thus, to prove Theorem 11, it suffices to establish that for some $C > 0$,

$$E(\mathbb{M}_n(b) - \mathbb{M}_n(\tau_{m,n})) \leq -C||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2 |b - \tau_{m,n}| \quad \text{and} \quad (7.6)$$

$$E \sup_{\substack{\delta/2 < d_n(b,\tau_{m,n}) < \delta, \\ b \in \mathcal{T}}} |\mathbb{M}_n(b) - \mathbb{M}_n(\tau_{m,n}) - E(\mathbb{M}_n(b) - \mathbb{M}_n(\tau_{m,n}))| \leq C \frac{\delta \sqrt{\rho_{m,n}}}{\sqrt{n}}. \quad (7.7)$$

As the right side of (7.6) and (7.7) are independent of z_1, z_2, w_1, w_2, it suffices to show

$$E^*(\mathbb{M}_n(b) - \mathbb{M}_n(\tau_{m,n})) \leq -C||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2 |b - \tau_{m,n}| \quad \text{and} \quad (7.8)$$

$$E^* \sup_{\substack{\delta/2 < d_n(b,\tau_{m,n}) < \delta, \\ b \in \mathcal{T}}} |\mathbb{M}_n(b) - \mathbb{M}_n(\tau_{m,n}) - E(\mathbb{M}_n(b) - \mathbb{M}_n(\tau_{m,n}))| \leq C \frac{\delta \sqrt{\rho_{m,n}}}{\sqrt{n}}. \quad (7.9)$$

where $E^*(\cdot) = E(\cdot | z_1, w_1, z_2, w_2)$. Similarly denote $V^* = V(\cdot | z_1, z_2, w_1, w_2)$ and $\mathrm{Cov}^*(\cdot) = \mathrm{Cov}(\cdot | z_1, z_2, w_1, w_2)$.

Note that the left hand side of (7.9) is dominated by

$$\left(E^* \sup_{\delta/2 < d_n(b,\tau_{m,n}) < \delta, \ b \in \mathcal{T}} (\mathbb{M}_n(b) - \mathbb{M}_n(\tau_{m,n}) - E(\mathbb{M}_n(b) - \mathbb{M}_n(\tau_{m,n})))^2 \right)^{1/2}. \quad (7.10)$$

By Doob's martingale inequality, (7.10) is further dominated by

$$(\mathrm{V}^*(\mathbb{M}_n(b) - \mathbb{M}_n(\tau_{m,n})))^{1/2} \quad \text{where } d_n(b, \tau_{m,n}) = \delta. \quad (7.11)$$

Thus, to prove Theorem 11, it suffices to show that for some $C > 0$,

$$\mathrm{V}^*(\mathbb{M}_n(b) - \mathbb{M}_n(\tau_{m,n})) \leq C n^{-1} d_n^2(b, \tau_{m,n}) \rho_{m,n}. \quad (7.12)$$

Hence, it sufficies to prove (7.8) and (7.12) to establish Theorem 11. We shall prove these for $b < \tau_{m,n}$. Similar arguments work when $b \geq \tau_{m,n}$.

Denote by $L_1 = \tilde{L}(b, z_1, w_1, \Lambda_1, \Delta_1)$ and $L_2 = \tilde{L}(\tau, z_2, w_2, \Lambda_2, \Delta_2)$. Hence,

$$L_1 - L_2 = A(b) + B(b) + D(b), \quad (7.13)$$

where

$$A(b) = \frac{1}{n} \sum_{i,j=1}^{m} \sum_{t=1}^{nb} \left[(A_{ijt} - \lambda_{z_1(i) z_1(j),1})^2 - (A_{ijt} - \lambda_{z_2(i) z_2(j),2})^2 \right],$$

$$
\begin{aligned}
B(b) &= \frac{1}{n}\sum_{i,j=1}^{m}\sum_{t=nb+1}^{n\tau}\Big[(A_{ijt}-\delta_{w_1(i)w_1(j),1})^2-(A_{ijt}-\delta_{z_2(i)z_2(j),2})^2\Big],\\
D(b) &= \frac{1}{n}\sum_{i,j=1}^{m}\sum_{t=n\tau+1}^{n}\Big[(A_{ijt}-\delta_{w_1(i)w_1(j),1})^2-(A_{ijt}-\delta_{w_2(i)w_2(j),2})^2\Big].
\end{aligned}
$$

Consider the first term of $A(b)$ as follows.

$$
\begin{aligned}
&\frac{1}{nb}\sum_{t=1}^{nb}\sum_{i,j=1}^{m}(A_{ijt}-\lambda_{z_1(i)z_1(j),1})^2\\
=\;&\frac{1}{nb}\sum_{t=1}^{nb}\sum_{i,j=1}^{m}A_{ijt}^2+\sum_{u,v=1}^{K}s_{u,z_1}s_{v,z_1}(\lambda_{uv,1})^2-\sum_{u,v=1}^{K}s_{u,z_1}s_{v,z_1}(\lambda_{z_1(i)z_1(j),1})\sum_{\substack{i:z_1(i)=u\\ j:z_1(j)=v}}\sum_{t=1}^{nb}\frac{A_{ijt}}{nb}\\
=\;&\frac{1}{nb}\sum_{t=1}^{nb}\sum_{i,j=1}^{m}A_{ijt}^2-\sum_{u,v=1}^{K}s_{u,z_1}s_{v,z_1}(\lambda_{uv,1})^2.
\end{aligned}
$$

Similarly, the second term of $A(b)$ is

$$
\begin{aligned}
\frac{1}{nb}\sum_{t=1}^{nb}\sum_{i,j=1}^{m}(A_{ijt}-\lambda_{z_2(i)z_2(j),2})^2 &= \frac{1}{nb}\sum_{t=1}^{nb}\sum_{i,j=1}^{m}A_{ijt}^2+\sum_{u,v=1}^{K}s_{u,z_2}s_{v,z_2}(\lambda_{uv,2})^2\\
&\quad -2\sum_{u,v=1}^{K}s_{u,z_2}s_{v,z_2}\lambda_{uv,2}\lambda_{uv,3}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
bE^*(A(b)) &= -\sum_{u,v=1}^{K}s_{u,z_2}s_{v,z_2}E^*(\lambda_{uv,2})^2-b\sum_{u,v=1}^{K}s_{u,z_1}s_{v,z_1}E^*(\lambda_{uv,1})^2\\
&\quad +2\sum_{u,v=1}^{K}s_{u,z_2}s_{v,z_2}E^*(\lambda_{uv,2}\lambda_{uv,3}).
\end{aligned}
$$

Let $S((u,v,f),(a,b,g))$ be the total number of edges which connect communities u and v under community structure f, and also communities a and b under community structure g. Therefore,

$$
\begin{aligned}
E^*(\lambda_{uv,1})^2 &= V^*(\lambda_{uv,1})+(E(\lambda_{uv,1}))^2\\
&= \frac{1}{nb}\frac{1}{(s_{u,z_1}s_{v,z_1})^2}\sum_{a,b=1}^{K}S((u,v,z_1),(a,b,z))\lambda_{ab}(1-\lambda_{ab})\\
&\quad +\left(\frac{1}{s_{u,z_1}s_{v,z_1}}\sum_{a,b=1}^{K}S((u,v,z_1),(a,b,z))\lambda_{ab}\right)^2,
\end{aligned}
$$

$$
\begin{aligned}
E^*(\lambda_{uv,2})^2 &= V^*(\lambda_{uv,2}) + (E(\lambda_{uv,2}))^2 \\
&= \frac{1}{n\tau}\frac{1}{(s_{u,z_2}s_{v,z_2})^2}\sum_{a,b=1}^{K} S((u,v,z_2),(a,b,z))\lambda_{ab}(1-\lambda_{ab}) \\
&\quad + \left(\frac{1}{s_{u,z_2}s_{v,z_2}}\sum_{a,b=1}^{K} S((u,v,z_2),(a,b,z))\lambda_{ab}\right)^2, \\
E^*(\lambda_{uv,2}\lambda_{uv,3}) &= \mathrm{Cov}^*(\lambda_{uv,2},\lambda_{uv,3}) + (E(\lambda_{uv,2}))(E(\lambda_{uv,3})) \\
&= \frac{1}{n\tau}\frac{1}{(s_{u,z_2}s_{v,z_2})^2}\sum_{a,b=1}^{K} S((u,v,z_2),(a,b,z))\lambda_{ab}(1-\lambda_{ab}) \\
&\quad + \left(\frac{1}{s_{u,z_2}s_{v,z_2}}\sum_{a,b=1}^{K} S((u,v,z_2),(a,b,z))\lambda_{ab}\right)^2.
\end{aligned}
\tag{7.14}
$$

Hence,

$$
bE^*(A(b)) = b(A_1(b) + A_2(b))
$$

where

$$
\begin{aligned}
A_1(b) &= -\sum_{u,v=1}^{K}\frac{1}{nb}\frac{1}{s_{u,z_1}s_{v,z_1}}\sum_{a,b=1}^{K} S((u,v,z_1),(a,b,z))\lambda_{ab}(1-\lambda_{ab}) \\
&\quad + \sum_{u,v=1}^{K}\frac{1}{n\tau}\frac{1}{s_{u,z_2}s_{v,z_2}}\sum_{a,b=1}^{K} S((u,v,z_2),(a,b,z))\lambda_{ab}(1-\lambda_{ab}), \\
A_2(b) &= -\sum_{u,v=1}^{K}\frac{1}{s_{u,z_1}s_{v,z_1}}\left(\sum_{a,b=1}^{K} S((u,v,z_1),(a,b,z))\lambda_{ab}\right)^2 \\
&\quad + \sum_{u,v=1}^{K}\frac{1}{s_{u,z_2}s_{v,z_2}}\left(\sum_{a,b=1}^{K} S((u,v,z_2),(a,b,z))\lambda_{ab}\right)^2.
\end{aligned}
\tag{7.15}
$$

Note that

$$
\begin{aligned}
A_1(b) &\geq -\sum_{u,v=1}^{K}\left(\frac{1}{nb}-\frac{1}{n\tau}\right)\frac{1}{s_{u,z_1}s_{v,z_1}}\sum_{a,b=1}^{K} S((u,v,z_1),(a,b,z))\lambda_{ab}(1-\lambda_{ab}) \qquad (7.16)\\
&\quad - \sum_{u,v=1}^{K}\frac{1}{n\tau}\frac{1}{s_{u,z_1}s_{v,z_1}}\sum_{a,b=1}^{K} S((u,v,z_1),(a,b,z))\lambda_{ab}(1-\lambda_{ab}) \\
&\quad + \sum_{u,v=1}^{K}\frac{1}{n\tau}\frac{1}{s_{u,z_2}s_{v,z_2}}\sum_{a,b=1}^{K} S((u,v,z_2),(a,b,z))\lambda_{ab}(1-\lambda_{ab}) \\
&\geq -C\left(\frac{1}{nb}-\frac{1}{n\tau}\right)K^2\rho_{m,n}
\end{aligned}
$$

$$
\begin{aligned}
& -C\sum_{u,v=1}^{K}\frac{1}{n}\frac{1}{s_{u,z_1}s_{v,z_1}}\sum_{(a,b)\neq(u,v)}S((u,v,z_1),(a,b,z)) \\
& -C\sum_{u,v=1}^{K}\frac{1}{n}\frac{1}{s_{u,z_2}s_{v,z_2}}\sum_{(a,b)\neq(u,v)}S((u,v,z_2),(a,b,z)) \\
& -C\frac{1}{n}\sum_{u,v=1}^{K}(S((u,v,z_1),(u,v,z)))|(s_{u,z_1}s_{v,z_1})^{-1}-(s_{u,z}s_{v,z})^{-1}| \\
& -C\frac{1}{n}\sum_{u,v=1}^{K}(S((u,v,z_2),(u,v,z)))|(s_{u,z_2}s_{v,z_2})^{-1}-(s_{u,z}s_{v,z})^{-1}| \\
& -C\frac{1}{n}\sum_{u,v=1}^{K}(s_{u,z}s_{v,z})^{-1}|(S((u,v,z_1),(u,v,z)))-(S((u,v,z_2),(u,v,z))| \\
\geq\ & -C(\tau-b)\frac{K^2}{n}\rho_{m,n}-C(\tau-b)\mathcal{M}_{b,n,m}^2\rho_{m,n}^2. \qquad (7.17)
\end{aligned}
$$

Further,

$$
\begin{aligned}
A_2(b)\ \geq\ & -C\sum_{u,v=1}^{K}\frac{1}{s_{u,z_1}s_{v,z_1}}\left(\sum_{(a,b)\neq(u,v)}S((u,v,z_1),(a,b,z))\right)^2\rho_{m,n}^2 \\
& -C\sum_{u,v=1}^{K}\frac{1}{s_{u,z_1}s_{v,z_1}}\left(S((u,v,z_1),(u,v,z))\right)^2\rho_{m,n}^2 \\
& -C\sum_{u,v=1}^{K}\frac{1}{s_{u,z_2}s_{v,z_2}}\left(\sum_{(a,b)\neq(u,v)}S((u,v,z_2),(a,b,z))\right)^2\rho_{m,n}^2 \\
& -C\sum_{u,v=1}^{K}\frac{1}{s_{u,z_2}s_{v,z_2}}\left(S((u,v,z_2),(u,v,z))\right)^2\rho_{m,n}^2 \\
\geq\ & -C(\tau-b)n^2\mathcal{M}_{b,n,m}^2\rho_{m,n}^2 \\
& -C\sum_{u,v=1}^{K}(S((u,v,z_1),(u,v,z)))^2|(s_{u,z_1}s_{v,z_1})^{-1}-(s_{u,z}s_{v,z})^{-1}|\rho_{m,n}^2 \\
& -C\sum_{u,v=1}^{K}(S((u,v,z_2),(u,v,z)))^2|(s_{u,z_2}s_{v,z_2})^{-1}-(s_{u,z}s_{v,z})^{-1}|\rho_{m,n}^2 \\
& -C\sum_{u,v=1}^{K}(s_{u,z}s_{v,z})^{-1}|(S((u,v,z_1),(u,v,z)))^2-(S((u,v,z_2),(u,v,z))^2|\rho_{m,n}^2 \\
\geq\ & -C(\tau-b)n^2\mathcal{M}_{b,n,m}^2\rho_{m,n}^2. \qquad (7.18)
\end{aligned}
$$

This proves

$$
E^*(A(b))\geq -C(\tau-b)(\frac{K^2}{n}+n^2\mathcal{M}_{b,n,m}^2\rho_{m,n}^2). \qquad (7.19)
$$

Next, consider $B(b)$. Define

$$\mu_{uv,1} = \frac{1}{n(\tau-b)} \sum_{t=nb+1}^{n\tau} \frac{1}{s_{u,z_2}s_{v,z_2}} \sum_{\substack{i:z_2(i)=u \\ j:z_2(j)=v}} A_{ijt},$$

$$\mu_{uv,2} = \frac{1}{n(\tau-b)} \sum_{t=nb+1}^{n\tau} \frac{1}{s_{u,w_1}s_{v,w_1}} \sum_{\substack{i:w_1(i)=u \\ j:w_1(j)=v}} A_{ijt}.$$

Note that

$$\begin{aligned}
B(b) &= \frac{1}{n} \sum_{t=nb+1}^{n\tau} \sum_{i,j=1}^{m} \left[(A_{ijt} - \delta_{w_1(i)w_1(j),1})^2 - (A_{ijt} - \lambda_{z_2(i)z_2(j),2})^2\right] \\
&= \frac{1}{n} \sum_{t=nb+1}^{n\tau} \sum_{i,j=1}^{m} \Big[(\delta_{w_1(i)w_1(j),1})^2 - (\lambda_{z_2(i)z_2(j),2})^2 - 2A_{ijt}(\delta_{w_1(i)w_1(j),1}) \\
&\qquad + 2A_{ijt}(\lambda_{z_2(i)z_2(j),2}))\Big] \\
&= (\tau-b) \sum_{u,v=1}^{K} (s_{u,w_1}s_{v,w_1}(\delta_{uv,1})^2 - s_{u,z_2}s_{v,z_2}(\lambda_{uv,2})^2 - 2\mu_{uv,2}\delta_{uv,1} + 2\mu_{uv,1}\lambda_{uv,2}).
\end{aligned}$$

Therefore,

$$E^*(B(b)) = B_1(b) + B_2(b) \tag{7.20}$$

where

$$\begin{aligned}
B_1(b) &= (\tau-b) \sum_{u,v=1}^{K} \Big[s_{u,w_1}s_{v,w_1}V^*(\delta_{uv,1}) - s_{u,z_2}s_{v,z_2}V^*(\lambda_{uv,2}) \\
&\qquad - 2s_{u,w_1}s_{v,w_1}\mathrm{Cov}^*(\mu_{uv,2}, \delta_{uv,1}) + 2s_{u,z_2}s_{v,z_2}\mathrm{Cov}^*(\mu_{uv,1}, \lambda_{uv,2})\Big], \\
B_2(b) &= (\tau-b) \sum_{u,v=1}^{K} \Big[s_{u,w_1}s_{v,w_1}(E^*(\delta_{uv,1}))^2 - s_{u,z_2}s_{v,z_2}(E^*(\lambda_{uv,2}))^2 \\
&\qquad - 2s_{u,w_1}s_{v,w_1}E^*(\mu_{uv,2})E^*(\delta_{uv,1}) + 2s_{u,z_2}s_{v,z_2}E^*(\mu_{uv,1})E^*(\lambda_{uv,2})\Big] \\
&= (\tau-b)(B_{21} + B_{22} + B_{23} + B_{24}). \qquad (7.21)
\end{aligned}$$

Now, $V^*(\lambda_{uv,2})$ is given in (7.14) and

$$\begin{aligned}
V^*(\delta_{uv,1}) &= \frac{1}{(n(1-b))^2}\Bigg[\frac{n(\tau-b)}{s_{u,w_1}s_{v,w_1}} \sum_{a,b=1}^{K} S((u,v,w_1),(a,b,z))\lambda_{ab}(1-\lambda_{ab}) \\
&\qquad + \frac{n(1-\tau)}{s_{u,w_1}s_{v,w_1}} \sum_{a,b=1}^{K} S((u,v,w_1),(a,b,w))\delta_{ab}(1-\delta_{ab})\Bigg],
\end{aligned}$$

$$\text{Cov}^*(\mu_{uv,2},\delta_{uv,1}) = \frac{1}{n^2(\tau-b)(1-b)}\left[\frac{n(\tau-b)}{s_{u,w_1}s_{v,w_1}}\sum_{a,b=1}^{K}S((u,v,w_1),(a,b,z))\lambda_{ab}(1-\lambda_{ab})\right],$$
$$\text{Cov}^*(\mu_{uv,1},\lambda_{uv,2}) = \frac{1}{n^2(\tau-b)\tau}\left[\frac{n(\tau-b)}{s_{u,z_2}s_{v,z_2}}\sum_{a,b=1}^{K}S((u,v,z_2),(a,b,z))\lambda_{ab}(1-\lambda_{ab})\right].$$

Using similar calculations as in (7.16), we obtain

$$B_1(b) \geq -C(\tau-b)\left(\frac{K^2}{n}\rho_{m,n} + \mathcal{M}_{b,n,m}^2\rho_{m,n}^2\right). \tag{7.22}$$

Next, consider $B_2(b)$.

$$\begin{aligned}
B_{21} &= \sum_{u,v=1}^{K}(s_{u,w_1}s_{v,w_1}-s_{u,w}s_{v,w})(E^*(\delta_{uv,1}))^2 \\
&\quad + \sum_{u,v=1}^{K}s_{u,w}s_{v,w}((E^*(\delta_{uv,1}))^2-(E^*(\delta_{uv,w}))^2) + \sum_{u,v=1}^{K}s_{u,w}s_{v,w}(E^*(\delta_{uv,w}))^2 \\
&\geq -C\mathcal{M}_{b,n,m}^2 - C\sum_{u,v=1}^{K}s_{u,w}s_{v,w}|(E^*(\delta_{uv,1}))-(E^*(\delta_{uv,w}))| + \sum_{u,v=1}^{K}s_{u,w}s_{v,w}\delta_{uv}^2 \\
&\geq -C\mathcal{M}_{b,n,m}^2\rho_{m,n}^2 - B_{211} + \sum_{u,v=1}^{K}s_{u,w}s_{v,w}\delta_{uv}^2. \qquad (7.23)
\end{aligned}$$

We then get

$$\begin{aligned}
B_{211} = &\sum_{u,v=1}^{K}s_{u,w}s_{v,w}|(E^*(\delta_{uv,1}))-(E^*(\delta_{uv,w}))| \\
\leq &\sum_{u,v=1}^{K}s_{u,w}s_{v,w}\Bigg|\frac{1}{n(1-b)}\Bigg[n(\tau-b)\frac{1}{s_{u,w_1}s_{v,w_1}}\sum_{a,b=1}^{K}S((u,v,w_1),(a,b,z))\lambda_{ab} \\
&+n(1-\tau)\frac{1}{s_{u,w_1}s_{v,w_1}}\sum_{a,b=1}^{K}S((u,v,w_1),(a,b,w))\delta_{ab}\Bigg]-\delta_{uv}\Bigg| \\
\leq\ & C\frac{\tau-b}{1-b}\left|\sum_{u,v=1}^{K}\left[\frac{s_{u,w}s_{v,w}}{s_{u,w_1}s_{v,w_1}}\sum_{a,b=1}^{K}S((u,v,w_1),(a,b,z))\lambda_{ab}-s_{u,w}s_{v,w}\delta_{uv}\right]\right| \\
&+C\frac{1-\tau}{1-b}\left|\sum_{u,v=1}^{K}\left[\frac{s_{u,w}s_{v,w}}{s_{u,w_1}s_{v,w_1}}\sum_{a,b=1}^{K}S((u,v,w_1),(a,b,w))\delta_{ab}-s_{u,w}s_{v,w}\delta_{uv}\right]\right| \\
=\ & B_{211a}+B_{211b}, \quad \text{say.} \qquad (7.24)
\end{aligned}$$

It is easy to see that

$$B_{211a} \leq C\rho_{m,n}^2\sum_{u,v=1}^{K}\left|\frac{s_{u,w}s_{v,w}}{s_{u,w_1}s_{v,w_1}}-1\right|s_{u,w_1}s_{v,w_1}$$

$$
\begin{aligned}
& +C\rho_{m,n}^2 \sum_{u,v=1}^{K}\sum_{a,b=1}^{K} |S((u,v,w_1),(a,b,z)) - S((u,v,w),(a,b,z))| \\
\leq & \quad C\mathcal{M}_{b,n,m}^2 \rho_{m,n}^2.
\end{aligned}
$$

Similarly, $B_{211b} \leq C\mathcal{M}_{b,n,m}^2\rho_{m,n}^2$. Thus, by (7.23) and (7.24), we get

$$
B_{21} \geq -C\mathcal{M}_{b,n,m}^2\rho_{m,n}^2 + \sum_{u,v=1}^{K} s_{u,w}s_{v,w}\delta_{uv}^2.
$$

Using similar arguments as above, we also have

$$
\begin{aligned}
B_{22} &\geq -C\mathcal{M}_{b,n,m}^2\rho_{m,n}^2 - \sum_{u,v=1}^{K} s_{u,z}s_{v,z}\lambda_{uv}^2, \\
B_{23} &\geq -C\mathcal{M}_{b,n,m}^2\rho_{m,n}^2 - 2\sum_{u,v=1}^{K} s_{u,w}s_{v,w}\delta_{uv}\sum_{a,b=1}^{K} S((u,v,w),(a,b,z))\lambda_{ab} \\
&\geq -C\mathcal{M}_{b,n,m}^2\rho_{m,n}^2 - 2\sum_{i,j=1}^{m} \lambda_{z(i)z(j)}\delta_{w(i)w(j)}, \\
B_{24} &\geq -C\mathcal{M}_{b,n,m}^2\rho_{m,n}^2 + 2\sum_{u,v=1}^{K} s_{u,z}s_{v,z}\lambda_{uv}^2.
\end{aligned}
$$

Hence, by (7.21)

$$
B_2(b) \geq -C(\tau-b)\mathcal{M}_{b,n,m}^2\rho_{m,n}^2 + C(\tau-b)||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2.
$$

Consequently, by (7.20) and (7.22), we have

$$
\begin{aligned}
E^*(B(b)) \;\; \geq \;\; & -C(\tau-b)\frac{K^2}{n}\rho_{m,n} - C(\tau-b)\mathcal{M}_{b,n,m}^2\rho_{m,n}^2 \\
& +C(\tau-b)||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2. \qquad (7.25)
\end{aligned}
$$

Recall $D(b)$ in (7.13). Similar arguments as above also lead us to conclude

$$
E^*(D(b)) \geq -C(\tau-b)\frac{K^2}{n}\rho_{m,n} - C(\tau-b)n^2\mathcal{M}_{b,n,m}^2\rho_{m,n}^2. \qquad (7.26)
$$

Hence by (7.13), (7.19), (7.25) and (7.26), we have

$$
\begin{aligned}
E^*(L_1 - L_2) \;\; \geq \;\; & -C(\tau-b)\frac{K^2}{n}\rho_{m,n} - C(\tau-b)n^2\mathcal{M}_{b,n,m}^2\rho_{m,n}^2 \\
& +C(\tau-b)||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2 \\
\geq \;\; & -C(\tau-b)||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2, \;\; \text{by SNR-DSBM and (A1).} \qquad (7.27)
\end{aligned}
$$

This proves (7.8).

Next, we compute variances. By (7.13),

$$V^*(L_1 - L_2) = V^*(A(b)) + V^*(B(b)) + V^*(D(b)).$$

We only show the computation for $V^*(A(b))$. Other terms can be handled similarly.

$$\begin{aligned} V^*(A(b)) &\leq C \sum_{u,v=1}^{K} s_{u,z_2} s_{v,z_2} V^*(\lambda_{uv,2})^2 + C \sum_{u,v=1}^{K} s_{u,z_1} s_{v,z_1} V^*(\lambda_{uv,1})^2 \\ &\quad + 2C \sum_{u,v=1}^{K} s_{u,z_2} s_{v,z_2} V^*(\lambda_{uv,2}\lambda_{uv,3}). \end{aligned} \tag{7.28}$$

Let $\mathrm{Cum}_r(X)$ denote the r-th order cumulant of X. Then,

$$\begin{aligned} V^*(\lambda_{uv,2})^2 &\leq \frac{C}{n^4 (s_{u,z_2} s_{v,z_2})^4} E\Big[\sum_{t=1}^{n\tau} \sum_{\substack{i:z_2(i)=u \\ j:z_2(j)=v}} (A_{ijt} - EA_{ijt})\Big]^4 \\ &\leq \frac{C}{n^4 (s_{u,z_2} s_{v,z_2})^4} \Big[\sum_{t=1}^{n\tau} \sum_{\substack{i:z_2(i)=u \\ j:z_2(j)=v}} \mathrm{Cum}_4(A_{ijt} - EA_{ijt}) \\ &\quad + \Big(\sum_{t=1}^{n\tau} \sum_{\substack{i:z_2(i)=u \\ j:z_2(j)=v}} \mathrm{Cum}_4(A_{ijt} - EA_{ijt})\Big)\Big(\sum_{t=1}^{n\tau} \sum_{\substack{i:z_2(i)=u \\ j:z_2(j)=v}} \mathrm{Cum}_4(A_{ijt} - EA_{ijt})\Big)\Big] \\ &\leq \frac{C\rho_{m,n}}{n^2 (s_{u,z_2} s_{v,z_2})^2}. \end{aligned}$$

Similarly, $V^*(\lambda_{uv,1}) \leq \frac{C\rho_{m,n}}{n^2(s_{u,z_1}s_{v,z_1})^2}$ and $V^*(\lambda_{uv,2}\lambda_{uv,3}) \leq \frac{C\rho_{m,n}}{n^2(s_{u,z_2}s_{v,z_2})^2}$. Hence,

$$V^*(A(b)) \leq C\rho_{m,n} \frac{K^2}{n^2} \leq C(\tau - b)\rho_{m,n} \frac{K^2}{n}.$$

Using similar arguments as above, we also have

$$V^*(B(b)), V^*(D(b)) \leq C\rho_{m,n} \frac{K^2}{n^2} \leq C(\tau - b)\rho_{m,n} \frac{K^2}{n}.$$

Hence,

$$V^*(L_1 - L_2) \leq C \frac{K^2}{n^2} \rho_{m,n} \leq C(\tau - b)\rho_{m,n} \frac{||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2}{n}. \tag{7.29}$$

This proves (7.12).

Therefore, by Lemma 32 the proof of Theorem 11 is complete. ■

Remark 35 *Following the proof of Theorem 11 (see (7.27)), it is easy to see that Assumption (A9) $n^2\mathcal{M}_{b,n,m}^2 \rho_{m,n}^2 ||Ed_z(\Lambda) - Ed_w(\Delta)||_F^{-2} \to 0 \ \forall b \in (c^*, 1-c^*)$ on the misclassification* *rate due to clustering is required for achieving consistency of* $\tilde{\tilde{\tau}}_{m,n}$. *The rate of* $\mathcal{M}_{b,n,m}$ *varies for different clustering procedures. For Clustering Algorithm I presented in Section 2, the rate of* $\mathcal{M}^2_{b,n,m}$ *is given in Theorem 8 and hence (A9) reduces to (A1). Details are given before stating (A1). Two variants together with Assumption (A9) and their corresponding misclassification error rates have also been presented and discussed in Section 7.5.*

Note that Assumption (A9) is needed when used in conjunction with the every time point algorithm. However, the assumption can be weakened if we only cluster the nodes once before and after the change pont. Note that we assume that the change-point lies in the interval $(c^*, 1-c^*)$, *which implies that we can cluster nodes using all time points before* c *and obtain* z *and similarly cluster nodes using all time points after* $(1-c^*$ *to obtain* w. *Then,* $A_2(b) = 0$ *and as a consequence* $E^*(A(b)) \geq -C(\tau - b)(\frac{K^2}{n}\rho_{m,n} + \mathcal{M}^2_{b,n,m}\rho^2_{m,n})$ *holds which is a sharper lower bound for* $E^*(A(b))$ *than the one provided in (7.19). Analogously, for* w *we get* $E^*(D(b)) \geq -C(\tau - b)(\frac{K^2}{n}\rho_{m,n} + \mathcal{M}^2_{b,n,m}\rho^2_{m,n})$, *This provides* $E^*(L_1 - L_2) \geq -C(\tau - b)\frac{K^2}{n}\rho_{m,n} - C(\tau - b)\mathcal{M}^2_{b,n,m}\rho^2_{m,n} + C(\tau - b)||Ed_z(\Lambda) - Ed_w(\Delta)||^2_F$ *and a weaker version (A9*) is needed* $[\mathcal{M}^2_{b,n,m}\rho^2_{m,n}||Ed_z(\Lambda) - Ed_w(\Delta)||^{-2}_F \to 0 \ \forall\ b \in (c^*, 1-c^*)$ *(instead of (A9))] along with SNR-DSBM to establish (7.8).*

7.4. Proof of Theorem 13

Next, we focus on establishing the convergence rate for $\tilde{\tilde{\Lambda}}$, while analogous arguments are applicable for $\tilde{\tilde{\Delta}}$.

Without loss of generality, assume $\tilde{\tilde{\tau}}_{m,n} > \tau_{m,n}$. For some clustering function f and $b \in (c^*, 1-c^*)$, recall that $\tilde{\lambda}_{uv,f,(b,n),m} = \frac{1}{nb}\sum_{t=1}^{nb}\frac{1}{s_{u,f}s_{v,f}}\sum_{\substack{f(i)=u\\ f(j)=v}} A_{ij,(t,n)}$.

For some $C > 0$, we have

$$
\begin{aligned}
||\mathrm{Ed}_{\tilde{\tilde{z}}}(\tilde{\tilde{\Lambda}}) - \mathrm{Ed}_z(\Lambda)||^2_F &= \sum_{i,j=1}^{m}(\tilde{\lambda}_{\tilde{\tilde{z}}(i)\tilde{\tilde{z}}(j),\tilde{\tilde{z}},(\tilde{\tilde{\tau}}_{m,n},n),m} - \lambda_{z(i)z(j)})^2 \\
&\leq 2\sum_{i,j=1}^{m}(\tilde{\lambda}_{\tilde{\tilde{z}}(i)\tilde{\tilde{z}}(j),\tilde{\tilde{z}},(\tilde{\tilde{\tau}}_{m,n},n),m} - \tilde{\lambda}_{\tilde{\tilde{z}}(i)\tilde{\tilde{z}}(j),\tilde{\tilde{z}},(\tau_{m,n},n),m})^2 \\
&\quad +2\sum_{i,j=1}^{m}(\tilde{\lambda}_{\tilde{\tilde{z}}(i)\tilde{\tilde{z}}(j),\tilde{\tilde{z}},(\tau_{m,n},n),m} - \lambda_{z(i)z(j)})^2 \\
&\leq Cm^2(\hat{\tilde{\tau}}_{m,n} - \tau_{m,n})^2 \\
&\qquad +C\sum_{i,j=1}^{m}(\tilde{\lambda}_{\tilde{\tilde{z}}(i)\tilde{\tilde{z}}(j),\tilde{\tilde{z}},(\tau_{m,n},n),m} - E^*\tilde{\lambda}_{\tilde{\tilde{z}}(i)\tilde{\tilde{z}}(j),\tilde{\tilde{z}},(\tau_{m,n},n),m})^2 \\
&\qquad\qquad C\sum_{i,j=1}^{m}(E^*\tilde{\lambda}_{\tilde{\tilde{z}}(i)\tilde{\tilde{z}}(j),\tilde{\tilde{z}},(\tau_{m,n},n),m} - \lambda_{z(i)z(j)})^2 \\
&= \mathbb{T}_1 + \mathbb{T}_2 + \mathbb{T}_3 \ \ \text{(say)}.
\end{aligned}
$$

Note that by Theorem 11 we have $m^{-2}\mathbb{T}_1 = O_{\mathrm{P}}(I(n > 1)n^{-2}||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||^{-4}_F\rho^2_{m,n})$.

Let $P^*(\cdot) = P(\cdot|\tilde{\tilde{z}}, \tilde{\tilde{w}})$. By the sub-Gaussian property of Bernoulli random variables and since for some positive sequence $\{\tilde{C}_{m,n}\}$, $\hat{\mathcal{S}}_{m,n} \geq \tilde{C}_{m,n}\ \forall\ n$ with probability 1, we get

$$
\begin{aligned}
P^*(m^{-2}\mathbb{T}_2 \geq t) &= P^*\left(\frac{1}{m^2}\sum_{i,j=1}^{m}(\tilde{\lambda}_{\tilde{\tilde{z}}(i)\tilde{\tilde{z}}(j),\tilde{\tilde{z}},(\tau_{m,n},n),m} - E^*\tilde{\lambda}_{\tilde{\tilde{z}}(i)\tilde{\tilde{z}}(j),\tilde{\tilde{z}},(\tau_{m,n},n),m})^2 \geq t\right) \\
&\leq \sum_{i,j=1}^{m} P^*\left(|\tilde{\lambda}_{\tilde{\tilde{z}}(i)\tilde{\tilde{z}}(j),\tilde{\tilde{z}},(\tau_{m,n},n),m} - E^*\tilde{\lambda}_{\tilde{\tilde{z}}(i)\tilde{\tilde{z}}(j),\tilde{\tilde{z}},(\tau_{m,n},n),m}| \geq C\sqrt{t}\right) \\
&\leq m^2 C_1 e^{-C_2 n \hat{\mathcal{S}}^2_{m,n} t} \leq m^2 C_1 e^{-C_2 n \tilde{C}^2_{m,n} t}.
\end{aligned}
$$

Therefore, $P\left(m^{-2}\mathbb{T}_2 \geq t\right) \leq m^2 C_1 e^{-C_2 n \tilde{C}_{m,n} t} \to 0$ for $t = \frac{\log m}{n\tilde{C}^2_{m,n}}\rho_{m,n}$. Hence, $m^{-2}\mathbb{T}_2 = O_{\mathrm{P}}\left(\frac{\log m}{n\tilde{C}^2_{m,n}}\rho_{m,n}\right)$. Finally,

$$
\begin{aligned}
m^{-2}\mathbb{T}_3 &= \frac{1}{m^2}\sum_{i,j=1}^{m}(E^*\tilde{\lambda}_{\tilde{\tilde{z}}(i)\tilde{\tilde{z}}(j),\tilde{\tilde{z}},(\tau_{m,n},n),m} - \lambda_{z(i)z(j)})^2 \\
&= \frac{1}{m^2}\sum_{i,j=1}^{m}\left(\frac{1}{s_{\tilde{\tilde{z}}(i),\tilde{\tilde{z}}(j)} s_{\tilde{\tilde{z}}(j),\tilde{\tilde{z}}}}\sum_{a,b=1}^{K} S((\tilde{\tilde{z}}(i), \tilde{\tilde{z}}(j), \tilde{\tilde{z}}), (a,b,z))(\lambda_{ab} - \lambda_{z(i)z(j)})\right)^2 \\
&\leq C\mathcal{M}^2_{\tilde{\tilde{\tau}}_{m,n},n,m} = O_{\mathrm{P}}\left(\left(\frac{Km}{n\nu^2_{m,n}}\right)^2\right).
\end{aligned}
$$

Thus, combining the convergence rate of $\mathbb{T}_1$, $\mathbb{T}_2$ and $\mathbb{T}_3$ derived above, establishes the convergence rate of $\mathrm{Ed}_z(\tilde{\tilde{\Lambda}})$ when $\tilde{\tilde{\tau}}_{m,n} > \tau_{m,n}$. Similar arguments work for $\tilde{\tilde{\tau}}_{m,n} \leq \tau_{m,n}$. This completes the proof of Theorem 13. ■

7.5. More on Remark 19

A key step in using the "all time point clustering" algorithm involves clustering. A specific clustering procedure proposed in Bhattacharyya and Chatterjee (2017) was used to identify the communities and for locating the change-point in Section 2. Nevertheless, other clustering algorithms proposed in the literature [Pensky and Zhang (2019); Rohe et al. (2011)] could be employed. For any given clustering algorithms, (a) and (b) in Remark 19 hold.

However, for (A9) to hold, the corresponding misclassification rate in the dense regime needs to satisfy $\mathcal{M}_{b,n,m}n||\mathrm{Ed}_z(\Lambda)-\mathrm{Ed}_w(\Delta)||_F^{-1} \to 0$ for consistency of the estimators. Next, we elaborate on these alternative clustering algorithms.

Clustering Algorithm II. Instead of doing a spectral decomposition of the average adjacency matrices B_1 and B_2, the spectral decompositon is applied to their corresponding Laplcian matrices.

An appropriate modification of the Proof of Theorem 2.1 in Rohe et al. (2011) implies

$$\mathcal{M}_{b,n,m} = O_{\mathrm{P}}\left(\frac{P_{m,n}}{\xi^4_{K_{m,n}}}\left(\frac{(\log m)^2}{nm} + m^2|\tau_{m,n} - b|||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2\right)\right), \tag{7.30}$$

where $P_{m,n} = \max\{s_{u,z}, s_{u,w} : u = 1, 2, \ldots, K_{m,n}\}$ is the maximum community size and $\xi_{K_{m,n}}$ is the minimum between the $K_{m,n}$-th smallest eigenvalue of the Laplacians of B_1 and B_2. A proof is given in Section 7.5.1. Therefore, to satisfy $\mathcal{M}_{b,n,m}n||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^{-1} \to 0$ for the above spectral clustering, we need $\frac{n^{1/2}P_{m,n}(\log m)^2}{\xi^4_{K_{m,n}}mK_{m,n}} = O(1)$ and $\frac{P_{m,n}m^3n}{\xi^4_{K_{m,n}}} \to 0$. However, the latter condition seems excessively stringent in practical settings. For example, suppose $\Lambda = (p_1 - q_1)I_{K_{m,n}} + q_1J_{K_{m,n}}$ and $\Delta = (p_2 - q_2)I_{K_{m,n}} + q_2J_{K_{m,n}}$ where $I_{K_{m,n}}$ is the identity matrix of order $K_{m,n}$ and $J_{K_{m,n}}$ is the $K_{m,n} \times K_{m,n}$ matrix whose entries all equal 1. Further, suppose $0 < C < p_1, q_1, p_2, q_2 < 1 - C < 1$ and that the communities are of equal size. Then, $P_n = O(m/K_{m,n})$. Moreover, Rohe et al. (2011) established that $\xi_{K_{m,n}} = O({K_{m,n}}^{-1})$. Hence, $\frac{n^{1/2}P_{m,n}(\log m)^2}{\xi^4_{K_{m,n}}mK_{m,n}} = O(\sqrt{n}K^2_{m,n}(\log m)^2) \to \infty$ and $\frac{P_{m,n}m^3n}{\xi^4_{K_{m,n}}} = O(m^4n{K_{m,n}}^3) \to \infty$. On the other hand, as we have seen in Example 1, (A1) is satisfied for this example.

Clustering Algorithm III. In this case, the following modification of Rohe et al. (2011)'s algorithm for community detection is employed as follows. Define

$$\begin{aligned} D_{i,(t,n)} &= \sum_{j=1}^{m} A_{ij,(t,n)}, \quad D_{(t,n)} = \mathrm{Diag}\{D_{i,(t,n)} : 1 \le i \le n\}, \qquad (7.31) \\ L_{(t,n)} &= D^{-1/2}_{(t,n)}A_{t,n}D^{-1/2}_{(t,n)}, \quad L_{\Lambda,(b,n)} = \frac{1}{nb}\sum_{t=1}^{nb} L_{(t,n)}, \quad L_{\Delta,(b,n)} = \frac{1}{n(1-b)}\sum_{t=nb+1}^{n} L_{(t,n)}. \end{aligned}$$

Note that $I - L_{(t,n)}$ is the Laplacian of $A_{t,n}$. Next, run the spectral clustering algorithm introduced in Rohe et al. (2011) after replacing L respectively by $L_{\Lambda,(b,n)}$ and $L_{\Delta,(b,n)}$ for estimating z, w.

In this case,

$$\begin{aligned} \mathcal{M}_{b,n,m} &= O_{\mathrm{P}}\Bigg(\frac{P_{m,n}}{\xi^4_{K_{m,n}}}\Bigg(\frac{(\log m\sqrt{n})^2}{\sqrt{n}m} + m^2|\tau_{m,n} - b|||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2 \\ &\qquad + |\tau_{m,n} - b|\frac{(\log m)^2}{m}\Bigg)\Bigg), \end{aligned} \tag{7.32}$$

where $P_{m,n}$ and $\xi_{K_{m,n}}$ are as described after (7.30). A proof is given in Section 7.5.2. Therefore, to satisfy $\mathcal{M}_{b,n,m}n||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^{-1} \to 0$ for this variant of the spectral clustering algorithm, we require $\frac{n^{3/2}P_{m,n}(\log m)^2}{\xi^4_{K_{m,n}}mK_{m,n}} = O(1)$, $\frac{nP_{m,n}(\log m\sqrt{n})^2}{\xi^4_{K_{m,n}}mK_{m,n}} = O(1)$ and $\frac{P_{m,n}m^3n}{\xi^4_{K_{m,n}}} \to 0$. However, these are much stronger conditions that the one required for Clustering Algorithm II.

The upshot of the previous discussion is that Clustering Algorithm I requires a milder assumption (A1) on the misclassification rate compared to Clustering Algorithms II and

III. This is the reason that the results established in Sections 2 and 3 leverage the former algorithm.

7.5.1. Justification of (7.30)

Let $\mathcal{L}(A)$ denote the Laplacian of A. Also without loss of generality, assume $b > \tau$. Using similar arguments as in Appendix B, C and D of Rohe et al. (2011), we can easily show that for some $C > 0$ and with probability tending 1,

$$\begin{aligned}
\mathcal{M}_{b,n,m} &\leq C\frac{P_{m,n}}{\xi^4_{K_{m,n}}}||(\mathcal{L}(\frac{1}{n}\sum_{t=1}^{nb}A_{t,n}))^2 - (\mathcal{L}(\mathrm{Ed}_z(\Lambda)))^2||_F^2 \\
&\leq C\frac{P_{m,n}}{\xi^4_{K_{m,n}}}\Big(||(\mathcal{L}(\frac{1}{n}\sum_{t=1}^{n\tau}A_{t,n}))^2 - (\mathcal{L}(\mathrm{Ed}_z(\Lambda)))^2||_F^2 \\
&\quad +||(\mathcal{L}(\frac{1}{n}\sum_{t=n\tau+1}^{nb}A_{t,n}))^2 - (\mathcal{L}(\mathrm{Ed}_z(\Delta)))^2||_F^2 \\
&\quad +|\tau - b|\ ||(\mathcal{L}(\mathrm{Ed}_z(\Delta)))^2 - (\mathcal{L}(\mathrm{Ed}_z(\Lambda)))^2||_F^2\Big) \\
&= C\frac{P_{m,n}}{\xi^4_{K_{m,n}}}(A_1 + A_2 + A_3),\ \text{say.}
\end{aligned} \tag{7.33}$$

Then,using similar arguments as in Lemma $A.1$ of Rohe et al. (2011), we obtain

$$A_1, A_2 = O_{\mathrm{P}}(\frac{(\log m)^2}{mn}).$$

Define,

$$\begin{aligned}
\mathcal{D}_{i,\Lambda} &= \sum_{j=1}^{m}\lambda_{z(i)z(j)}, \quad \mathcal{D}_\Lambda = \mathrm{Diag}\{\mathcal{D}_{i,\Lambda}:\ 1 \leq i \leq m\}, \\
\mathcal{D}_{i,\Delta} &= \sum_{j=1}^{m}\delta_{w(i)w(j)}, \quad \mathcal{D}_\Delta = \mathrm{Diag}\{\mathcal{D}_{i,\Delta}:\ 1 \leq i \leq m\}.
\end{aligned}$$

Then,

$$\begin{aligned}
A_3 &\leq Cm||\mathcal{L}(\mathrm{Ed}_z(\Lambda)) - \mathcal{L}(\mathrm{Ed}_w(\Delta))||_F^2 \\
&\leq Cm||\mathcal{D}_\Lambda^{-1/2}\mathrm{Ed}_z(\Lambda)\mathcal{D}_\Lambda^{-1/2} - \mathcal{D}_\Delta^{-1/2}\mathrm{Ed}_z(\Delta)\mathcal{D}_\Delta^{1/2}||_F^2 \\
&\leq Cm\Big[||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_z(\Delta)||_F^2||\mathcal{D}_\Lambda^{1/2}||_F^4 + 2||\mathcal{D}_\Lambda^{-1/2} - \mathcal{D}_\Delta^{-1/2}||_F^2||\mathrm{Ed}_z(\Delta)||_F^2||\mathcal{D}_\Delta^{-1/2}||_F^2\Big] \\
&\leq Cm(||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_z(\Delta)||_F^2 + \frac{C}{m}||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_z(\Delta)||_F^2 m^2) \\
&\leq Cm^2||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_z(\Delta)||_F^2.
\end{aligned}$$

Hence,

$$\mathcal{M}_{b,n,m} = O_{\mathrm{P}}\left(\frac{P_{m,n}}{\xi^4_{K_{m,n}}}\left(\frac{(\log m)^2}{mn} + |\tau - b|m^2||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_z(\Delta)||_F^2\right)\right).$$

This completes the justification of (7.30).

7.5.2. Justification of (7.32)

Using similar arguments to those presented in Section 7.5.1, with probability tending to 1, we have

$$\begin{aligned}
\mathcal{M}_{b,n,m} &\leq C\frac{P_{m,n}}{\xi^4_{K_{m,n}}}||(L_{\Lambda,(b,n)})^2 - (\mathcal{L}(\mathrm{Ed}_z(\Lambda)))^2||_F^2 \\
&\leq C\frac{P_{m,n}}{\xi^4_{K_{m,n}}}\Bigg[||(L_{\Lambda,(\tau,n)})^2 - (\mathcal{L}(\mathrm{Ed}_z(\Lambda)))^2||_F^2 + \frac{1}{n}\sum_{t=n\tau+1}^{nb}||(\mathcal{L}(A_{t,n}))^2 - (\mathcal{L}(\mathrm{Ed}_w(\Delta)))^2||_F^2 \\
&\qquad + |\tau - b|||(\mathcal{L}(\mathrm{Ed}_z(\Delta)))^2 - (\mathcal{L}(\mathrm{Ed}_z(\Lambda)))^2||_F^2\Bigg] \\
&\leq C\frac{P_{m,n}}{\xi^4_{K_{m,n}}}(A_1 + A_2 + |\tau - b|m^2||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_z(\Delta)||_F^2), \text{ say.}
\end{aligned}$$

Then, by Theorem 2.1 in Rohe et al. (2011), we have

$$A_1 = O_{\mathrm{P}}(\frac{(\log m\sqrt{n})^2}{m\sqrt{n}}) \text{ and } A_2 = O_{\mathrm{P}}(|\tau - b|\frac{(\log m)^2}{m}).$$

Hence,

$$\mathcal{M}_{b,n,m} = O_{\mathrm{P}}\left(\frac{P_{m,n}}{\xi^4_{K_{m,n}}}\left(\frac{(\log m\sqrt{n})^2}{\sqrt{n}m} + m^2|\tau - b|||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2 + |\tau - b|\frac{(\log m)^2}{m}\right)\right).$$

This completes the justification of (7.32).

7.6. Proof of Theorem 20

To prove Theorem 20, note that the proof of Theorem 11 in Section 7.3 goes through once we use $\mathcal{M}_{b,n,m} = 0$, $z_1 = z_2 = z$, $w_1 = w_2 = w$ and $K = m$. In this case, (7.27) and (7.29) implies

$$\begin{aligned}
E^*(L_1 - L_2) &\geq -C(\tau - b)\frac{m^2\rho_{m,n}}{n} + C(\tau - b)||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2, \\
V^*(L_1 - L_2) &\leq C(\tau - b)\rho_{m,n}||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2.
\end{aligned}$$

Therefore, by SNR-ER and Lemma 32, Theorem 20 follows. ∎

7.7. More on Remark 23

In this section, we focus on the dense network regime. A similar discussion is applicable for the sparse regime as well. As pointed out in Remark 23, one may wonder regarding settings where SNR-DSBM holds, but neither (A1) nor SNR-ER do. The following Examples 5 and 6 introduce such settings in the context of changes in the connection probabilities and in the community structures, respectively.

Example 5 ***(Change in connection probabilities)*** *Consider a DSBM where*

$$z = w \qquad \text{and} \qquad \Lambda = \Delta - \frac{1}{\sqrt{n}}. \tag{7.34}$$

In this case $||Ed_z(\Lambda) - Ed_w(\Delta)||_F^2 = \frac{m^2}{n}$. Therefore, SNR-ER does not hold. However, SNR-DSBM holds if $K = o(m)$. Cases (a)-(c) presented below provide settings where (A1) does not hold, but SNR-DSBM does.

(a) Consider the setting in Example 1 with $K = Cm^{0.5-\delta}$ (that is $K = o(m)$) and $n = Cm^{4\delta}$ for some $C > 0$ and $\delta < 1/6$. It can easily be seen that (A1) does not hold, but SNR-DSBM does.

(b) Suppose all assumptions in Example 2 hold, K is finite and $m = Cn^{2\delta}$. In this case, (A1) does not hold, but SNR-DSBM does.

(c) Finally, consider the setup in Example 3 with $K = o(m)$, $m_{\min} = Cm^{\delta}$, $n = m^{\lambda}$ for some $\lambda > 0$, $\delta \in [0, 1]$ and $-\lambda/2 \le 2\delta - 1 < \lambda/2$. The same conclusion on (A1) failing to hold, while SNR-DSBM holding is reached.

Therefore, in each of the (a)-(c) cases, together with (7.34) do not satisfy (A1) and SNR-ER, whereas SNR-DSBM holds.

Example 6 ***(Change in communities)*** *Consider a DSBM where for $0 < p < 1$,*

$$K = 2, \qquad z(i) = \begin{cases} 1 & \text{if } i \text{ is odd} \\ 2 & \text{if } i \text{ is even,} \end{cases} \qquad w(i) = \begin{cases} 1 & \text{if } 1 \le i \le [m/2] \\ 2 & \text{if } [m/2] < i \le m, \end{cases} \tag{7.35}$$
$$\Lambda = \Delta = \begin{pmatrix} p & p - \frac{1}{\sqrt{n}} \\ p - \frac{1}{\sqrt{n}} & p \end{pmatrix}.$$

This gives $||Ed_z(\Lambda) - Ed_w(\Delta)||_F^2 = \frac{m^2}{n}$. Hence, SNR-ER does not hold, but SNR-DSBM does. Also suppose $m = Cn^{\delta}$ for some $C > 0$ and $\delta \in [1, 1.5)$. In this case (A1) is not satisfied.

The methods discussed in Sections 2 and 3 fail to detect the change-point under the above presented settings. Therefore, alternative strategies not based on clustering and hence assumption (A1) need to be investigated.

One possibility for the case of a single change-point being present was discussed in Remark 17.

Example 7 *As the true change-point $\tau_{m,n} \in (c^*, 1-c^*)$, we can use $\tau^*_{m,n}$ to estimate $\tau_{m,n}$ and its consistency follows from SNR-DSBM and (A1*) $\frac{m}{\sqrt{n}\nu^2_{m,n}} = O(1)$ which is much weaker than (A1). As we have seen before, (A1) and SNR-ER do not hold in Examples 6 and 7 whereas SNR-DSBM is satisfied. Based on the discussion in Remark 17, it is easy to see that (A1*) holds for these examples. Therefore, for the settings posited in Examples 6 and 7, $\tau^*_{m,n}$ estimates $\tau_{m,n}$ consistently. Nevertheless, as mentioned in Remark 17, this strategy is not easy to extend to a setting involving multiple change-points.*

Another setting that does not require clustering is presented next and builds on the model discussed in Gao et al. (2015).

Example 8 *Consider a DSBM with $K = 2$ communities. Further, let B_{1z} and B_{1w} be the blocks where node 1 belongs to under z and w, respectively, and let $\Lambda = \begin{pmatrix} a_1 & d_1 \\ d_1 & a_1 \end{pmatrix}$, $\Delta = \begin{pmatrix} a_2 & d_2 \\ d_2 & a_2 \end{pmatrix}$ with $0 < c < a_1, a_2, d_1, d_2 < 1-c < 1$, $a_1 > d_1, a_2 > d_2$, $a_1 - d_1 = a_2 - d_2$ and the true change-point $\tau \in (c^*, 1-c^*)$. Recall $\hat{p}_{ij,(b,n)}$ and $\hat{q}_{ij,(b,n)}$ from (3.2). Let $\gamma_j = \hat{p}_{11,(b,n)} - \hat{p}_{1j,(b,n)}$ and $\delta_j = \hat{q}_{11,(b,n)} - \hat{q}_{1j,(b,n)}$. One can use the following algorithm to detect communities. Chose $B, B^* > 0$ and $\delta \in (0,1)$ such that $\frac{B}{\sqrt{n^\delta}} \leq \frac{c^*}{1-c^*}(a_1 - d_1)$.*

1. *If $\gamma_j \leq \frac{B}{\sqrt{n^\delta}}$ and $\delta_j \leq \frac{B}{\sqrt{n^\delta}}$, then put node j in $B_{1z} \cap B_{1w}$.*
2. *If $\gamma_j \leq \frac{B}{\sqrt{n^\delta}}$ and $\delta_j > \frac{B}{\sqrt{n^\delta}}$, then put node j in $B_{1z} \cap B^c_{1w}$.*
3. *If $\gamma_j > \frac{B}{\sqrt{n^\delta}}$ and $\delta_j \leq \frac{B}{\sqrt{n^\delta}}$, then put node j in $B^c_{1z} \cap B_{1w}$.*
4. *If $\gamma_j > \frac{B}{\sqrt{n^\delta}}$ and $\delta_j > \frac{B}{\sqrt{n^\delta}}$, then we need further investigation.*
   - *(4a) If $\frac{\gamma_j}{\delta_j} \leq 1 - \frac{B^*}{\sqrt{n^\delta}}$, then put node j in $B_{1z} \cap B^c_{1w}$.*
   - *(4b) If $\frac{\gamma_j}{\delta_j} > 1 + \frac{B^*}{\sqrt{n^\delta}}$, then put node j in $B^c_{1z} \cap B_{1w}$.*
   - *(4c) If $\frac{\gamma_j}{\delta_j} \in (1 - \frac{B^*}{\sqrt{n^\delta}}, 1 + \frac{B^*}{\sqrt{n^\delta}})$, then put node j in $B^c_{1z} \cap B^c_{1w}$.*

In this algorithm, it is easy to see that $P(\text{no node is misclassifed}) \to 1$. Therefore, an alternative condition (A9) is satisfied (see details about it in Section 7.7.1) and $\tilde{\tilde{\tau}}_n$ estimates τ_n consistently.

However, the setting in Example 8 is very specific involving two parameters only for each connection probability matrix), which in turn allows one to use statistics based on the degree connectivity of each node and thus avoid using a clustering algorithm. Nevertheless, a generally applicable strategy is currently lacking for the regime where SNR-DSBM holds, but neither SNR-ER or (A1) do. This constitutes an interesting direction of further research.

7.7.1. justification of Example 8

It is easy to see that the following results (a)-(d) hold under the assumptions in Example 8.

(a) $\gamma_j, \delta_j = O_{\mathrm{P}}(\frac{1}{\sqrt{n}})$ when $j \in B_{1z} \cap B_{1w}$.

(b) $\gamma_j - \frac{b-\tau}{b}(a_2 - d_2),\ \ \delta_j - (a_2 - d_2),\ \ \frac{\gamma_j}{\delta_j} - \frac{b-\tau}{b} = O_{\mathrm{P}}(\frac{1}{\sqrt{n}})$ when $j \in B_{1z} \cap B_{1w}^c$.

(c) $\gamma_j - \frac{\tau}{b}(a_1 - d_1),\ \ \delta_j = O_{\mathrm{P}}(\frac{1}{\sqrt{n}})$ when $j \in B_{1z}^c \cap B_{1w}$.

(d) $\gamma_j - (a_1 - d_1),\ \ \delta_j - (a_1 - d_1),\ \ \frac{\gamma_j}{\delta_j} - 1 = O_{\mathrm{P}}(\frac{1}{\sqrt{n}})$ when $j \in B_{1z}^c \cap B_{1w}^c$.

Using the above results, we have

(a) $P(j \text{ is classified in } B_{1z} \cap B_{2w} \mid j \in B_{1z} \cap B_{2w}) \leq P(\gamma_j < \frac{B}{\sqrt{n^\delta}}, \delta_j < \frac{B}{\sqrt{n^\delta}} \mid j \in B_{1z} \cap B_{2w}) \to 1$.

(b) $P(j \text{ is classified in } B_{1z}^c \cap B_{2w} \mid j \in B_{1z}^c \cap B_{2w}) \leq P(\gamma_j > \frac{B}{\sqrt{n^\delta}}, \delta_j < \frac{B}{\sqrt{n^\delta}} \mid j \in B_{1z}^c \cap B_{2w}) + P(\frac{\gamma_j}{\delta_j} > 1 + \frac{B^*}{\sqrt{n^\delta}} \mid j \in B_{1z}^c \cap B_{2w}) \to 1$.

(c) $P(j \text{ is classified in } B_{1z} \cap B_{2w}^c \mid j \in B_{1z} \cap B_{2w}^c) \leq P(\gamma_j < \frac{B}{\sqrt{n^\delta}}, \delta_j > \frac{B}{\sqrt{n^\delta}} \mid j \in B_{1z} \cap B_{2w}^c) + P(\frac{\gamma_j}{\delta_j} < 1 - \frac{B^*}{\sqrt{n^\delta}} \mid j \in B_{1z} \cap B_{2w}^c) \to 1$.

(d) $P(j \text{ is classified in } B_{1z}^c \cap B_{2w}^c \mid j \in B_{1z}^c \cap B_{2w}^c) \leq P(\frac{\gamma_j}{\delta_j} \in (1 - \frac{B^*}{\sqrt{n^\delta}}, 1 + \frac{B^*}{\sqrt{n^\delta}}) \mid j \in B_{1z}^c \cap B_{2w}^c) \to 1$.

These all together implies $P(\text{no node is misclassified}) \to 1$.

7.8. Proof of Theorem 26

Next, we prove Theorem 26 for $\tilde{\tilde{\tau}}_{m,n}$.Note that the proof for $\hat{\tau}_{m,n}$ is much simpler, once we use $z_1 = z_2 = z$, $w_1 = w_2 = w$, $K = m$ and $\mathcal{M}_{b,n,m} = 0$ in the following proof.

Suppose $||\mathrm{Ed}_z(\lambda) - \mathrm{Ed}_w(\Delta)||_F \to \infty$. Then by Theorem 11, it is easy to see that $P(\tilde{\tilde{\tau}}_{m,n} = \tau_{m,n}) \to 1$.

Lemma 33 from van der Vaart and Wellner (1996) proves useful for establishing the asymptotic distribution of the change-point estimate, when $||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F \to c \geq 0$.

Next, suppose $||\mathrm{Ed}_z(\lambda) - \mathrm{Ed}_w(\Delta)||_F \to c \geq 0$. Take $h = n|\tau - b|||\mathrm{Ed}_z(\lambda) - \mathrm{Ed}_w(\Delta)||_F^2$.

Recall the definitions of $A(b)$ and $D(b)$ from (7.13). Using expectations in (7.19) and (7.26), it is easy to see that by SNR-DSBM, (A1) and as $||\mathrm{Ed}_z(\lambda) - \mathrm{Ed}_w(\Delta)||_F \to c \geq 0$, we have

$$E \sup_{h \in \mathcal{C}} |nA(b)|, E \sup_{h \in \mathcal{C}} |nD(b)| \leq C \frac{K^2}{n||\mathrm{Ed}_z(\lambda) - \mathrm{Ed}_w(\Delta)||_F^2} + C \frac{n\mathbb{M}_{b,n,m}^2}{|\mathrm{Ed}_z(\lambda) - \mathrm{Ed}_w(\Delta)||_F^2} \to 0$$

for some compact set $\mathcal{C} \subset \mathbb{R}$.

This establishes that if $||\mathrm{Ed}_z(\lambda) - \mathrm{Ed}_w(\Delta)||_F \to c \geq 0$ and SNR-DSBM and (A1) hold, then

$$\sup_{h \in \mathcal{C}} |nA(b)|, \quad \sup_{h \in \mathcal{C}} |nD(b)| \xrightarrow{\mathrm{P}} 0. \tag{7.36}$$

Next, recall the definition of $B(b)$ from (7.13). Using similar arguments in Section 7.3, it is easy to show that

$$\begin{aligned} B(b) &= \sum_{t=nb+1}^{n\tau} \sum_{i,j=1}^{m} (2A_{ijt} - 2\lambda_{z(i)z(j)})(\lambda_{z(i)z(j)} - \delta_{w(i)w(j)}) \\ &\qquad + n|\tau - b| ||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2 + R(b) \end{aligned}$$

where $R(b) \leq C|\tau - b| n^2 \mathcal{M}_{b,n,m}^2$. Therefore, by (A1)

$$\sup_{h \in \mathcal{C}} |R(b)| \xrightarrow{\mathrm{P}} 0. \tag{7.37}$$

Suppose $||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F \to 0$. Applying the Central Limit Theorem, it is easy to see that

$$\sup_{h \in \mathcal{C}} |B(b) - R(b) + |h| + 2\gamma B_h| \xrightarrow{\mathrm{P}} 0. \tag{7.38}$$

Thus, by (7.36)-(7.38) and Lemma 33,

$$\begin{aligned} n||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F^2 (\tilde{\tilde{\tau}}_{m,n} - \tau_{m,n}) \xrightarrow{\mathcal{D}} \arg\min_{h \in \mathbb{R}} (|h| + 2\gamma B_h) &\stackrel{\mathcal{D}}{=} \arg\max_{h \in \mathbb{R}} (-0.5|h| + \gamma B_h) \\ &\stackrel{\mathcal{D}}{=} \gamma^2 \arg\max_{h \in \mathbb{R}} (-0.5|h| + B_h). \end{aligned}$$

This proves Part(b) of Theorem 26.

Suppose $||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F \to c > 0$. Then,

$$\begin{aligned} B(b) - R(b) &= \sum_{i,j=1}^{m} \sum_{t=nb+1}^{n\tau} \Big[-(A_{ij,(t,n)} - \lambda_{z(i)z(j)})^2 + (A_{ij,(t,n)} - \delta_{w(i)w(j)})^2 \Big] \\ &= \sum_{i,j \in \mathcal{K}_n} \sum_{t=nb+1}^{n\tau} \Big[-(A_{ij,(t,n)} - \lambda_{z(i)z(j)})^2 + (A_{ij,(t,n)} - \delta_{w(i)w(j)})^2 \Big] \\ &\quad + \sum_{i,j \in \mathcal{K}_0} \sum_{t=nb+1}^{n\tau} \Big[-(A_{ij,(t,n)} - \lambda_{z(i)z(j)})^2 + (A_{ij,(t,n)} - \delta_{w(i)w(j)})^2 \Big] \\ &= T_a + T_b \quad \text{(say)}. \end{aligned} \tag{7.39}$$

By (A6) and (A7) and if $||\mathrm{Ed}_z(\Lambda) - \mathrm{Ed}_w(\Delta)||_F \to c > 0$, we obtain

$$\sup_{h \in \mathcal{C}} |T_b - A^*(h)| \xrightarrow{\mathrm{P}} 0 \tag{7.40}$$

where each $h \in \mathbb{Z}$, $A^*(c^2(h+1)) - A^*(c^2h) = \sum_{k \in \mathcal{K}_0} \left[(Z_{ij,h} - a^*_{ij,1})^2 - (Z_{ij,h} - a^*_{ij,2})^2 \right]$ and $\{Z_{ij,h}\}$ are independently distributed with $Z_{ij,h} \stackrel{d}{=} A^*_{ij,1} I(h < 0) + A^*_{ij,2} I(h \geq 0)$ for all $(i,j) \in \mathcal{K}_0$.

Next, $T_a = 2 \sum_{t=nb+1}^{n\tau} \sum_{i,j \in \mathcal{K}_n} (A_{ij,(t,n)} - \lambda_{z(i)z(j)})(\delta_{w(i)w(j)} - \lambda_{z(i)z(j)}) + |h|$. An application of the Central Limit Theorem together with (A4) and (A5) yields

$$\sup_{h \in \mathcal{C}} |T_a - D^*(h) - C^*(h)| \stackrel{\text{P}}{\to} 0. \tag{7.41}$$

where for each $h \in \mathbb{Z}$, $D^*(c^2(h+1)) - D^*(c^2h) = 0.5\text{Sign}(-h)c_1^2$ and $C^*(c^2(h+1)) - C^*(c^2h) = \tilde{\gamma}_{\text{LSE}} W_h, \quad W_h \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$.

Therefore, by (7.36), (7.37), (7.40), (7.41) and Lemma 33, Part (c) of Theorem 26 is established.

This completes the proof of Theorem 26. ■

7.9. Proof of Theorem 28

Suppose $h > 0$. Then,

$$\begin{aligned} \tilde{L}^*(\hat{\tau}_{m,n} + h/n, \hat{z}, \hat{w}, \hat{\hat{\Lambda}}, \hat{\hat{\Delta}}) \quad = \quad & \frac{1}{n} \sum_{i,j=1}^{m} \Bigg[\sum_{t=1}^{n\hat{\tau}_{m,n}+h} (A_{ij,(t,n),\text{DSBM}} - \hat{\hat{\lambda}}_{\hat{z}(i)\hat{z}(j)})^2 \\ & + \sum_{t=n\hat{\tau}_{m,n}+h+1}^{n} (A_{ij,(t,n),\text{DSBM}} - \hat{\hat{\delta}}_{\hat{w}(i)\hat{w}(j)})^2 \Bigg]. \end{aligned}$$

We then have

$$\begin{aligned} & \tilde{L}^*(\hat{\tau}_{m,n} + h/n, \hat{z}, \hat{w}, \hat{\hat{\Lambda}}, \hat{\hat{\Delta}}) - \tilde{L}^*(\hat{\tau}_{m,n}, \hat{z}, \hat{w}, \hat{\hat{\Lambda}}, \hat{\hat{\Delta}}) \\ = \quad & \frac{1}{n} \sum_{i,j=1}^{m} \sum_{t=n\hat{\tau}_{m,n}+1}^{n\hat{\tau}_{m,n}+h} \left[(A_{ij,(t,n),\text{DSBM}} - \hat{\hat{\lambda}}_{\hat{z}(i)\hat{z}(j)})^2 - (A_{ij,(t,n),\text{DSBM}} - \hat{\hat{\delta}}_{\hat{w}(i)\hat{w}(j)})^2 \right] \\ = \quad & \frac{1}{n} \sum_{i,j=1}^{m} \sum_{t=n\hat{\tau}_{m,n}+1}^{n\hat{\tau}_{m,n}+h} \Bigg[(\hat{\hat{\delta}}_{\hat{w}(i)\hat{w}(j)} - \hat{\hat{\lambda}}_{\hat{z}(i)\hat{z}(j)})^2 \\ & \qquad + 2(A_{ij,(t,n),\text{DSBM}} - \hat{\hat{\delta}}_{\hat{w}(i)\hat{w}(j)})(\hat{\hat{\delta}}_{\hat{w}(i)\hat{w}(j)} - \hat{\hat{\lambda}}_{\hat{z}(i)\hat{z}(j)}) \Bigg]. \end{aligned}$$

Let $E^{**}(\cdot) = E(\cdot|\hat{z}, \hat{w})$, $V^{**} = V(\cdot|\hat{z}, \hat{w})$ and $\text{Cov}^{**}(\cdot) = \text{Cov}(\cdot|\hat{z}, \hat{w})$.

Therefore,

$$E^{**}(\tilde{L}^*(\hat{\tau}_n + h/n, \hat{z}, \hat{w}, \hat{\hat{\Lambda}}, \hat{\hat{\Delta}}) - \tilde{L}^*(\hat{\tau}_n, \hat{z}, \hat{w}, \hat{\hat{\Lambda}}, \hat{\hat{\Delta}})) \quad = \quad \frac{h}{n} ||\text{Ed}_{\hat{z}}(\hat{\hat{\Lambda}}) - \text{Ed}_{\hat{w}}(\hat{\hat{\Delta}})||_F^2.$$

Note that all entries of $\hat{\hat{\Lambda}}$ and $\hat{\hat{\Delta}}$ are bounded away from 0 and 1, since $\log m = o(\sqrt{n})$. Therefore,

$$\begin{aligned}
& V^{**}(\tilde{L}^*(\hat{\tau}_{m,n}+h/n, \hat{z}, \hat{w}, \hat{\hat{\Lambda}}, \hat{\hat{\Delta}}) - \tilde{L}^*(\hat{\tau}_{m,n}, \hat{z}, \hat{w}, \hat{\hat{\Lambda}}, \hat{\hat{\Delta}})) \\
= \ & \frac{h}{n^2}\sum_{i,j=1}^{m}(\hat{\hat{\lambda}}_{\hat{z}(i)\hat{z}(j)} - \hat{\hat{\delta}}_{\hat{w}(i)\hat{w}(j)})^2 \hat{\hat{\delta}}_{\hat{w}(i)\hat{w}(j)}(1-\hat{\hat{\delta}}_{\hat{w}(i)\hat{w}(j)}) \\
\leq \ & \frac{h}{n^2}||\mathrm{Ed}_{\hat{z}}(\hat{\hat{\Lambda}}) - \mathrm{Ed}_{\hat{w}}(\hat{\hat{\Delta}})||_F^2.
\end{aligned}$$

Hence, by Lemma 32 and similar arguments to those made at the beginning of Section 7.3, we have

$$||\mathrm{Ed}_{\hat{z}}(\hat{\hat{\Lambda}}) - \mathrm{Ed}_{\hat{w}}(\hat{\hat{\Delta}})||_F^2 h = O_{\mathrm{P}}(1).$$

Then, by Lemma 34(a),

$$||\mathrm{Ed}_{z}(\Lambda) - \mathrm{Ed}_{w}(\Delta)||_F^2 h = O_{\mathrm{P}}(1).$$

This implies Theorem 28(a).

Next, we establish Theorem 28(b). Note that

$$\begin{aligned}
& n(\tilde{L}^*(\hat{\tau}_{m,n}+h||\mathrm{Ed}_{\hat{z}}(\hat{\hat{\Lambda}}) - \mathrm{Ed}_{\hat{w}}(\hat{\hat{\Delta}})||_F^{-2}/n, \hat{z}, \hat{w}, \hat{\hat{\Lambda}}, \hat{\hat{\Delta}}) - \tilde{L}^*(\hat{\tau}_{m,n}, \hat{z}, \hat{w}, \hat{\hat{\Lambda}}, \hat{\hat{\Delta}})) \\
= \ & -|h| - 2\sum_{t=n\hat{\tau}_{m,n}+1}^{n\hat{\tau}_{m,n}+h}\sum_{i,j=1}^{m}(\hat{\hat{\lambda}}_{\hat{z}(i)\hat{z}(j)} - \hat{\hat{\delta}}_{\hat{w}(i)\hat{w}(j)})(A_{ij,(t,n),\mathrm{DSBM}} - \hat{\hat{\delta}}_{\hat{w}(i)\hat{w}(j)}) \\
& \qquad + o_{\mathrm{P}}(1), \qquad (7.42)
\end{aligned}$$

Further, note that given $\{A_{t,n}\}$, $\{\sum_{i,j=1}^{m}(\hat{\hat{\lambda}}_{\hat{z}(i)\hat{z}(j)} - \hat{\hat{\delta}}_{\hat{w}(i)\hat{w}(j)})(A_{ij,(t,n),\mathrm{DSBM}} - \hat{\hat{\delta}}_{\hat{w}(i)\hat{w}(j)})\}$ is a collection of independent random variables. By Lemma 34(b), we have

$$\begin{aligned}
& E\sum_{t=n\hat{\tau}_{m,n}+1}^{n\hat{\tau}_{m,n}+h}\sum_{i,j=1}^{m}(\hat{\hat{\lambda}}_{\hat{z}(i)\hat{z}(j)} - \hat{\hat{\delta}}_{\hat{w}(i)\hat{w}(j)})(A_{ij,(t,n),\mathrm{DSBM}} - \hat{\hat{\delta}}_{\hat{w}(i)\hat{w}(j)}) \\
= \ & E\Big[\sum_{t=n\hat{\tau}_{m,n}+1}^{n\hat{\tau}_{m,n}+h}\sum_{i,j=1}^{m}(\hat{\hat{\lambda}}_{\hat{z}(i)\hat{z}(j)} - \hat{\hat{\delta}}_{\hat{w}(i)\hat{w}(j)})E^{**}(A_{ij,(t,n),\mathrm{DSBM}} - \hat{\hat{\delta}}_{\hat{w}(i)\hat{w}(j)})\Big] = 0,
\end{aligned}$$

$$\begin{aligned}
& \mathrm{V}\Big(\sum_{t=n\hat{\tau}_{m,n}+1}^{n\hat{\tau}_{m,n}+h}\sum_{i,j=1}^{m}(\hat{\hat{\lambda}}_{\hat{z}(i)\hat{z}(j)} - \hat{\hat{\delta}}_{\hat{w}(i)\hat{w}(j)})(A_{ij,(t,n),\mathrm{DSBM}} - \hat{\hat{\delta}}_{\hat{w}(i)\hat{w}(j)})\Big) \\
= \ & hE\Big(||\mathrm{Ed}_{\hat{z}}(\hat{\hat{\Lambda}}) - \mathrm{Ed}_{\hat{w}}(\hat{\hat{\Delta}})||_F^{-2}\sum_{i,j=1}^{m}(\hat{\hat{\Lambda}}_{\hat{z}(i)\hat{z}(i)} - \hat{\hat{\Delta}}_{\hat{w}(i)\hat{w}(i)})^2\hat{\hat{\delta}}_{\hat{w}(i)\hat{w}(i)}(1-\hat{\hat{\delta}}_{\hat{w}(i)\hat{w}(i)})\Big) \to h\gamma^2,
\end{aligned}$$

$$\frac{E\Big[\sum_{t=n\hat{\tau}_{m,n}+1}^{n\hat{\tau}_{m,n}+h}\sum_{i,j=1}^{m}(\hat{\hat{\lambda}}_{\hat{z}(i)\hat{z}(j)}-\hat{\hat{\delta}}_{\hat{w}(i)\hat{w}(j)})(A_{ij,(t,n),\text{DSBM}}-\hat{\hat{\delta}}_{\hat{w}(i)\hat{w}(j)})\Big]^3}{\Big[\mathrm{V}\Big(\sum_{t=n\hat{\tau}_{m,n}+1}^{n\hat{\tau}_{m,n}+h}\sum_{i,j=1}^{m}(\hat{\hat{\lambda}}_{\hat{z}(i)\hat{z}(j)}-\hat{\hat{\delta}}_{\hat{w}(i)\hat{w}(j)})(A_{ij,(t,n),\text{DSBM}}-\hat{\hat{\delta}}_{\hat{w}(i)\hat{w}(j)})\Big)\Big]^{3/2}}$$

$$\begin{aligned}
&\leq CE\left(\frac{\sum_{i,j=1}^{m}|\hat{\hat{\lambda}}_{\hat{z}(i)\hat{z}(j)}-\hat{\hat{\delta}}_{\hat{w}(i)\hat{w}(j)}|^3}{||\mathrm{Ed}_{\hat{z}}(\hat{\hat{\Lambda}})-\mathrm{Ed}_{\hat{w}}(\hat{\hat{\Delta}})||_F^2}\right)\\
&\leq C(E||\mathrm{Ed}_{\hat{z}}(\hat{\hat{\Lambda}})-\mathrm{Ed}_{\hat{w}}(\hat{\hat{\Delta}})||_F^2)^{1/2}\\
&\leq C\left(E(||\mathrm{Ed}_z(\Lambda)-\mathrm{Ed}_{\hat{z}}\hat{\hat{\Lambda}}||_F^2)+E(||\mathrm{Ed}_w(\Delta)-\mathrm{Ed}_{\hat{w}}(\hat{\hat{\Delta}})||_F^2)+||\mathrm{Ed}_z(\Lambda)-\mathrm{Ed}_w(\Delta)||_F^2\right)\to 0.
\end{aligned}$$

Hence, an application of Lyapunov's Central Limit Theorem together with (A1)-(A4) and SNR**-DSBM-ADAP yields

$$\begin{aligned}
n(\tilde{L}^*(\hat{\tau}_{m,n}+h||\mathrm{Ed}_{\hat{z}}(\hat{\hat{\Lambda}})-\mathrm{Ed}_{\hat{w}}(\hat{\hat{\Delta}})||_F^{-2}/n,\hat{z},\hat{w},\hat{\hat{\Lambda}},\hat{\hat{\Delta}})-\tilde{L}^*(\hat{\tau}_{m,n},\hat{z},\hat{w},\hat{\hat{\Lambda}},\hat{\hat{\Delta}}))&\\
\Rightarrow -|h|+\gamma B_h& \qquad (7.43)
\end{aligned}$$

Similar arguments are applicable for the case of $h<0$.

Finally, (7.43) in conjunction with Lemma 33 establish Theorem 28(b).

An analogous argument to that in the proof of Theorem 26(c) together with similar approximations as in the proof of Theorem 28(b) establish Theorem 28(c) and hence they are omitted. Hence, Theorem 28 is established. ■

7.10. Assumptions for the asymptotic distribution of change-point estimators

Next, we provide precise statements of Assumptions (A3)-(A7) required for establishing the asymptotic distribution of the change point estimators in Theorem 26. A brief comment on these assumptions is given after stating them. We refer to Bhattacharjee et al. (2017) for more in depth explanation.

For Regime II, we define

$$\begin{aligned}
\gamma^2 &= \lim\frac{\sum_{i,j=1}^{m}(\lambda_{z(i)z(j)}-\delta_{w(i)w(j)})^2\lambda_{z(i)z(j)}(1-\lambda_{z(i)z(j)})}{\sum_{i,j=1}^{m}(\lambda_{z(i)z(j)}-\delta_{w(i)w(j)})^2}\\
&= \lim\frac{\sum_{i,j=1}^{m}(\lambda_{z(i)z(j)}-\delta_{w(i)w(j)})^2\delta_{w(i)w(j)}(1-\delta_{w(i)w(j)})}{\sum_{i,j=1}^{m}(\lambda_{z(i)z(j)}-\delta_{w(i)w(j)})^2},
\end{aligned}$$

and assume that

(A3) γ^2 exists.

In Regime II, the asymptotic variance of the change-point estimator is proportional to γ^2. Hence, we require (A3) for its existence and (A2) for the non-degeneracy of the asymptotic distribution.

In Regime III, we consider the following set of edges

$$\mathcal{K}_n = \{(i,j) : 1 \leq i,j \leq m, \ \ |\lambda_{z(i)z(j)} - \delta_{w(i)w(j)}| \to 0\}. \tag{7.44}$$

Define

$$\begin{aligned} c_1^2 &= \lim \sum_{i,j \in \mathcal{K}_n} (\lambda_{z(i)z(j)} - \delta_{w(i)w(j)})^2 \ \text{ and} \\ \tilde{\gamma}^2 &= \lim \sum_{i,j \in \mathcal{K}_n} (\lambda_{z(i)z(j)} - \delta_{w(i)w(j)})^2 \lambda_{z(i)z(j)}(1-\lambda_{z(i)z(j)}) \\ &= \lim \sum_{i,j \in \mathcal{K}_n} (\lambda_{z(i)z(j)} - \delta_{w(i)w(j)})^2 \delta_{w(i)w(j)}(1-\delta_{w(i)w(j)}). \end{aligned} \tag{7.45}$$

Consider the following assumptions.
(A4) c_1 and $\tilde{\gamma}$ exist.
(A5) $\sup_{ij \in \mathcal{K}_n} |\lambda_{z(i)z(j)} - \delta_{w(i)w(j)}| \to 0$.
(A6) $\mathcal{K}_0 = \mathcal{K}_n^c$ does not vary with n.
(A7) For some $\tau^* \in (c^*, 1-c^*)$, $\tau_n \to \tau^*$ as $n \to \infty$. Suppose $\lambda_{z(i)z(j)} \to a^*_{ij,1}$ and $\delta_{w(i)w(j)} \to a^*_{ij,2}$ for all $(i,j) \in \mathcal{K}_0$.

In Regime III, we need to treat edges in $\mathcal{K}_n$ and $\mathcal{K}_0 = \mathcal{K}_n^c$ separately. Note that in Regime II, $\mathcal{K}_n = \{(i,j) : \ 1 \leq i,j \leq m\}$ is the set of all edges. Hence, we can treat $\mathcal{K}_n$ in a similar way as in Regime II and hence we need (A4) in Regime III analogous to (A3) in Regime II. (A5) is a technical assumption and is required for establishing asymptotic normality on $\mathcal{K}_n$. Moreover, $\mathcal{K}_0$ is a finite set. (A6) guarantees that $\mathcal{K}_0$ does not vary with n. Consider the collection of independent Bernoulli random variables $\{A^*_{ij,l} : (i,j) \in \mathcal{K}_0, l = 1,2\}$ with $E(A^*_{ij,l}) = a^*_{ij,l}$. (A7) ensures that $A_{ij,(\lfloor nf \rfloor, n)} \stackrel{\mathcal{D}}{\to} A^*_{ij,1} I(f < \tau^*) + A^*_{ij,2} I(f > \tau^*) \ \forall (i,j) \in \mathcal{K}_0$.

Remark 36 *Note that (A2) is a crucial assumption for establishing the asymptotic distribution of the change-point estimator. It indicates that the resulting random graphi's topology. However, another regime of interest is that where the expected degree of each node grows slower than the total number of nodes in the graph, which gives rise to a* sparse regime. *A number of technical results both from probabilistic and statistical viewpoints have been considered in the recent literature - see, for examples Sarkar and Bickel (2015) and Le et al. (2017). Note that results strongly diverge in their conclusions under these two regimes. For example, Oliveira (2009) showed that the inhomogeneous Erdős-Rényi model satisfies*

$$||L(A) - L(EA)|| = O\left(\sqrt{\frac{\log m}{d_0}}\right)$$

with high probability, where m is the total number of nodes in the graph, $d_0 = \min_i \sum_{j=1}^m EA_{ij}$, A is the observed adjacency matrix, $L(\cdot)$ is the Laplacian and $||\cdot||$ is the operator norm. Therefore, if the expected degrees are growing slower than $\log m$, $L(A)$ will not be concentrated around $L(EA)$. Le et al. (2017) established a different concentration inequality for the case $d_0 = o(\log m)$ after appropriate regularization on the Laplacian and the edge probability matrix. Sarkar and Bickel (2015) also established the convergence rate of the eigenvectors *of the Laplacian for SBM with two communities, which deviates from existing results for dense random graphs. The upshot is that results for the sparse regime are markedly different than those for the dense one.*

It is worth noting that the convergence rate results established in Sections 2 and 3 hold also for the sparse setting; however, establishing the asymptotic distribution of the change-point estimate in a sparse setting, together with issues of adaptive inference will require further work.

References

A. Alexander and L. Horváth. Structural breaks in time series. *Journal of Time Series Analysis*, 34(1):1–16, 2013.

D. Aloise, A. Deshpande, P. Hansen, and P. Popat. NP-hardness of euclidean sum-of-squares clustering. *Machine learning*, 75(2):245–248, 2009.

I. E. Auger and C. E. Lawrence. Algorithms for the optimal identification of segment neighborhoods. *Bulletin of mathematical biology*, 51(1):39–54, 1989.

W. Bao and G. Michailidis. Core community structure recovery and phase transition detection in temporally evolving networks. *Scientific reports*, 8(1):12938, 2018.

M. Bhattacharjee, M. Banerjee, and G. Michailidis. Common change point estimation in panel data from the least-squares and maximum likelihood viewpoints. *arXiv preprint arXiv:1708.05836*, 2017.

S. Bhattacharyya and S. Chatterjee. Spectral clustering for dynamic stochastic block model. 2017.

E. Brodsky and B. S. Darkhovsky. *Nonparametric methods in change point problems*, volume 243. Springer Science & Business Media, 2013.

D. Choi and P. J. Wolfe. Co-clustering separately exchangeable network data. *The Annals of Statistics*, 42(1):29–63, 2014.

H. Crane. *Probabilistic Foundations of Statistical Network Analysis.* Chapman and Hall/CRC, 2018.

M. Csörgö and L. Horváth. *Limit theorems in change-point analysis*, volume 18. John Wiley & Sons Inc, 1997.

D. Durante and D. B. Dunson. Locally adaptive dynamic networks. *The Annals of Applied Statistics*, 10(4):2203–2232, 2016.

D. Durante, D. B. Dunson, and J. T. Vogelstein. Nonparametric bayes modeling of populations of networks. *Journal of the American Statistical Association*, 112(520):1516–1530, 2017.

C. Gao, Y. Lu, and H. H. Zhou. Rate-optimal graphon estimation. *The Annals of Statistics*, 43(6):2624–2652, 2015.

C. Gao, Z. Ma, A. Y. Zhang, and H. H. Zhou. Achieving optimal misclassification proportion in stochastic block models. *The Journal of Machine Learning Research*, 18(1):1980–2024, 2017.

Q. Han, K. Xu, and E. Airoldi. Consistent estimation of dynamic and multi-layer block models. *Proceedings of the 32nd International Conference on Machine Learning (ICML-15)*, pages 1511–1520, 2015.

P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5:109–137, 1983.

J. Jin. Fast community detection by score. *The Annals of Statistics*, 43(1):57–89, 2015.

A. Joseph and B. Yu. Impact of regularization on spectral clustering. *The Annals of Statistics*, 44(4):1765–1791, 2016.

O. Klopp, A. B. Tsybakov, and N. Verzelen. Oracle inequalities for network models and sparse graphon estimation. *The Annals of Statistics*, 45(1):316–354, 2017.

E. D. Kolaczyk and G. Csárdi. *Statistical analysis of network data with R, Use R! book series*, volume 65. Springer, 2014.

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

M. R. Kosorok. *Introduction to empirical processes and semiparametric inference.* Springer, 2008.

A. Kumar, Y. Sabharwal, and S. Sen. A simple linear time $(1+\epsilon)$-approximation algorithm for k-means clustering in any dimensions. *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 454–462, 2004.

C. M. Le, E. Levina, and R. Vershynin. Concentration and regularization of random graphs. *Random Structures & Algorithms*, 51(3):538–561, 2017.

J. Lei and A. Rinaldo. Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43(1):215–237, 2015.

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

S. Minhas, P. D. Hoff, and M. D. Ward. Relax, tensors are here: Dependencies in international processes. *arXiv preprint arXiv:1504.08218*, 2015.

J. Moody and P. J. Mucha. Portrait of political party polarization. *Network Science*, 1(1): 119–121, 2013.

M. Newman, A. L. Barabasi, and D. J. Watts. *The Structure and Dynamics of Networks (Princeton Studies in Complexity)*. Princeton University Press, 2006.

R. I. Oliveira. Concentration of the adjacency matrix and of the laplacian in random graphs with independent edges. *arXiv preprint arXiv:0911.0600*, 2009.

L. Peel and A. Clauset. Detecting change points in the large-scale structure of evolving networks. *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, pages 2914–2920, 2015.

M. Pensky. Dynamic network models and graphon estimation. *The Annals of Statistics*, 47 (4):2378–2403, 2019.

M. Pensky and T. Zhang. Spectral clustering in the dynamic stochastic block model. *Electronic Journal of Statistics*, 13(1):678–709, 2019.

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

P. Sarkar and P. J. Bickel. Role of normalization in spectral clustering for stochastic blockmodels. *The Annals of Statistics*, 43(3):962–990, 2015.

A. W. van der Vaart and J. A. Wellner. *Weak convergence and empirical processes: with applications to statistics*. Springer, 1996.

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

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

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

K. Xu. Stochastic block transition models for dynamic networks. In *Artificial Intelligence and Statistics*, pages 1079–1087, 2015.

T. Yang, Y. Chi, S. Zhu, Y. Gong, and R. Jin. Detecting communities and their evolutions in dynamic social networksa bayesian approach. *Machine learning*, 82(2):157–189, 2011.

A. Y. Zhang and H. H. Zhou. Minimax rates of community detection in stochastic block models. *The Annals of Statistics*, 44(5):2252–2280, 2016.

Y. Zhang, E. Levina, and J. Zhu. Estimating network edge probabilities by neighborhood smoothing. *arXiv preprint arXiv:1509.08588*, 2015.

Y. Zhao, E. Levina, and J. Zhu. Consistency of community detection in networks under degree-corrected stochastic block models. *The Annals of Statistics*, 40(4):2266–2292, 2012.