*Biometrika* (?), **?**, ?, *p*. 1
*Printed in Great Britain*

# Spectral clustering for heterophilic stochastic block models with dynamic node memberships

BY K. Z. LIN
*Department of Statistics & Data Science, University of Pennsylvania*
kevinl1@wharton.upenn.edu

J. LEI
*Department of Statistics & Data Science, Carnegie Mellon University*
jinglei@andrew.cmu.edu

SUMMARY

We consider a time-ordered sequence of networks stemming from stochastic block models where nodes are gradually changing memberships over time and no network contains sufficient signal strength to recover its community structure. To estimate the time-varying community structure, we develop KD-SoS (kernel debiased sum-of-square), a method performing spectral clustering after a debiased sum-of-squared aggregation of adjacency matrices. Our theory demonstrates via a novel bias-variance decomposition that KD-SoS achieves consistent community detection of each network even when the networks are heterophilic, and do not require smoothness in the time-varying dynamics of between-community connectivities. We also prove the identifiability of aligning community structures across time based on how rapidly nodes change communities, and develop a data-adaptive bandwidth tuning procedure for KD-SoS. We demonstrate the utility and advantages of KD-SoS through simulations and a novel analysis on the time-varying dynamics in gene coordination in the human developing brain system.

*Some key words*: Gene co-expression network, human brain development, network analysis, non-parametric analysis, single-cell RNA-seq, time-varying model

## 1. INTRODUCTION

Longitudinal analyses of a network reveals insights on how communities of nodes lose or create new creations over time. Due to the complexity of most networks, statistical methods are a necessity to uncover these broad dynamics. Simply put, suppose we observe a time-ordered sequence of networks among the same $n$ nodes represented as symmetric binary matrices $A^{(0)}, \ldots, A^{(1)} \in \{0,1\}^{n\times n}$, where for time $t \in [0,1]$, the $(i,j)$-entry of $A$ denotes the presence or absence of interaction between two nodes. Due to the non-Euclidean nature of the data, it is often difficult to assess if the larger-scale community structures changed over time, and if so, which specific nodes were changing communities at what rate. Sarkar & Moore (2006) developed one of the first methods to investigate these time-varying dynamics. However, research on the statistical properties of such estimators is recent by comparison (Han et al., 2015). See Kim et al. (2018); Pensky & Zhang (2019) for a comprehensive overview. Our goal in this paper is to provide new theoretical advancements in this area by providing a method that is computationally-efficient and can handle a wide range of network dynamics.

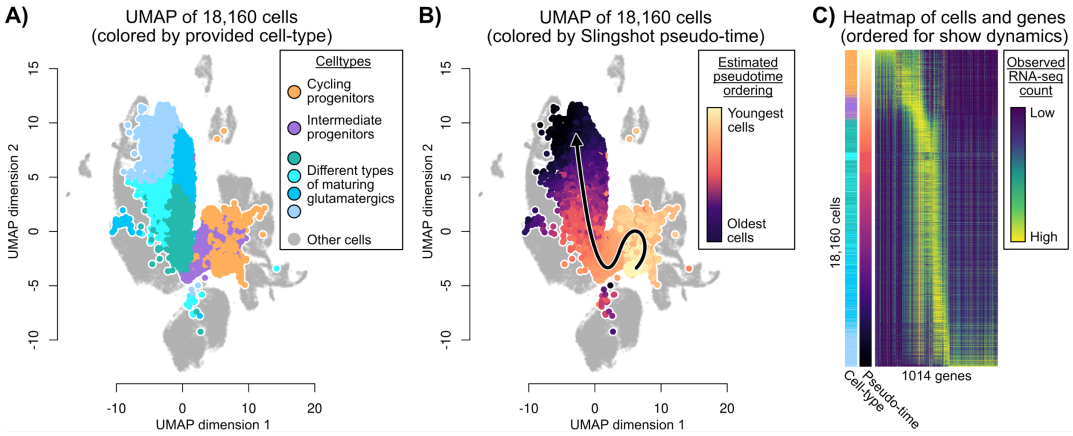


Fig. 1. A) UMAP of the cells among the human developing brain, highlighting the 18,160 cells relevant to our analysis. These denotes cell types such as cycling progenitors (orange) and maturing glutamatergics (shades of teal). B) The 18,160 cells colored based on their estimated pseudotime using Slingshot (Street et al., 2018), colored from youngest (bright yellow) to oldest (dark purple). C) Heatmap ordering the cells based on their estimated pseudotime, and ordering the 1014 relevant genes for this development. The gene expression for each cell is colored based on their expression (high as yellow, low as dark blue).

In this work, we focus on understanding the dynamics of gene coordination over human brain development, but our methods are applicable more broadly to investigate any time-ordered sequence of networks. Consider the single-cell RNA-seq (scRNA-seq) dataset initially published in Trevino et al. (2021), where the authors delineated a specific set of 18,160 cells representing how cycling progenitors (orange) that develop into numerous types of maturing glutamatergics (shades of teal). These cells were annotated by the authors, and the authors also discovered a set 1014 genes that were associated with these cells' development. This data can be visualized through a UMAP (McInnes et al., 2018), a non-linear dimension-reduction method (Figure 1A). Using typical tools in the single-cell analysis toolbox such as Slingshot (Street et al., 2018), we can order the cells in this lineage from the youngest to oldest cells (Figure 1B) and visualize how the gene expression evolves across this lineage (Figure 1C). However, while this simple analysis shows apparent dynamics of the mean gene expression across pseudo-time, the evolution of the gene coordination patterns is unknown. Do the genes that are tightly-coordinated at the beginning of development remain tightly-coordinated at the end of development, and are there tightly-coordinated genes that are not highly expressed?

As reviewed in Kim et al. (2018), there are a plethora of statistical models for time-varying networks. In this work, we focus on time-varying stochastic block models (SBMs). SBMs are a class of prototypical networks that reveal insightful theory while being flexible enough to model many networks in practice. Broadly speaking, a SBM represents each node as part of $K$ (unobserved) communities, and the presence of an edge between two nodes is determined solely by the nodes' community label. Previous work have proven that there is a fundamental limit on how sparse the SBM can be before recovering the communities is impossible (Abbe, 2017). However, when there is a collection of SBMs, this fundamental sparsity can become even smaller. This has led to many different lines of work. For example, one line of work studies the fixed community structure, where $T$ SBMs are observed with all the same community structure (Bhattacharyya & Chatterjee, 2020; Paul & Chen, 2020; Arroyo et al., 2021; Lei & Lin, 2022). Another line of work is when $T$ time-ordered SBMs are observed but there is a changepoint – all the networks before or all the networks after the changepoint share the same community

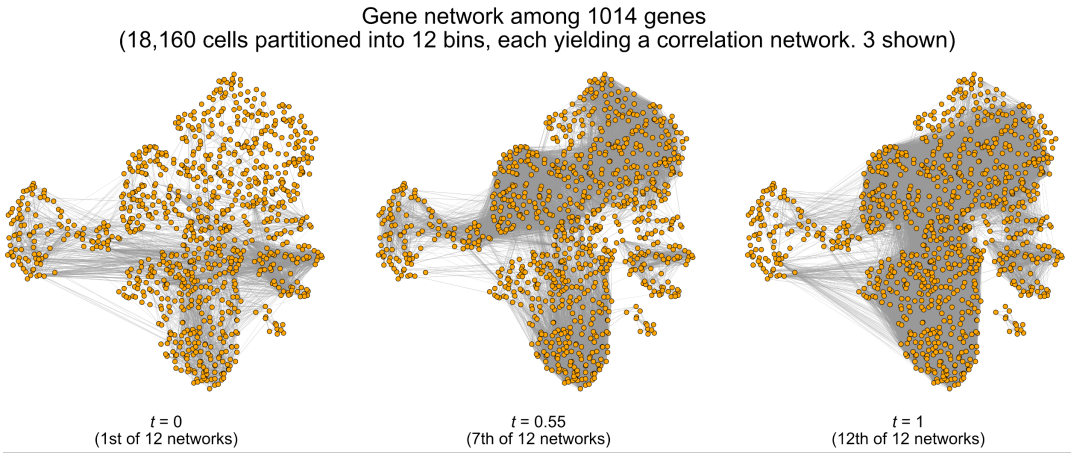


Fig. 2. Three of 12 networks, for $t = 0$ (i.e., gene network among the youngest cells), $t = 0.55$ and $t = 1$ (i.e., gene network among the oldest cells). These are constructed based on thresholding the correlation matrix among the 1014 genes. The visual position of each gene is fixed for each network, but the edges among the gene varies.

structure (Liu et al., 2018; Wang et al., 2021). One last noteworthy line of work is when no temporal structure is imposed across the $T$ networks, but instead each network slightly deviates from a common community structure at random (Chen et al., 2020).

Despite the abundance of aforementioned SBM models equipped with rigorous theory, they are not quite applicable for our intended analysis of the human developing brain. To provide the reader with a scope of the analysis, we plot the correlation network among the 1014 genes for three different time points in Figure 2. These networks were constructed from 12 non-overlapping partitioning of cells across the estimated time, and we observe potentially gradual changes in community structure across time. See Appendix for more details on the preprocessing. Hence, we turn towards time-varying SBM models where the community structure changes slowly over time. To date, Pensky & Zhang (2019) and Keriven & Vaiter (2022) are among the only works that study this setting. This difficulty is induced by the simple observation that changes in community structure are discrete, which prevent typical non-parameteric techniques from being easily applied. However, as we will discuss later, we take a different theoretical approach to analyze this problem and prove consistent estimation of each network's community under broader assumptions. We briefly note that beyond time-varying SBMs, there are works that studying time-varying latent-position graphs (Gallagher et al., 2021; Athreya et al., 2022). Latent-position graphs are more general than SBMs, as they do not impose a community structure. However, these are not the focus of our work, and also not immediately applicable for understanding the gene coordination dynamics in the developing brain.

The main contribution of this paper is a novel and computationally-efficient method equipped with theoretical guarantees regarding community estimation. Specifically, our method is inspired by Lei & Lin (2022), where a debiased sum-of-squared estimator was proven to consistently estimate of communities for fixed-community multi-layer networks, allowing for both homophilic and heterophilic networks. We adapt this to the time-varying setting by introducing a kernel smoother, and prove through a novel bias-variance decomposition that it can consistently estimate the time-varying communities, holding all other assumptions the same. In particular, while the nodes are gradually changing communities, we impose almost no conditions on how the communities are related within a network or how the community-relations change across networks.

We also formalize the information-theoretic relation between the number of networks and the rate nodes change communities as an identifiability condition.

Our second contribution is a tuning procedure for an appropriate kernel bandwidth that also does not impose restrictions on how the community-relations change across networks. Leave-one-out tuning procedures designed in other matrix applications (Yang & Peng, 2020) where network $t$ is predicted using other temporally-surrounding networks are inappropriate since these procedures require community-relations to change smoothly over time. This also precludes Lepskii-based procedures (Pensky & Zhang, 2019). In contrast, our procedure is designed based on the cosine distance between eigenspaces – for network $t \in \mathcal{T}$, the cosine distance is computed between the eigenspaces of kernel-weighted networks for time less than $t$ and of kernel-weighted networks for time greater than $t$ respectively. The bandwidth that minimizes this distance, averaged over all $t \in \mathcal{T}$, is deemed the most appropriate bandwidth. We show through simulation studies and a thorough investigation of the scRNA-seq data that this procedure selects a desirable bandwidth.

## 2. DYNAMIC STOCHASTIC BLOCK MODEL

We define the statistical model and assumptions for the dynamic stochastic block model (SBM) throughout this section, which models the evolution of the same set of nodes through time. Let $n$ denote the number of nodes, and $m^{(0)} \in \{1, \ldots, K\}^n$ denote the initial membership vector, where $K$ is a fixed number of communities. That is, $m_i^{(0)} = k$ for $k \in \{1, \ldots, K\}$ if node $i$ starts in community $k$. We posit that each of the $n$ nodes changes communities according to a $\text{Poisson}(\gamma)$ process with $\gamma > 0$, independent of all other nodes. This means node $i \in \{1, \ldots, n\}$ changes communities at random times $0 < x_{i,1} < x_{i,2} < \ldots < 1$ where the expected difference between consecutive times is $1/\gamma$, and the node changes to one of the $K - 1$ other communities with arbitrary probability. This process generates membership vectors $m^{(t)}$ for $t \in [0, 1]$.

Although each node can potentially change communities at multiple times throughout $t \in [0, 1]$, we assume that only $T$ graphs at fixed time points are observed, for

$$\mathcal{T} = \Big\{0,\ \frac{1}{T-1},\ \frac{2}{T-1},\ \ldots,\ 1\Big\}.$$

The generative model for a specific graph $A^{(t)} \in \{0, 1\}^{n \times n}$ for a time $t \in \mathcal{T}$ is as follows. Let $B^{(t)} \in [0, 1]^{K \times K}$ be a symmetric matrix that denotes the connectivity matrix among the $K$ communities for a fixed positive integer $K$, and let $m^{(t)}$ be the random membership vector based on the above $\text{Poisson}(\gamma)$ process. Each membership vector $m^{(t)}$ can be encoded as one-hot membership matrix $M^{(t)} \in \{0, 1\}^{n \times K}$ where $M_{ik}^{(t)} = 1$ if and only if node $i$ is in community $k$, and 0 otherwise. Then, the probability matrix $Q^{(t)} \in [0, 1]^{n \times n}$ is defined as

$$Q^{(t)} = \rho_n \cdot M^{(t)} B^{(t)} \big(M^{(t)}\big)^\top, \tag{1}$$

for a network density parameter $\rho_n \in (0, 1)$, and $P^{(t)} = Q^{(t)} - \text{diag}(Q^{(t)})$. The observed graph $A^{(t)}$ for time $t \in \mathcal{T}$ is then sampled according to

$$A_{ij}^{(t)} = \begin{cases} \text{Bernoulli}\left(P_{ij}^{(t)}\right), & \text{if } i > j, \\ 0 & \text{if } i = j, \\ A_{ji}^{(t)} & \text{otherwise.} \end{cases} \tag{2}$$

This implies the following relation:

$$\mathbb{E}\big[A^{(t)}\big] = P^{(t)} = Q^{(t)} - \operatorname{diag}\big(Q^{(t)}\big).$$

We define additional to facilitate our theoretical development. Let $n_1^{(t)}, \ldots, n_K^{(t)}$ denote the number of nodes in each community at time $t$, and $n_{\min}^{(t)} = \min\{n_1^{(t)}, \ldots, n_K^{(t)}\}$. Define $L(M^{(s)}, M^{(t)})$ as the relative Hamming distance between the two membership matrices $M^{(s)}$ and $M^{(t)}$,

$$L(M^{(s)}, M^{(t)}) = \min_{R \in \mathbb{Q}_K} \frac{1}{n} \big\| M^{(s)} R - M^{(t)} \big\|_0, \tag{3}$$

where $\mathbb{Q}_K$ is the set of $K \times K$ permutation matrices. For the rest of the paper, we reserve the letters $i, j \in \{1, \ldots, n\}$ for indices of nodes, and $s, t \in [0, 1]$ for indices of time. We reserve $K$ to denote the fixed number of clusters, and the letter $k, \ell \in \{1, \ldots, K\}$ for indices of cluster.

Our theoretic goal is to show the interplay between the number of nodes $n$, the number of observed networks $T$, community-switching rate $\gamma$, the network-sparsity parameter $\rho_n$, and smallest eigenvalue across all $T$ connectivity matrices squared,

$$\lambda_{\min}^2 = \min_{t \in \{1, \ldots, T\}} \lambda_{\min}\big[\big(B^{(t)}\big)^2\big]$$

needed to consistently estimate the $T$ membership matrices across time. Existing theory of single-layer SBMs have already shown that for single network, if $\rho_n \gtrsim \log(n)$, then spectral clustering can asymptotically recover the community structure, while no method can achieve exact recovery if $\rho_n \lesssim \log(n)$ (Bickel & Chen, 2009; Lei & Rinaldo, 2015; Abbe, 2017). Hence, we are interested primarily in the latter setting, where we show that the time-varying community structures can nonetheless be estimated in this challenging setting by aggregating information across networks. Previous methods and theoretical analyses for this setting require strict assumptions on connectivity matrices $\{B^{(t)}\}$ (Pensky & Zhang, 2019; Keriven & Vaiter, 2022) – these matrices are required to smoothly-vary across time and strictly positive eigenvalues, i.e., cannot display patterns of heterophily where edges between communities are more frequent than edges within communities. We seek to develop a method that does not require these assumptions.

## 3. Debiasing and kernel smoothing

### *3·1. Estimator*

Our estimator, the kernelized debiased sum-of-squared (KD-SoS), is motivated by Lei & Lin (2022), where we adopt using de-biased sum of squared adjacency matrices to handle heterophilic networks. We describe our method using the box kernel for simplicity, but the method and theory can be extended to any kernels that are bounded, continuous, symmetric, non-negative and integrate to 1. Provided a bandwidth $r \in [0, 1]$ and number of clusters $K$, our estimator applies the following procedure for any $t \in \mathcal{T}$. First compute the de-biased sum of squared adjacency matrices, where the summation is over all networks within a bandwidth $r$,

$$Z^{(t;r)} = \sum_{s \in \mathcal{S}(t;r)} (A^{(s)})^2 - D^{(s)}, \tag{4}$$

where $\mathcal{S}(t;r) = \mathcal{T} \cap [t-r, t+r]$ and $D^{(t)} \in \mathbb{R}^{n\times n}$ is the (random) diagonal matrix encoding the degrees of the $n$ nodes, i.e.,

$$\left[D^{(t)}\right]_{ii} = \sum_{j=1}^{n} A_{ij}^{(t)}, \quad \text{for all } i \in \{1,\dots,n\}.$$

Second, compute eigen-decomposition of $\widehat{Z}^{(t;r)}$,

$$\widehat{Z}^{(t;r)} = \widehat{U}^{(t;r)}\widehat{\Lambda}^{(t;r)}(\widehat{U}^{(t;r)})^\top,$$

where the diagonal entries of $\widehat{\Lambda}^{(t;r)}$ are in descending order, and lastly apply K-means clustering row-wise on the first $K$ columns of $\widehat{U}^{(t;r)}$. This yields the estimated clustering $\widetilde{m}^{(t)} \in \{1,\dots,K\}^n$. This debiased sum-of-squared estimator is proven in Lei & Lin (2022) to consistently estimate communities under the fixed-community setting, where the squaring of adjacency matrices enable the population connectivity matrices $\{B^{(t)}\}$ to be semidefinite, and the debiasing corrects for the additive noise incurred by this squaring.

After estimating the communities for all $T$ networks, we align the clusters across time. Specifically, initialize $\widehat{m}^{(0)} = \widetilde{m}^{(0)}$, $\delta = 1/(T-1)$ and for $t = 0$, let $C^{(t,t+\delta)}$ be the confusion matrix between $\widehat{m}^{(t)}$ and $\widetilde{m}^{(t+\delta)}$ where

$$C_{k\ell}^{(t,t+\delta)} = \left|\left\{i \in \{1,\dots,n\} : \widehat{m}_i^{(t)} = k \text{ and } \widetilde{m}_i^{(t+\delta)} = \ell\right\}\right|. \tag{5}$$

We then solve the following assignment problem,

$$\widehat{R}^{(t+\delta)} = \max_{R\in\mathbb{Q}_K} \left\| \operatorname{diag}(C^{(t,t+\delta)}R)\right\|_1 \tag{6}$$

This can be formulated as an *Hungarian assignment problem*, which can be solved via linear programming. Then, we let the estimated clustering for time $t$ to be $\widehat{m}^{(t)}$ where $\widehat{m}^{(t)} = k$ if and only if $\widetilde{m}^{(t)} = \ell$ and $\widehat{R}_{\ell k}^{(t+\delta)} = 1$. We repeat this procedure for $t \in \mathcal{T}\backslash\{0\}$, and return the final estimated clusterings $\widehat{m}^{(t)}$ for $t \in \mathcal{T}$.

### *3·2. Bias-variance tradeoff for spectral clustering*

We first describe the bias-variance decomposition foundational to our work. Let $\Delta^{(t)} \in \mathbb{R}^{K\times K}$ denote the diagonal matrix where

$$\operatorname{diag}\left(\Delta^{(t)}\right) = \left\{n_1^{(t)},\dots,n_K^{(t)}\right\}.$$

We then define $\widetilde{M}^{(t)} = M^{(t)}\left(\Delta^{(t)}\right)^{-1/2}$, which is an orthonormal matrix, as well as its projection matrix $\Pi^{(t)} = \widetilde{M}^{(t)}(\widetilde{M}^{(t)})^\top$. Additionally, define the noise matrix $X^{(t)} = P^{(t)} - A^{(t)}$. Observe the following bias-variance decomposition.

LEMMA 1. *Given the model in Section 2, the following deterministic equality holds,*

$$\sum_{s\in\mathcal{S}(t;r)}(A^{(s)})^2 - D^{(s)} = \underbrace{\Big[\sum_{s\in\mathcal{S}(t;r)}(Q^{(s)})^2 - \Pi^{(t)}(Q^{(s)})^2\Pi^{(t)}\Big]}_{I} \tag{7}$$

$$+ \underbrace{\Big[\sum_{s\in\mathcal{S}(t;r)}\big[\operatorname{diag}(Q^{(t)})\big]^2 - Q^{(t)}\operatorname{diag}(Q^{(t)}) - \operatorname{diag}(Q^{(t)})Q^{(t)}\Big]}_{II}$$

$$+ \underbrace{\Big[\sum_{s\in\mathcal{S}(t;r)}X^{(s)}P^{(s)} + P^{(s)}X^{(s)}\Big]}_{III} + \underbrace{\Big[\sum_{s\in\mathcal{S}(t;r)}(X^{(s)})^2 - D^{(s)}\Big]}_{IV}$$

$$+ \underbrace{\Big[\sum_{s\in\mathcal{S}(t;r)}\Pi^{(t)}(Q^{(s)})^2\Pi^{(t)}\Big]}_{V}.$$

We deem this decompose as the *bias-variance decomposition* for dynamic SBMs, since term $I$ represents the deterministic bias dictated by nodes changing communities, term $II$ represents the deterministic diagonal bias, term $III$ represents a random error term centered around 0, term $IV$ represents the random variance term, and term $V$ represents the determistic desired matrix with the community information. We note that this decomposition differs from those used in Pensky & Zhang (2019) and Keriven & Vaiter (2022), which instead yield a decomposition that requires smoothness assumptions in $\{B^{(t)}\}$ in order to derive community-consistency.

### *3·3. Consistency of time-varying communities*

*Assumption* 1 *(Asymptotic regime).* Assume a sequence where $n$ and $T$ are increasing and $n = o(\log(T))$. Additionally, $\rho_n$ and $\gamma$ can vary with $n$ and $T$, but there exists a constant $c_0$ such that $n\rho_n \leq c_0$. Furthermore, assume $K$ is fixed.

We codify the membership dynamics described in Section 2 with the following assumption.

*Assumption* 2 *(Independent Poisson community changing rate).* Assume for a given community-switching rate $\gamma \geq 0$, each node changes memberships at random times between $t \in [0,1]$ according to a $\operatorname{Poisson}(\gamma)$ process, independent of all other nodes.

*Assumption* 3 *(Minimum eigenvalue of aggregated connectivity matrix).* Assume that the sequence $\{B^{(t)}\}$ from $t \in [0,1]$ is an integrable process across each $(i,j) \in \{1,\dots,K\}^2$ coordinate. Additionally, assume that for any $t_1, t_2 \in [0,1]$ for $t_1 \leq t_2$ that

$$\lambda_{\min}\Big(\int_{s=t_1}^{t_2}(B^{(s)})^2 ds\Big) \geq c_1 \cdot (t_2 - t_1)$$

for some constant $c_1 > 0$ independent of $n, T, \gamma, \rho_n$ and $t_1, t_2$.

*Assumption* 4 *(Roughly-equal community sizes).* Assume that across all $t \in [0,1]$ and all clusters $k \in \{1,\dots,K\}$, there exists a constant $c$ independent of $n, T, \lambda, \rho_n$ satisfying $1 \leq c_2$ such that with probability 1 almost surely,

$$n_k^{(t)} \in \Big[\frac{1}{c_2 K}\cdot n,\ \frac{c_2}{K}\cdot n\Big], \quad \text{for all } t \in [0,1].$$

*Assumption* 5 *(Identifiability).* Assume that along the sequence of $\gamma$ and $T$,

$$\gamma/T = o(1). \tag{8}$$

Assumption 1 identifies the regime of interesting, where $n\rho_n \leq c_0$ ensures no network has enough information for consistent estimation of the community structure. (The assumption $n = o(\log(T))$ is for expositional simplicity.) Assumption 3 ensures that the signal can be aggregated across networks, but the specific rate of $c_1(t_2 - t_1)$ can be replaced any suitable rate, as long as the resulting relative Hamming estimation error rates are appropriately adjusted. Assumption 4 is standard for the multilayer network literature, but it has additional significance in our paper since highly imbalanced cluster sizes would result in difficult alignment of communities across time. Notice that this is only an assumption on cluster sizes – we do not enforce the transition matrix of node's communities (for example, such as following a Markov transition matrix). Lastly, as we will show later, Assumption 5 an identifiability assumption. Without it, KD-SoS can still estimate each network's community structure. However, it would be difficult to align the communities across time. Recall that since each node changes memberships independently of one another according to Poisson($\lambda$) process, the expected number of nodes to change memberships within a time interval of $1/(T-1)$ (i.e., the time elapsed between two consecutively observed networks) is roughly $n\gamma/(T-1)$ if $\gamma/(T-1) \lesssim 1$. Hence, a more explicit equivalent statement of (8) is

$$n\gamma/T = o(n/K).$$

This demonstrates the intuition that the networks' communities are identifiable across time if the number of changes between consecutive networks is less than the smallest community size.

Provided these assumptions, KD-SoS's estimated communities have the following pointwise relative Hamming estimation error for the network at time $t \in \mathcal{T}$.

THEOREM 1. *Given Assumptions 1, 2, 3, and 4 for the model in Section 2, for a bandwidth $r \in [0,1]$ satisfying $(rT)^{1/2}n\rho_n \geq C_1 \log^{1/2}(rT+n)$ and $\gamma r < 1/4$ for some constant $C_1 > 1$, then at any particular $t \in \mathcal{T}$,*

$$L\big(G^{(t)}, \widehat{G}^{(t)}\big) \leq C \cdot \frac{1}{(1-\gamma r)^2} \cdot \Big(\gamma r + \frac{1}{n^2} + \frac{\log(rT+n)}{rTn^2\rho_n^2}\Big), \tag{9}$$

*with probability at least $1 - O((rT+n)^{-1})$ for some constant $C > 0$ that depends $c_0$, $c_1$, $c_2$, $C_1$ and $K$.*

We now derive an upper-bound for the relative Hamming error by optimizing the bandwidth $r$. This highlights two regimes, one where the community-switching rate $\gamma$ goes to 0 and the bandwidth $r$ goes to 1 (i.e., the largest possible bandwidth), or where the community-switching rate $\gamma$ is constant or growing and the bandwidth $r$ goes to 0.

COROLLARY 1 (OPTIMAL BANDWIDTH). *Given Assumptions 1, 2, 3, and 4 for the model in Section 2, assume $T^{1/2}n\rho_n = \omega(\log(T+n))$. The bandwidth $r^*$ minimizing the rate in Theorem 1 lies within a range,*

$$r^* \in \Big[\min\Big\{C \cdot \frac{1}{(\gamma T)^{1/2}n\rho_n}, 1\Big\}, \min\Big\{C \cdot \frac{\log^{1/2}(T+n)}{(\gamma T)^{1/2}n\rho_n}, 1\Big\}\Big], \tag{10}$$

*for some constant $C > 0$ that depends $c_0$, $c_1$, $c_2$, $C_1$ and $K$. Here, we interpret $x/0 = \infty$ for $x > 0$.*

Observe that $r^*$ in (10) captures an intuitive behavior. If the number of nodes $n$ or network density $\rho_n$ increases, then there is more signal in each network, reducing the bandwidth $r^*$. If the

community-switching rate $\gamma$ increases, then there is less incentive to aggregate across networks, reducing $r^*$. Observe that box kernel roughly averages $O(r^*T)$ networks, meaning that the number of networks relevant for computing the community structure of network $t$ is approximately $O(\sqrt{T})$ networks. This means the bandwidth grows slower than the total number of networks $T$, which is reasonable. Next, we state the resulting relative Hamming error stemming from the optimal bandwidth $r^*$. This yields two regimes, which can be observed by whether $r^* \to 0$ or $r^* \to 1$.

COROLLARY 2 (FAST COMMUNITY-CHANGING REGIME). *Assume the setting and bandwidth $r^*$ in Corollary 1. With this bandwidth $r^*$, if*

$$C \cdot \frac{\log^{1/2}(T+n)}{T^{1/2}n\rho_n} \leq \gamma \leq C \cdot \frac{T(n\rho_n)^2}{\log^{1/2}(T+n)}, \tag{11}$$

*then $r^* \to 0$ and KD-SoS has a relative Hamming error of*

$$L\Big(G^{(t)}, \widehat{G}^{(t)}\Big) \leq C \cdot \Big(\frac{1}{n^2} + \frac{\gamma^{1/2}\log(T+n)}{T^{1/2}n\rho_n}\Big) \to 0,$$

*with probability $1 - O((r^*T+n)^{-1})$ for any particular $t \in \mathcal{T}$.*

COROLLARY 3 (SLOW COMMUNITY-CHANGING REGIME). *Assume the setting and bandwidth $r^*$ in Corollary 1. With this bandwidth $r^*$, if*

$$\gamma \leq C \cdot \frac{\log^{1/2}(T+n)}{T^{1/2}n\rho_n}, \tag{12}$$

*then $r^* \to 1$ and KD-SoS has a relative Hamming error of*

$$L\Big(G^{(t)}, \widehat{G}^{(t)}\Big) \leq C \cdot \Big(\gamma + \frac{1}{n^2} + \frac{\log(T+n)}{T(n\rho_n)^2}\Big) \to 0,$$

*with probability $1 - O((r^*T+n)^{-1})$ for any particular $t \in \mathcal{T}$.*

Observe that the two conditions (11) and (12) dichotomize the settings in a "fast community-switching regime" and "slow community-switching regime" respectively. In the former setting, the bandwidth converges to 0 due to the nodes changing communities too quickly relative to the other parameters $T$, $n$ and $\rho_n$. In the latter setting, the nodes become less likely to change communities along the asymptotic sequence of $\{T, n, \rho_n\}$, eventually resulting in KD-SoS averaging over all $T$ networks and recovering the rate of the fixed-community setting from Theorem 1 of Lei & Lin (2022).

### *3·4. Identifiability bound for aligning communities across time*

The following proposition demonstrates how Assumption 5 is an identifiability assumption. Consider the Hungarian assignment in (6) that relates the community at time $s$ to time $t$ for $t - s = 1/(T-1)$ (i.e., two consecutively-observed networks). We say that the communities are identifiable if the communities evolving according the Poisson($\lambda$) process yields a Hungarian assignment in (6) of $I_K$ for any two consecutively-observed networks with high probability. Certainly, if the optimal assignment between the (unobserved) communities at time $s$ and $t$ is not the identity, then there is no hope to consistently recover the alignment of communities from random networks.

PROPOSITION 1. *Given Assumption 2 and 4 for the model in Section 2, the addition of Assumption 5 ensures that with probability at least $1 - O(n+T)^{-1}$, simultaneously across all*

*communities, no more than $n/(2c_2K)$ nodes in each community change communities between time $s$ and $t$ for $t - s = 1/(T-1)$. Under this event, the memberships at time $s$ and $t$ are aligned, i.e.,*

$$I_K = \max_{R \in \mathbb{Q}_K} \left\| \operatorname{diag}(C^{(s,t)}R) \right\|_1,$$

*for the confusion matrix $C^{(s,t)}$ defined in* (5) *via the two membership matrices $M^{(s)}$ and $M^{(t)}$.*

## 4. NUMERICAL EXPERIMENTS

In this section, we describe the tuning procedure to choose $r$ in a data-adaptive manner. Our simulations demonstrate that 1) the tuning procedure is reflective of the oracle bandwidth, and 2) KD-SoS and the tuning procedure combined outperforms other estimators for time-varying SBMs.

### *4·1. Tuning procedure*

We design the following procedure to tune the bandwidth $r$ in practice. Observe that typical tuning procedures for time-varying scalar or matrix-valued data often rely on a local smoothness of the observed data across time. For example, this may be predicting the network $A^{(t)}$ using all other networks $\{A^{(s)}\}$ for $s \in \mathcal{S}(t;r)\backslash\{t\}$, but such a procedure would necessarily require additional smoothness assumptions on the connectivity matrices $\{B^{(t)}\}$ on top of our weaker integrability assumption in Assumption 3. Since our estimation theory in Theorem 1 does not require these additional assumptions, we seek to design a tuning procedure that also does not.

Recall that while Theorem 1 does not require smoothness across $\{B^{(t)}\}$, we assume that the community structure is gradually changing via a Poisson($\lambda$) process where $\lambda/T = o(1)$ (Assumption 5). Our theory also demonstrates that changes to the community structure is reflected in the eigenspaces of the probability matrices $\{P^{(t)}\}$. This inspires our method – for a particular time $t \in \mathcal{T}$ and choice of bandwidth $r$, we kernel-average the networks earlier than $t$ (i.e., $\{A^{(s)}\}$ for $s < t$) and compute its leading eigenspace. We then compute the $\sin\theta$ distance (defined below) of this eigenspace from the kernel-average the networks later than $t$ (i.e., $\{A^{(s)}\}$ for $s > t$). A small $\sin\Theta$ distance for an appropriate choice of the bandwidth $\widehat{r}$ would be indicative of two aspects, relative to other choices of $r$: 1) the community structure among the networks in $\mathcal{S}(t;\widehat{r})\backslash[0,t)$ are not too dissimilar to those in networks in $\mathcal{S}(t;\widehat{r})\backslash(t,1]$, and 2) $\widehat{r}$ is large enough to produce stably-estimated eigenspaces among the networks in $\mathcal{S}(t;\widehat{r})\backslash[0,t)$ or $\mathcal{S}(t;\widehat{r})\backslash(t,1]$. Reflecting back to our bias-variance decomposition in (7), the first regards the bias caused by community dynamics, and the second regards the variance due to sparsely-observed networks.

Recall that for two orthonormal matrices $U, V \in [-1,1]^{n\times K}$, the $\sin\Theta$ distance (measured via Frobenius norm) is defined as,

$$\left\| \sin\Theta(U,V) \right\|_F = \sqrt{K - \left\| U^\top V \right\|_F^2}. \tag{13}$$

(See references such as Stewart & Sun (1990) and Cai et al. (2018).) Formally, our procedure is as follows. Suppose a grid possible bandwidths $r_1, \ldots, r_m$ are provided, in addition to the observed networks $\{A^{(t)}\}$.

1. For each bandwidth $r \in \{r_1, \ldots, r_m\}$, compute the score of the bandwidth $\theta(r)$ in the following way.

a. For each time $t \in \mathcal{T}$, compute the leading eigenspaces of $\sum_{s\in\mathcal{S}}(A^{(s)})^2 - D^{(s)}$, where $\mathcal{S}$ is either $\mathcal{S}(t; c\cdot r)\backslash[0,t)$ or $\mathcal{S}(t; c\cdot r)\backslash(t,1]$. Then, compute the $\sin\Theta$ distance between these two eigenspaces via (13), denoted as $\theta(t; r)$.
b. Average $\theta(t; r)$ over $t$. That is, $\theta(r) = \sum_t \theta(t; r)/T$.

2. Choose the optimal bandwidth with the smallest score, i.e., $\widehat{r} = \arg\min_{r\in\{r_1,\ldots,r_m\}} \theta(r)$.

Observe the presence of a small adjustment factor $c > 0$ when deploying the above tuning strategy. This is to account for the fact the size of the sets $\mathcal{S}(t; c\cdot r)\backslash[0,t)$ and $\mathcal{S}(t; c\cdot r)\backslash(t,1]$ are both roughly $c\cdot r$, while the usage of $\widehat{r}$ in KD-SoS would use $\mathcal{S}(t;\widehat{r})$, a set of size roughly $2\widehat{r}+1$. Hence, the adjustment factor $c$ is to scale the bandwidths when tuning to be reflective of its performance when used by KD-SoS. We have found $c = 2$ to be a reasonable choice in practice.

### *4·2. Simulation*

We provide numerical experiments that demonstrate that our estimator described in Section 3·1 is equipped with a tuning procedure which: 1) selects the bandwidth based on data that mimics the oracle that minimizes the Hamming error, and 2) improves upon other methods designed to estimate the community structure for the model (2). We describe our simulation setup. Consider $T = 50$ networks, each consisting of a network among $n = 500$ nodes partitioned into $K = 3$ communities. The first layer set 200 nodes to the first community, 50 nodes to the second community, and 250 nodes to the third community. Then, for each consecutive layer, the nodes switch communities according to the following Markov transition matrix,

$$\begin{bmatrix} 1-\gamma & 0 & \gamma \\ 0 & 1-\gamma & \gamma \\ \frac{4\gamma}{5} & \frac{\gamma}{5} & 1-\gamma \end{bmatrix}. \tag{14}$$

Observe that $100\cdot(1-\gamma)$ percent of the nodes change communities between any two consecutive layers in expectation, and for the given initial community partition, this transition matrix ensures that the community sizes are stationary in expectation. The connectivity matrix is set to alternate between two possible matrices,

$$B^{(t)} = \begin{cases} B^{(\text{odd})} & \text{if } t\cdot(T-1) \bmod 2 = 1, \\ B^{(\text{even})} & \text{otherwise} \end{cases} \quad \text{for } t \in \Big\{0, \frac{1}{T-1}, \frac{2}{T-1}, \ldots, 1\Big\},$$

where

$$B^{(\text{odd})} = \begin{bmatrix} 0.62 & 0.22 & 0.46 \\ 0.22 & 0.62 & 0.46 \\ 0.46 & 0.46 & 0.85 \end{bmatrix}, \quad \text{and} \quad B^{(\text{even})} = \begin{bmatrix} 0.22 & 0.62 & 0.46 \\ 0.62 & 0.22 & 0.46 \\ 0.46 & 0.46 & 0.85 \end{bmatrix}. \tag{15}$$

Then, the observed data is generated according to the model (2), for the desired network-density $\rho$ (varying between sparse networks with $\rho = 0.05$ to dense networks with $\rho = 1$) and the nodes' community-switching transition matrix (14) for a given rate $\gamma$ (varying between stable communities with $\gamma = 0$ to rapidly-changing communities with $\gamma = 0.1$). By considering connectivity matrices $B^{(t)}$ of the form (15), the graphs alternate between being either homophilic and heterophilic.

We first show that our tuning procedure to select the appropriate bandwidth $r$ of the box kernel, as demonstrated in Figure 3. In the left panel, we fix $\rho = 0.3$ and $\gamma = 0.01$ and plot the mean Hamming error across all networks as a function of applying our estimator with the bandwidth $r$ (black line) and the bandwidth alignment used to tune $r$ (blue line), both averaged across 25

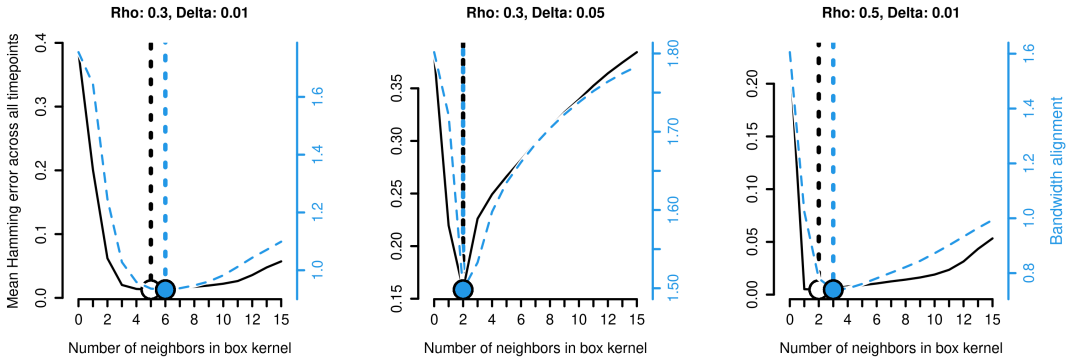


Fig. 3. Simulation across three different settings of the community-switching rate $\gamma$ and network density $\rho_n$ demonstrating KD-SoS's performance for different bandwidth $r$'s. The Hamming error (3) or the bandwidth score measured via $\sin\Theta$ (13) are averaged across 25 trials for each $r$ (black and blue respectively), and the the vertical dotted lines denote the oracle minimizer of the Hamming error (black) and the chosen bandwidth $\widehat{r}$ using the tuning procedure (blue).

trials. The minimum of both curves are marked by a dot of their respective color. We make two observations. First, the Hamming error follows a classical U-shape as a function of $r$. This demonstrates that although a single network does not contain information to accurately estimate the communities (i.e. $r = 0$), pooling information across too many networks is not ideal either since the community structures varies too much among the networks (i.e. $r = 15$). Second, we see that while a bandwidth of $r = 5$ achieves the minimum Hamming error, our tuning procedure would select $r = 6$ on average and its degradation of the Hamming error is not substantial. We also vary $\rho$ and $\gamma$. When we set $\gamma$ to be 0.05 instead of 0.1, we see that minimizing bandwidth becomes smaller, reflecting that less neighboring networks are relevant for estimating a particular network's community structure. Alternatively, when we set $\rho$ to be 0.5 instead of 0.3, we also see that minimizing bandwidth becomes smaller. However, as implied by the mean Hamming error on the y-axis, this is because there is more information contained within each denser network, lessening the need to pool information across networks.

We now compare our method against other methods designed to estimate communities for the model (2). Two natural candidates are our debiasing-and-smoothing method where the bandwidth is set to be $r = 0$ (i.e., "Singleton," where each network's community is estimated using only that network) and $r = 1$ (i.e., "All," where each network's community is estimated by equally weighting all the networks). We measure the performance of each of the three methods by computing the relative Hamming distance between $\widehat{M}^{(t)}$ and $M^{(t)}$, averaged across all time $t \in \mathcal{T}$ (i.e., a smaller metric implies better performance). Our results are shown in Figure 4. In the first simulation suite, we hold network density $\rho_n = 0.5$ but vary the community-switching rate $\delta$ from 0 to 0.1 (i.e., stable communities to rapidly changing communities). Across the 50 trials for each value of $\delta$, we see that KD-SoS (blue) is able to retain a small Hamming error below 0.2 across a wide range of $\delta$. In contrast, observe that "Singleton" (orange) has a relatively stable performance, which is intuitive as this method is not affected by the time-varying structure. Meanwhile, "All" (purple) degrades in performance quite rapidly as $\delta$ increases due to aggregating among all the networks despite large differences in community structure. In the second simulation suite, we hold community-switching rate $\delta = 0.05$, but vary the network density $\rho_n$ from 0.2 to 1 (i.e., sparse networks to dense networks). Across the 50 trials for each value of $\rho_n$, we see that KD-SoS (blue) has better performance as $\rho_n$ increases, and this performance

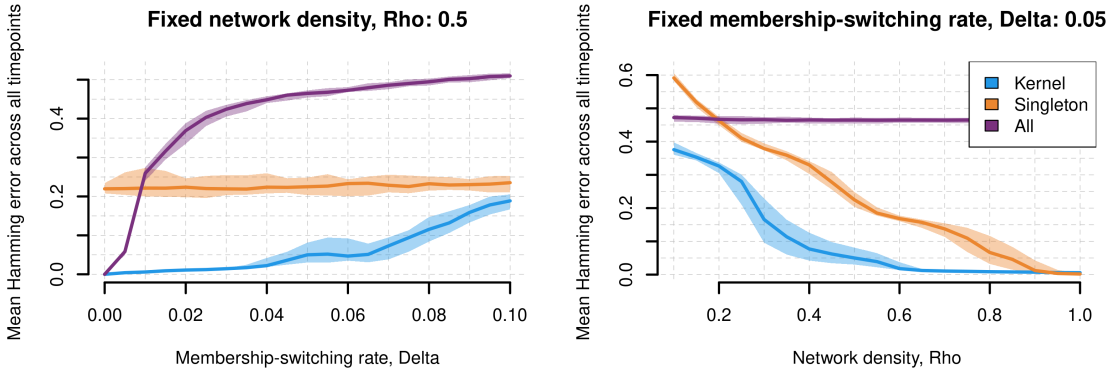


Fig. 4. Simulation suite across three different settings of the community-switching rate $\gamma$ and network density $\rho_n$ demonstrating KD-SoS with the bandwidth tuning procedure's performance ("Kernel," blue) compared to applying spectral clusterings to only one network at a time ("Singleton," orange) or aggregating across all networks, akin to Lei & Lin (2022) ("All," purple).

is uniformly better than the "Singleton" (orange). This is sensible, as KD-SoS aggregates information across networks with an appropriately chosen bandwidth $r$. Meanwhile, "All" (purple) does not change in performance as $\rho_n$ increases because the time-varying community structure obstructs good performance regardless of network sparsity.

### *4·3. Application to gene co-expression networks along developmental trajectories*

We now return the analysis of the developing brain introduced in Section 1. We first enumerate descriptive summary statistics of these twelve networks.

We now describe the results when applying KD-SoS on the dataset. Due to the presence of only twelve networks with very different degrees, we use a Gaussian kernel normalized by each network's leading singular value, i.e.,

$$Z^{(t;r)} = \sum_{s \in \mathcal{T}} \frac{w(s,t;r)}{\|A^{(s)}\|_{\text{op}}} \cdot \left[ (A^{(s)})^2 - D^{(s)} \right], \quad \text{where } w(s,t;r) = \exp\left( \frac{-(t-s)^2}{r^2} \right)$$

instead of the aggregation used in (4). While our theoretical developments in Theorem 1 are not using this estimator, our techniques would apply similarly to such estimators. Based on a scree plot among $\{A^{(t)}\}$, we chose $K = 10$ as the dimensionality and number of communities. The bandwidth is chosen using our procedure in Section 4·1. The clustering results for three of the twelve networks is shown in Figure 5, where nodes of different colors are in different communities. Already, we can see gradual shifts in communities within these three networks. For example, both the purple and red communities grow in size as time progresses. Meanwhile, genes starting in the olive community eventually become part of the pink or white community.

It is hard to discern the broad summary of how communities are related across time from Figure 5. Hence, we plot the percentage of genes that exit from one community to join a different community between the first three networks in Figure 6. We see that our tuning bandwidth procedure chose an $r$ that yielded relatively stable communities across time. Meanwhile, Figure 6 also visualizes the latent 10-dimensional embedding among all 1014 genes for the first three networks. We observe that: 1) the SBM model is appropriate for modeling the dataset at hand since the heatmaps demonstrate strong block structure, and 2) a choice of $K = 10$ seems visually appropriate, as none of the 10 communities seem to represent sub-communities based on the 10 latent dimensions.

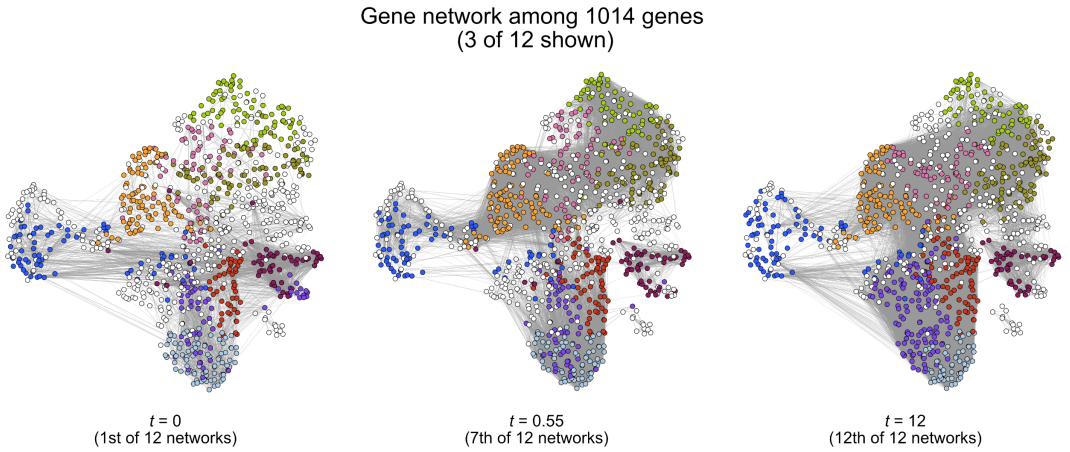


Fig. 5. Three networks, as displayed in Figure 2, but with genes colored by the $K = 10$ different communities via $K$ different colors as estimated by KD-SoS and the bandwidth tuning procedure.

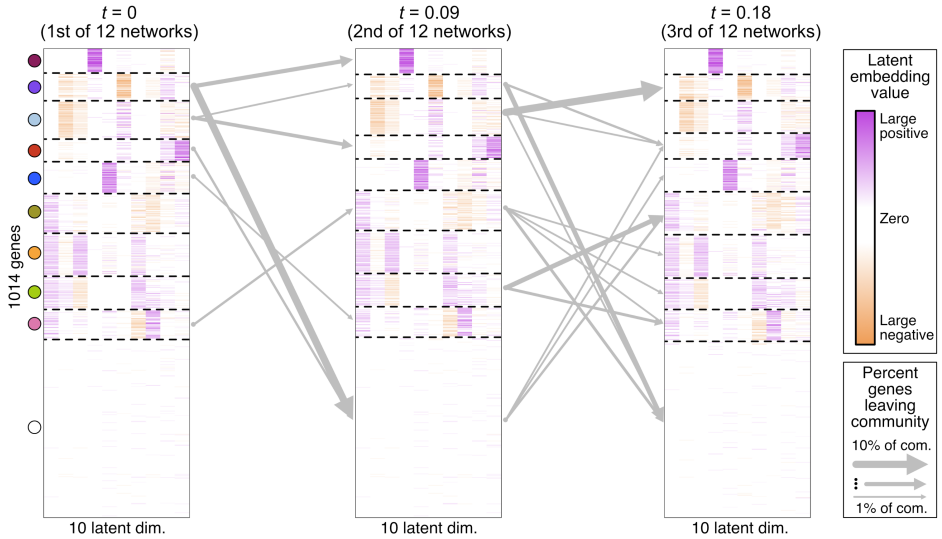


Fig. 6. The heatmap of the first three network's leading $K = 10$ eigenvectors, where the 1014 genes are ordered based on their assigned communities, with their colors (left) corresponding to those in Figure 5. Let $s, t \in [0, 1]$ denote two consecutive two timepoints where $s < t$. The size of the arrow connecting two different communities, one at $s$ and another at $t$ denotes the percentage of genes that leave the community at time $s$ to a different community at time $t$, ranging from 1% of the genes in the community (thin arrow) to 10% (thick arrow).

Now that we have investigated the appropriateness of the time-varying SBM model as well as the KD-SoS estimated community structure, we now address the motivating biological questions asked in Section 1. We focus on specifically the second and twelfth network here, but other investigations are included in the Appendix. Starting with the second network (Figure 7A), we observe that genes in purple communities are highly expressed for cells in this pseudotime bin, while genes in the blue communities are not. However, when looking at the corresponding network, we see that genes in the purple and blue communities are highly coordinated among themselves. Likewise, when investigating the twelfth network (Figure 7B), we observe that certain genes in the purple genes and most genes in the orange and pink communities are highly expressed for the last pseudotime bin. However, we see that the purple, light blue and red com-

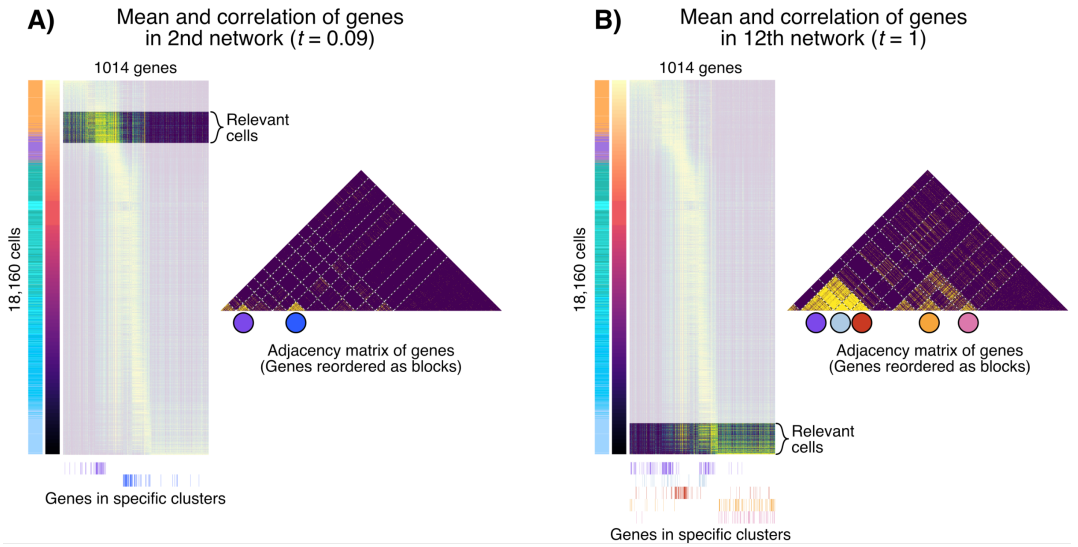


Fig. 7. Correlation networks for the second (A) or twelfth (B) timepoint, where the cells corresponding to the respective bin of pseudotimes are highlighted via the cell-gene heatmap (left) and the corresponding adjacency matrix among 1014 genes where the genes are organized based on their estimated clustering for the respective timepoint (right). The cell-gene heatmaps are the same as in Figure 1. Below the heatmaps marks the genes (i.e., columns) that are part of specifically-highlighted communities, which correspond to the marked entries of the adjacency matrices.

munities form one highly coordinated block of genes at this time point, while the orange and pink genes are less coordinated in comparison. Altogether, these results demonstrates that investigating the dynamics among gene coordination can give an alternative perspective of developing in the brain.

Additional plots corresponding to networks not shown in Figures 5 through 7 as well as additional visualizations of the time-varying dynamics are included in the Appendix. We also include a GO-term analysis to describe the pathways describing different communities for each of the time points there.

## 5. Discussion

We establish a bridge between time-varying network analysis and non-parametric analysis in this paper, demonstrating that smoothness across the connectivity matrices $\{B^{(t)}\}$ is not required for consistent community detection. We achieve this through a novel bias-variance decomposition, whereby we project networks close to time $t$ onto the leading eigenspace of the network at time $t$. While our paper has demonstrated how to relate the discrete changes in nodes' communities to the typically-continuous non-parameteric theory, there are two major theoretical directions we hope our work can aid for future research. The first is refining this relation between time-varying networks and non-parametric analyses. While previous work for time-varying networks such as Pensky & Zhang (2019) and Keriven & Vaiter (2022) derived rates reliant on the smoothness across $\{B^{(t)}\}$, it is unclear from a minimax perspective how the community estimation rates improve as $\{B^{(t)}\}$ evolve according to a smoother process. Additionally, there have been major historical developments in non-parametric analysis through local polynomials and trend filtering that address the so-called boundary-bias typical in nonparameteric regression as well as constructing estimators that inherently adaptive the smoothness in the data. We wonder if there are analogies for these estimators for the time-varying SBM setting. Secondly, as with

any non-parameteric estimator, there are unanswered questions on how to best tune estimators such as KD-SoS. As we described in Section 4·1, tuning procedures reliant on prediction such as cross-validation are unlikely to be fruitful for the setting we study. However, recent ideas using leave-one-out analysis or sharp $\ell_{2\to\infty}$ estimation bounds for the leading eigenspaces have been successfully used to derive cross-validation-like approaches in other network settings. We believe those ideas can be used similarly in our setting where $\{B^{(t)}\}$ is not assumed to be positive definite or smoothly-varying.

## REFERENCES


ABBE, E. (2017). Community detection and stochastic block models: Recent developments. *The Journal of Machine Learning Research* **18**, 6446–6531.

ARROYO, J., ATHREYA, A., CAPE, J., CHEN, G., PRIEBE, C. E. & VOGELSTEIN, J. T. (2021). Inference for multiple heterogeneous networks with a common invariant subspace. *Journal of Machine Learning Research* **22**, 1–49.

ATHREYA, A., LUBBERTS, Z., PARK, Y. & PRIEBE, C. E. (2022). Discovering underlying dynamics in time series of networks. *arXiv preprint arXiv:2205.06877* .

BHATTACHARYYA, S. & CHATTERJEE, S. (2020). General community detection with optimal recovery conditions for multi-relational sparse networks with dependent layers. *arXiv preprint arXiv:2004.03480* .

BICKEL, P. J. & CHEN, A. (2009). A nonparametric view of network models and newman–girvan and other modularities. *Proceedings of the National Academy of Sciences* **106**, 21068–21073.

CAI, T. T., ZHANG, A. et al. (2018). Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *The Annals of Statistics* **46**, 60–89.

CHEN, S., LIU, S. & MA, Z. (2020). Global and individualized community detection in inhomogeneous multilayer networks. *arXiv preprint arXiv:2012.00933* .

GALLAGHER, I., JONES, A. & RUBIN-DELANCHY, P. (2021). Spectral embedding for dynamic networks with stability guarantees. *arXiv preprint arXiv:2106.01282* .

HAN, Q., XU, K. & AIROLDI, E. (2015). Consistent estimation of dynamic and multi-layer block models. In *International Conference on Machine Learning*. PMLR.

KERIVEN, N. & VAITER, S. (2022). Sparse and smooth: Improved guarantees for spectral clustering in the dynamic stochastic block model. *Electronic Journal of Statistics* **16**, 1330–1366.

KIM, B., LEE, K. H., XUE, L. & NIU, X. (2018). A review of dynamic network models with latent variables. *Statistics surveys* **12**, 105.

LEI, J. & LIN, K. Z. (2022). Bias-adjusted spectral clustering in multi-layer stochastic block models. *Journal of the American Statistical Association* , 1–13.

LEI, J. & RINALDO, A. (2015). Consistency of spectral clustering in stochastic block models. *The Annals of Statistics* **43**, 215–237.

LIU, F., CHOI, D., XIE, L. & ROEDER, K. (2018). Global spectral clustering in dynamic networks. *Proceedings of the National Academy of Sciences* **115**, 927–932.

MCINNES, L., HEALY, J. & MELVILLE, J. (2018). UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426* .

PAUL, S. & CHEN, Y. (2020). Spectral and matrix factorization methods for consistent community detection in multi-layer networks. *The Annals of Statistics* **48**, 230–250.

PENSKY, M. & ZHANG, T. (2019). Spectral clustering in the dynamic stochastic block model. *Electronic Journal of Statistics* **13**, 678–709.

SARKAR, P. & MOORE, A. W. (2006). Dynamic social network analysis using latent space models. *Advances in neural information processing systems* **18**, 1145.

STEWART, G. & SUN, J. (1990). Matrix perturbation theory, acad. *Press, Boston MA* .

STREET, K., RISSO, D., FLETCHER, R. B., DAS, D., NGAI, J., YOSEF, N., PURDOM, E. & DUDOIT, S. (2018). Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477.

TREVINO, A. E., MÜLLER, F., ANDERSEN, J., SUNDARAM, L., KATHIRIA, A., SHCHERBINA, A., FARH, K., CHANG, H. Y., PAŞCA, A. M., KUNDAJE, A., PASCA, S. P. & GREENLEAF, W. J. (2021). Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution. *Cell* .

WANG, D., YU, Y. & RINALDO, A. (2018). Optimal change point detection and localization in sparse dynamic networks. *arXiv preprint arXiv:1809.09602* .

WANG, D., YU, Y. & RINALDO, A. (2021). Optimal change point detection and localization in sparse dynamic networks. *The Annals of Statistics* **49**, 203–232.

YANG, J. & PENG, J. (2020). Estimating time-varying graphical models. *Journal of Computational and Graphical Statistics* **29**, 191–202.

YU, Y., WANG, T. & SAMWORTH, R. J. (2014). A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika* **102**, 315–323.