# Efficient Inference of Spatially-varying Gaussian Markov Random Fields with Applications in Gene Regulatory Networks*

Visweswaran Ravikumar[†]    Tong Xu[‡]    Wajd N. Al-Holou [§]    Salar Fattahi[¶]

Arvind Rao[‖]

June 22, 2022

## Abstract

In this paper, we study the problem of inferring spatially-varying Gaussian Markov random fields (SV-GMRF) where the goal is to learn a network of sparse, context-specific GMRFs representing network relationships between genes. An important application of SV-GMRFs is in inference of gene regulatory networks from spatially-resolved transcriptomics datasets. The current work on inference of SV-GMRFs are based on the regularized maximum likelihood estimation (MLE) and suffer from overwhelmingly high computational cost due to their highly nonlinear nature. To alleviate this challenge, we propose a simple and efficient optimization problem in lieu of MLE that comes equipped with strong statistical and computational guarantees. Our proposed optimization problem is extremely efficient in practice: we can solve instances of SV-GMRFs with more than 2 million variables in less than 2 minutes. We apply the developed framework to study how gene regulatory networks in Glioblastoma are spatially rewired within tissue, and identify prominent activity of the transcription factor HES4 and ribosomal proteins as characterizing the gene expression network in the tumor peri-vascular niche that is known to harbor treatment resistant stem cells.

## 1 Introduction

The advent of high throughput sequencing technologies has transformed our understanding of biological systems, and catalyzed the adoption of a systems-level approach to studying biological processes . Networks have emerged as the intuitive framework for reasoning about complex biological systems [5, 90]. Nodes in the network represent individual components, and edges represent direct interactions between them. For example, gene regualtory networks (GRNs) represent the wiring diagram of the cell's information processing system, with network edges identifying regulatory

interactions between different genes. It has become clear that complex diseases like cancer must be understood at the level of this interactome, rather than the classical reductionist approach of studying individual components [23, 30]. As another example, with billions of neurons and hundreds of thousands of voxels, the human brain is considered as one of the most complex physiological networks, whose structure remains as a long-standing mystery [77, 50, 61, 71, 53]. The accurate inference of the brain connectivity network will have a far-reaching impact on understanding different neurological disorders [31, 8, 66]. According to the NIH's *BRAIN Initiative*, the development of "faster, less expensive, and scalable" technologies is the cornerstone for anatomic reconstruction of neural circuits at realistic scales [6].

Spatially resolved transcriptomics have emerged as a transformative technology in the recent past with immense potential to bolster our understanding of biology ant a tissue architecture level[65, 60, 76]. Depending on the technology used, we can measure gene expression profile at near single cell resolution at the transcriptome-scale in situ [3, 70]. In studying complex processes such as tumor growth, viewing cancer as a case of evolution within the tissue has provided the groundwork for building a comprehensive theoretical framework to understand tumor diversity [67]. Evolutionary trade-offs between proliferation and survival strategies amongst cancer cells are driven by spatial gradients in exposure to nutrients, oxygen, immune cells and environmental toxicity between the tumor core versus periphery [20, 21, 49]. The need to optimize growth of the tumor through evolution of the hallmark traits [46] leads individual cancer cells to adopt a continuum of transcriptional states, that maximize their performance given spatially-imposed metabolic and survival constraints [7, 42, 10]. There are therefore strong spatial trends in the gene expression profiles and the underlying regulatory networks even amongst differentiated cells of the same type in both homeostasis and diseased states. Being able to infer these dynamic regulatory networks would provide us with a new lens for understanding complex biological processes, and can lead to new hypotheses regarding molecular mechanisms that would inspire further experimental and theoretical investigations into the nature of regulatory interactions underlying disease states.

One popular approach to model these problems is based on *spatially-varying Markov Random Fields* (SV-MRFs). SV-MRFs are associated with a network of undirected *Markov graphs* $\mathcal{G}_k(V_k, E_k)$, where $V_k$ and $E_k$ are the set of nodes and edges in the graph at location $k$. The node set $V_k$ represents the random variables (e.g. genes) in the model, while the edge set $E_k$ captures the conditional dependency between these variables at location $k$. In the special case of Gaussian Markov Random Fields (GMRFs), the edge set of the Markov graphs can be fully characterized based on the inverse covariance matrix (also known as the precision matrix). In particular, if the entry $(i, j)$ of the precision matrix $\Theta_k$ is zero, then the variables $i$ and $j$ at location $k$ are independent conditioned on the remaining variables.

A widely-used method for the inference of SV-MRFs is based on the so-called regularized maximum-likelihood estimation (MLE). Intuitively, MLE seeks to find a graphical model based on which the observed data is most likely to occur. However, MLE-based methods suffer from major computational challenges that undermine their applicability in large-scale settings. For example, in the Gaussian setting, the MLE requires optimizing over the so-called log-determinant of the precision matrix, which are known to be intractable in large scales [13, 44, 33]. This drawback is further compounded in the spatially-varying regime, where the precision matrix must be estimated at each spatial location, leading to a dramatic increase in the size of the problem.

2

## 1.1 Our Contributions

To address the aforementioned challenges, we propose a *simple* estimation methods for the inference of spatially-varying GMRFs. Unlike MLE-based methods, our proposed approach is based on a class of simple and computationally efficient optimization methods that come equipped with strong statistical guarantees and are implementable in realistic scales. Our contributions are summarized as follows:

**Computational guarantee:** Our proposed method reduces to a series of decomposable convex quadratic optimization problems that can be solved efficiently using any off-the-shelf solvers. In addition, the decomposable nature of the proposed optimization problem makes it amenable to parallel and distributed implementation.

**Statistical guarantees:** In addition to the desirable computational guarantees, we show the statistical consistency of our proposed method—both theoretically and in practice. In particular, we characterize the non-asymptotic consistency of our proposed method, proving that it accurately recovers the underlying graphical model, even in the high-dimensional settings where number of available samples is significantly smaller than the number of unknown parameters. Moreover, it can efficiently reveal the correct sparsity information in the parameters and their differences.

**Application in inferring gene regulatory networks:** Glioblastoma (GBM) is an incurable malignancy of the brain, with a median survival time of only 12-18 months despite therapy with surgical resection, chemotherapy and radiation [47]. Despite aggressive treatment, these tumors inevitably recur and this recurrence is likely due to significant heterogeneity, which has been highlighted by single cell sequencing studies [87]. Heterogenous populations of treatment-resistant tumor cells with stem cell properties have been identified in GBM that have been shown to drive treatment recurrence. Furthermore, these resistant cells often reside within unique microenvironmental niches [1, 45, 72, 57]. The consequence of spatial context in regulating the tumor cell state, stemness properties, and treatment resistance in these tumors is increasingly appreciated [56, 51]. It is thus imperative that we understand how the gene networks of GBM cells are rewired as a function of their spatial environment, to identify context-specific upstream regulators of heterogenous tumor cell states. We thus employ our developed statistical framework to study how gene regulatory networks are spatially rewired in GBM.

We partition the tumor section into distinct micro-environmental niches and estimate networks involving genes showing significant spatial trends in their activity. We identify Transcription factors and hub genes that control tumor behavior in distinct local environments. We find that the perivascular tumor niche is characterized by high levels of activity of ribosomal genes and that HES4 is a prominant upstream regulator in this environment. Our findings have been previously reported to be particularly important aspects of the stemness features of Glioblastomas in recent literature [9, 79]. We feel that our ability to define context specific upstream regulators of tumor states is an important step in fighting tumor recurrence and developing targeted therapies for this disease.

## 1.2 Notations

The $i$-th element of a vector $v$ or $v_t$ is denoted as $v_i$ or $v_{t;i}$. For a matrix $M$, the notations $M_{i:}$ and $M_{:j}$ denote the $i$-th row and $j$-th column, respectively. Moreover, for an index set $S$ and a matrix $M$, the notations $M_{S:}$ and $M_{:S}$ refer to a submatrix of $M$ with rows and columns indexed by $M$, respectively. For a matrix $M$ or a vector $v$, the notations $\|M\|_{\ell_q}$ and $\|v\|_q$ correspond to

3

the element-wise $\ell_q$-norm of $M$ and $\ell_q$-norm of $v$, respectively. Moreover, $\|M\|_q$ and $\|M\|_{\max}$ are the induced $q$-norm and the element with the largest absolute value of the matrix $M$, respectively. Moreover, $\|M\|_0$ denotes the total number of nonzero elements in $M$. We use $M \succ 0$ to show that $M$ is positive definite. For a vector $v$ and matrix $M$, the notations $\text{supp}(v)$ and $\text{supp}(M)$ are defined as the index sets of their nonzero elements. Given two sequences $f(n)$ and $g(n)$ indexed by $n$, the notation $f(n) \lesssim g(n)$ implies $f(n) \leq Cg(n)$ for some constant $C < \infty$. Moreover, the notation $f(n) \asymp g(n)$ implies that $f(n) \lesssim g(n)$ and $g(n) \lesssim f(n)$. The sign function $\text{sign}(\cdot)$ is defined as $\text{sign}(x) = x/|x|$ if $x \neq 0$ and $\text{sign}(0) = 0$. Accordingly, when $x$ is a vector, the function $\text{sign}(x)$ is defined as $\begin{bmatrix} \text{sign}(x_1) & \text{sign}(x_2) & \dots & \text{sign}(x_n) \end{bmatrix}^\top$.

**Organization.** The rest of the paper is organized as follows. In Section 2, we formulate the inference of spatially-varying GMRFs and discuss the shortcomings of the existing techniques. Motivated by these shortcomings, we present a new formulation of the problem in Section 3. The related work is presented in Section 4. In Section 5, we delineate the statistical guarantees of our proposed formulation, and how to solve it efficiently. Finally, we showcase the performance of our proposed method on synthetically generated as well as the Glioblastoma spatial transcriptomics dataset in Section 7.

## 2 Problem Formulation

Consider data samples from $K$ different Gaussian distributions with covariance matrices $\Sigma_k^\star \in \mathbb{S}_+^d, k = 1,...,K$ and sparse precision matrices $\Theta_k^\star = \Sigma_k^{\star-1}, k = 1,...,K$. Let $\{x_i^k\}_{i=1}^{n_k}$ be $n_k$ independent samples drawn from the $k$-th distribution, i.e., $x_i^k \sim \mathcal{N}(0, \Sigma_k^\star)$, for every $i = 1, \dots, n_k$ and $k = 1, \dots, K$. Our goal is to estimate the precision matrices $\{\Theta_k^\star\}_{k=1}^K$ given the samples. The most commonly-used method to perform this task is via *maximum likelihood estimation* (MLE) with an $\ell_1$ regularizer (also known as *Graphical Lasso* [40]):

$$\widehat{\Theta}_k = \arg \min_{\Theta_k \succ 0} \text{Tr}(\Theta_k \widehat{\Sigma}_k) - \log \det(\Theta_k) + \lambda \|\Theta_k\|_{\ell_1}$$

where $\text{Tr}(\cdot)$ is the trace operator and $\widehat{\Sigma}_k := \frac{1}{n_k} \sum_{i=1}^{n_k} x_i^k x_i^{k\top}$ is the sample covariance matrix for distribution $k$. A major drawback of the above estimation method is that it ignores any common structure among different distributions. To address this issue, a common approach is to consider a joint estimation method (also known as *joint Graphical Lasso* [25]):

$$\{\widehat{\Theta}_k\} = \arg \min_{\Theta_k \succ 0} \sum_{k=1}^K \left( \text{Tr}(\Theta_k \widehat{\Sigma}_k) - \log \det(\Theta_k) + \lambda \|\Theta_k\|_{\ell_1} \right) + P\left( \{\Theta_k\}_{k=1}^K \right) \tag{1}$$

where the term $P\left( \{\Theta_k\}_{k=1}^K \right)$ is a penalty function that encourages similarity across different precision matrices. A major difficulty in solving joint Graphical Lasso is its computational complexity: in order to obtain an $\epsilon$-accurate solution, typical numerical solvers for (1) have complexity ranging from $\mathcal{O}(Kd^6 \log(1/\epsilon))$ (via general interior-point methods) [68, 74] to $\mathcal{O}(Kd^3/\epsilon)$ (via tailored first-order methods, such as ADMM) [44, 25, 63]. Solvers with such computational complexity fall short of any practical use in the large-scale settings. Indeed, the prohibitive worst-case complexity of methods based on Graphical Lasso is also exemplified in their practical performance [35, 91, 33, 36, 37, 34].

# 3 Proposed Method

To address the aforementioned issues, we propose the following surrogate optimization problem for estimating sparse precision matrices:

$$\{\widehat{\Theta}_k\} = \arg\min_{\Theta_k} \underbrace{\sum_{k=1}^K \|\Theta_k - \tilde{F}^*(\widehat{\Sigma}_k)\|_{\ell_2}^2}_{\text{backward mapping deviation}} + \underbrace{\mu \sum_{k=1}^K \|\Theta_k\|_{\ell_1}}_{\text{absolute regularization}} + \underbrace{\gamma \sum_{l>k} W_{kl} \|\Theta_k - \Theta_l\|_{\ell_q}^q}_{\text{spatial regularization}} \quad \text{(Elem-}q\text{)}$$

In the above optimization, the *backward mapping deviation* captures the distance between the estimated precision matrix and the so-called *approximate backward mapping* which will be described in Section 3.1. Moreover, the *absolute regularization* promotes sparsity in the estimated parameters, whereas *spatial regularization* encourages common spatial similarities among different parameters. For any given pair $(k,l)$, the weight $W_{kl}^{-1}$ can be interpreted as the "distance" between the $k$-th and $l$-th MRFs. Accordingly, a large value for $W_{kl}$ encourages similarity between $\Theta_k$ and $\Theta_l$. Two common choices of spatial similarities are sparsity and smoothness:

- *Smoothly-changing GMRF:* In smoothly-changing GMRFs, the adjacent precision matrices vary gradually. In this setting, $q = 2$ can be used as the spatial regularizer in Elem-$q$ to promote the smoothness in the parameter differences.

- *Sparsely-changing GMRF:* In sparsely-changing GMRFs, the adjacent precision matrices differ only in a few entries. In this setting, $q = 1$ is a natural choice for the spatial regularizer in Elem-$q$ since it promotes sparsity in the parameter differences.

## 3.1 Approximate Backward Mapping

Our proposed optimization problem is contingent upon the availability of an approximate backward mapping. For a GMRF, the backward mapping is defined as the inverse of the true covariance matrix, i.e., $F^*(\Sigma_k^\star) = \Sigma_k^{\star -1} = \Theta_k^\star$ [84]. Based on this definition, a natural surrogate for the backward mapping is $F^*(\widehat{\Sigma}_k) = \widehat{\Sigma}_k^{-1}$, where $\widehat{\Sigma}_k$ is the sample covariance matrix for distribution $k$. However, in the high-dimensional settings, the number of available samples is significantly smaller than the dimension, and as a result the sample covariance matrix $\widehat{\Sigma}_k$ is singular and non-invertible. To alleviate this issue, Yang et al. [88] introduce an approximation of the backward mapping based on soft-thresholding. Consider the operator $\text{ST}_\nu(M) : \mathbb{R}^{d \times d} \to \mathbb{R}^{d \times d}$, where $\text{ST}_\nu(M)_{ij} = M_{ij} - \text{sign}(M_{ij}) \min\{|M_{ij}|, \nu\}$ if $i \neq j$, and $\text{ST}_\nu(M)_{ij} = M_{ij}$ if $i = j$. Given this operator, the approximate backward mapping is defined as $\widetilde{F}^*(\widehat{\Sigma}_k) = \text{ST}_\nu(\widehat{\Sigma}_k)^{-1}$, for every $k = 1, \ldots, K$. An important property of this approximate backward mapping is that it is well-defined even in the high-dimensional setting $n_k \ll d$ with an appropriate choice of the threshold $\nu$ [88]. Given this approximate backward mapping, we will show that the estimated precision matrices from Elem-$q$ are close to their true counterparts with an appropriate choice of parameters.

## 3.2 Decomposability

An important property of Elem-$q$ is that it naturally decomposes over different coordinates of the precision matrices: for every $(i,j)$ with $1 \leq i \leq j \leq d$, the $ij$-th element of $\{\Theta_k\}_{k=1}^K$ can be obtained

by solving the following subproblem:

$$\{\widehat{\Theta}_{k;ij}\}_{k=1}^{K} = \arg \min_{\{\Theta_{k;ij}\}_{k=1}^{K}} \sum_{k=1}^{K} \left( \Theta_{k;ij} - [\tilde{F}^*(\widehat{\Sigma}_k)]_{ij} \right)^2 + \mu \sum_{k=1}^{K} |\Theta_{k;ij}| + \gamma \sum_{l>k} W_{kl} |\Theta_{k;ij} - \Theta_{l;ij}|^q,$$

$$(\text{Elem-}(i,j,q))$$

Recall that the original problem Elem-$q$ has $Kd(d+1)/2$ variables. The above decomposition implies that Elem-$q$ can be decomposed into $d(d+1)/2$ smaller subproblems, each with only $K$ variables that can be solved independently in parallel. This is in stark contrast with the joint Graphical Lasso, which requires a dense coupling among the elements of the precision matrices through the non-decomposable logdet function. Later, we will show how each subproblem can be solved efficiently for different choices of $q$.

## 4 Related Work

Recently, many approaches have been proposed for sparse precision matrix estimation in high dimensions. This line of work begins by the inference of a single precision matrix, which can be achieved by $\ell_1$-regularized MLE, also known as *Graphical Lasso (GL)* [39, 4, 89].

Extending beyond single precision matrix inference, a recent line of research has focused on estimating time-varying MRFs, where the relation among variables may change over time [94]. A common approach for estimating time-varying MRFs is based on kernel methods, where the sample covariance matrix at any given time is a weighted average of the samples over time, where the weights are collected from a predefined kernel [94, 41, 32].

In the context of spatially-varying graphical models, the main focus has been devoted to different variants of MLE-based techniques, such as *Fused Graphical Lasso* (FGL) and *Group Graphical Lasso* (GGL) [25]. FGL penalizes the pairwise difference of the precision matrices in $\ell_1$-norm, while GGL regularizes the $\ell_2$-norm of the $(i, j)$-th element across all $K$ precision matrices. Guo et al. [43] reparameterized each off-diagonal element as the product of a common factor and difference, then applied separate $\ell_1$ regularization to these two parts. Saegusa and Shojaie [78] proposed to regularize the MLE with a Laplacian-type penalty to exploit the information among different distributions. However, all these techniques are based on MLE, and consequently suffer from a notoriously high computational cost.

To alleviate the computational cost of MLE-based technique, Lee and Liu [59] proposed to estimate the joint precision matrices based on a *constrained $\ell_1$ minimization for inverse matrix estimation* (CLIME) technique [17]. Unlike GL, CLIME does not optimize over the complex logdet function and has shown more favorable theoretical properties than GL. Finally, our method is built upon the *Elementary Estimator* introduced by Yang et al. [88], where the proposed estimator admits a closed-form solution based on soft-thresholding. This method was later extended by Fattahi and Gomez [32] to time-varying setting, showing that it can be solved in near-linear time and memory.

## 5 Statistical Guarantees

In this section, we elucidate the statistical properties of Elem-$q$ for SV-GMRFs with two widely-used spatial structures, namely *smoothly-changing* and *sparsely-changing* GMRFs. To this goal, we first need to make two important assumptions on the true precision matrices.

**Assumption 1** (Bounded norm). *There exist constant numbers $\kappa_1 < \infty, \kappa_2 > 0$, and $\kappa_3 < \infty$ such that*

$$\|\Theta_k^\star\|_\infty \leq \kappa_1, \inf_{w:\|w\|_\infty=1} \|\Sigma_k^\star w\|_\infty \geq \kappa_2, \|\Sigma_k^\star\|_{\max} \leq \kappa_3$$

*for every $k = 1, \ldots, K$.*

Assumption 1 is fairly mild and implies that the true covariance matrices and their inverses have bounded norms.

**Assumption 2** (Weak sparsity). *Each covariance matrix $\Sigma_k^\star$ satisfies $\max_i \sum_{j=1}^d \left| [\Sigma_k^\star]_{ij} \right|^p \leq s(p)$, for some function $s : [1, \infty) \to \mathbb{R}$ and scalar $0 \leq p < 1$.*

Informally, we say "the true covariance matrices are weakly sparse" if $\{\Sigma_t^\star\}_{t=0}^T$ are $s(p)$-weakly sparse with $s(p) \ll d$ for some $0 \leq p < 1$. The notion of weak sparsity extends the classical notion of sparsity to dense matrices. Indeed, except for a few special cases, a sparse matrix does not have a sparse inverse. Consequently, a sparse precision matrix may not lead to a sparse covariance matrix. However, a large class of sparse precision matrices have weakly sparse inverses. For instance, if $\Theta_k^\star$ has a banded structure with small bandwidth, then it is known that the elements of $\Sigma_k^\star = \Theta_k^{\star-1}$ enjoy exponential decay away from the main diagonal elements [27, 52]. Under such circumstances, simple calculation implies that $s(p) \leq \frac{C}{1-\rho^p}$ for some constants $C > 0$ and $\rho < 1$. More generally, a similar statement holds for a class of inverse covariance matrices whose support graphs have large average path length [12, 11]; a large class of inverse covariance matrices with row- and column-sparse structures satisfy this condition.

Next, we introduce some notations that simplify our subsequent analysis. Let $\pi : \{1, 2, \ldots, K\}^2 \to \{1, 2, \ldots, K(K+1)/2\}$ be a fixed, predefined labeling function that assigns a label to each pair $(k, l)$ with $l \geq k$. Let $G$ be a diagonal matrix whose $k$-th diagonal entry is defined as $W_{\pi^{-1}(k)}^{1/q}$. Moreover, let $A \in \mathbb{R}^{K(K-1)/2 \times K}$ be the adjacency matrix defined as $A(\pi(k, l), k) = 1$ and $A(\pi(k, l), l) = -1$, for every $l > k$. Finally, define $\Theta_{ij} = [\Theta_{1;ij} \ \Theta_{2;ij} \ \ldots \ \Theta_{K;ij}]^\top$ and $\tilde{F}_{ij}^* = \left[ [\tilde{F}^*(\widehat{\Sigma}_1)]_{ij} \ [\tilde{F}^*(\widehat{\Sigma}_2)]_{ij} \ \ldots \ [\tilde{F}^*(\widehat{\Sigma}_K)]_{ij} \right]$, for every $j \geq i$. It is easy to see that $\|GA\Theta_{ij}\|_q^q = \sum_{l>k} W_{kl} |\Theta_{k;ij} - \Theta_{l;ij}|^q$ for every $j \geq i$, and accordingly, Elem-$(i, j, q)$ can be written concisely as

$$\widehat{\Theta}_{ij} = \arg\min_{\Theta_{ij}} \left\| \Theta_{ij} - \tilde{F}_{ij}^* \right\|_2^2 + \mu \|\Theta_{ij}\|_1 + \gamma \|GA\Theta_{ij}\|_q^q. \tag{2}$$

Next, we provide sharp statistical guarantees for our proposed method when the precision matrices $\{\Theta_k^\star\}_{k=1}^K$ change smoothly or sparsely across different distributions.

## 5.1 Smoothly-changing GMRF

We start with our main assumption on the smoothness of the precision matrices.

**Assumption 3** (Smoothly-changing SV-GMRFs). *There exists a constant $D \geq 0$ such that $\sum_{l \geq k} (\Theta_{k;ij}^\star - \Theta_{l;ij}^\star)^2 \leq D^2$ for every $(i, j)$.*

Informally, we say "SV-GMRF is smoothly-changing" if Assumption 3 is satisfied with a small $D$. For a smoothly-changing SV-GMRF, it is natural to choose $q = 2$ in Elem-$q$ to promote smoothness in the spatial difference of the precision matrices. Our next theorem characterizes the sample complexity of Elem-$q$ with $q = 2$ for smoothly-changing SV-GMRF. Let $\Theta_{\min} = \min\{|\Theta_{k;ij}^\star| : \Theta_{k;ij}^\star \neq 0\}$.

**Theorem 1** (Smoothly-changing SV-GMRF). *Consider a smoothly-changing SV-GMRF with parameter $D$, and weakly-sparse covariance matrices with parameter $s(p)$ for some $0 \leq p < 1$. Suppose that the number of samples satisfies*

$$n_k \gtrsim L \frac{\log d}{\Theta_{\min}^2}, \quad where \quad L = \max \left\{ \left(\frac{s(p)}{\kappa_2}\right)^{\frac{2}{1-p}} \kappa_3^2, \left(\frac{\kappa_1 \kappa_3}{\kappa_2} + D\right)^2 \right\}.$$

*Define $n_{\min} = \min_k\{n_k\}$. Moreover, suppose that $\widetilde{F}^*(\widehat{\Sigma}_k) = [ST_{\nu_k}(\widehat{\Sigma}_k)]^{-1}$ with $\nu_k \asymp \kappa_3 \sqrt{\log d / n_k}$. Then, the solution obtained from Elem-q with $q = 2$ and parameters*

$$\gamma \asymp \frac{1}{K \|W\|_{\max}} \sqrt{\frac{\log d}{n_{\min}}}, \qquad \mu \asymp D \sqrt{\frac{\log d}{n_{\min}}},$$

*satisfies the following statements with probability of $1 - Kd^{-10}$:*

- **Sparsistency:** *The solution is unique and satisfies $\mathrm{supp}(\widehat{\Theta}_k) = \mathrm{supp}(\Theta_k^\star)$ for every $k$.*

- **Estimation error:** *The solution satisfies*

$$\|\widehat{\Theta}_k - \Theta_k^\star\|_{\max} \lesssim \left(\frac{\kappa_1 \kappa_3}{\kappa_2} + D\right) \sqrt{\frac{\log d}{n_{\min}}}, \quad for \ every \ k.$$

For smoothly-changing SV-GMRF, the above theorem provides a non-asymptotic guarantee on the estimation error and sparsistency of the estimated precision matrices via Elem-$q$ with $q = 2$, proving that the required number of samples must scale only *logarithmically* with the dimension $d$. Moreover, both the estimation error and the required number of samples decrease with a smaller smoothness parameter $D$; this is expected since a small value of $D$ implies that the adjacent distributions share more information, and hence, the SV-GMRF is easier to estimate.

## 5.2 Sparsely-changing GMRF

In sparsely-changing SV-GMRFs, the precision matrices are assumed to change sparsely across different distributions; this is formalized in our next assumption.

**Assumption 4** (Sparsely-changing SV-GMRFs). *There exists a constant $D_0 \geq 0$ such that $\sum_{l \geq k} \|(\Theta_{k;ij}^\star - \Theta_{l;ij}^\star)\|_0 \leq D_0$ for every $(i, j)$.*

Similar to the smoothly-changing SV-GMRFs, we say "SV-GMRFs is sparsely-changing" if it satisfies Assumption 4 with a small $D_0$. For a sparsely-changing SV-GMRFs, it is natural to choose $q = 1$ in Elem-$q$ to promote sparsity in the spatial difference of the precision matrices. To analyze the statistical property of this problem, we first consider (2) with $q = 1$ and rewrite it as:

$$\min \ \|\widetilde{F}_{ij}^* - \Theta_{ij}\|_2^2 + \mu \|B\Theta_{ij}\|_1, \ \text{where} \ B = \begin{bmatrix} \frac{\gamma}{\mu} GA \\ I \end{bmatrix}. \tag{3}$$

The above reformulation is a special case of the *generalized Lasso problem* introduced by Lee et al. [58]. To show the model selection consistency of the above formulation, we next introduce the notion of *irrepresentability*.

For any fixed $(i, j)$, let $\mathcal{S}_B \subset \{1, 2, ..., K(K+1)/2\}$ be the support of $B\Theta_{ij}^{\star}$, i.e., $[B\Theta_{ij}^{\star}]_k \neq 0$ for every $k \in \mathcal{S}_B$. Moreover, let $\mathcal{S}_B^c = \{1, 2, ..., K(K+1)/2\} \backslash \mathcal{S}_B$. Evidently, we have $|\mathcal{S}_B| \leq D_0 + S_0$, where $D_0$ is introduced in Assumption 4 and $S_0$ is defined as the maximum number of nonzero elements in $\Theta_{ij}^{\star}$, i.e., $S_0 = \max_{i,j}\{\|\Theta_{ij}^{\star}\|_0\}$.

**Assumption 5** (Irrepresentability condition (IC), Lee et al. [58])**.** *We have*

$$\left\| B_{\mathcal{S}_B^c:} B_{\mathcal{S}_B:}^{\dagger} \operatorname{sign}\left((B\Theta_{ij}^{\star})_{\mathcal{S}_B:}\right) \right\|_{\infty} \leq 1 - \alpha \tag{4}$$

*for some $0 < \alpha \leq 1$, where $B_{\mathcal{S}_B:}^{\dagger}$ is the Moore-Penrose pseudo-inverse of a matrix $B_{\mathcal{S}_B:}$.*

The irrepresentability condition (IC) entails that the rows of $B$ corresponding to the zero elements of $B\Theta_{ij}^{\star}$ must be nearly orthogonal to the other rows. Despite the seemingly complicated nature of IC, classical results on Lasso have shown that it is a necessary condition for the exact sparsity recovery, and hence, cannot be relaxed [92, 82]. Later, we show that this condition is satisfied for our problem under a mild condition on the weight matrix $W$ and parameters $\mu$ and $\gamma$.

Another quantity that plays a central role in our derived bounds is the so-called *compatibility constant* defined as

$$\kappa_{\text{IC}} := \left\| B_{\mathcal{S}_B^c:} B_{\mathcal{S}_B:}^{\dagger} \right\|_{\infty} + 1.$$

The compatibility constant $\kappa_{\text{IC}}$ is closely related to IC. In particular, if $\left\| B_{\mathcal{S}_B^c:} B_{\mathcal{S}_B:}^{\dagger} \right\|_{\infty} \leq 1 - \alpha$ (which is a slightly stronger version of IC), then $\kappa_{\text{IC}} \leq 2 - \alpha$. Similar to IC, we will later show that $\kappa_{\text{IC}}$ remains bounded under a mild condition on the weight matrix $W$. Finally, we define $\Delta\Theta_{\min} = \min_{k,i,j}\{|\Theta_{k;ij}^{\star} - \Theta_{l;ij}^{\star}| : \Theta_{k;ij}^{\star} - \Theta_{l;ij}^{\star} \neq 0\}$.

**Theorem 2** (Sparsely-changing SV-GMRFs.)**.** *Consider a sparsely-changing SV-GMRFs with parameter $D_0$, and weakly-sparse covariance matrices with parameter $s(p)$ for some $0 \leq p < 1$. Suppose that the number of samples satisfies*

$$n_{\min} \gtrsim L \frac{\log d}{\min\{\Theta_{\min}^2, \Delta\Theta_{\min}^2\}}, \quad where \quad L = \left\{ \left(\frac{s(p)}{\kappa_2}\right)^{\frac{2}{1-p}} \kappa_3^2, \left(\frac{\kappa_{\text{IC}}\kappa_1\kappa_3}{\kappa_2\alpha}\right) (\|W\|_{\max} D_0 + S_0) \right\}.$$

*Define $n_{\min} = \min_k\{n_k\}$. Moreover, suppose that $\widetilde{F}^*(\widehat{\Sigma}_k) = [ST_{\nu_k}(\widehat{\Sigma}_k)]^{-1}$ with $\nu_k \asymp \kappa_3\sqrt{\log d / n_k}$. Moreover, suppose that the weight matrix $W$ and parameters $\mu$ and $\gamma$ are chosen such that IC (Assumption 5) is satisfied. Then, the solution obtained from Elem-q with $q = 1$ and parameter*

$$\mu \asymp \frac{\kappa_{\text{IC}}\kappa_1\kappa_3}{\kappa_2\alpha}\sqrt{\frac{\log d}{n_{\min}}},$$

*satisfies the following statements with probability $1 - Kd^{-10}$:*

- **Sparsistency.** *The solution is unique and satisfies $\operatorname{supp}(\widehat{\Theta}_k) = \operatorname{supp}(\Theta_k^{\star})$ for every $k$ and $\operatorname{supp}(\widehat{\Theta}_k - \widehat{\Theta}_l) = \operatorname{supp}(\Theta_k^{\star} - \Theta_l^{\star})$ for every $k > l$.*

- **Estimation error.** *For every $(i, j)$, the solution satisfies*

$$\left\| \widehat{\Theta}_{ij} - \Theta_{ij}^{\star} \right\|_2 \lesssim \left( \sqrt{\|W\|_{\max} D_0} + \sqrt{S_0} \right) \frac{\kappa_{\text{IC}}\kappa_1\kappa_3}{\kappa_2\alpha} \sqrt{\frac{\log d}{n_{\min}}}.$$

9

The above theorem characterizes the sample complexity of inferring sparsely-changing SV-GMRFs, showing that the sparsity pattern of the precision matrices and their differences can be recovered *exactly*, given that the number of samples scale logarithmically with the dimension and the problem satisfies IC. Evidently, our result crucially relies on the satisfaction of IC and $\kappa_{\mathrm{IC}}$ being small. This leads to a follow-up question: how restrictive are these conditions in practice? Our next proposition shows that both conditions hold if $\gamma$ and $\mu$ are selected such that $\mu \leq \gamma \leq 2\mu$ and $W_{kl}$ is the same for every $k > l$.

**Proposition 1.** *Suppose that* $0 < \mu \leq \gamma \leq 2\mu$ *and* $W_{kl}$ *is the same for every* $k > l$. *Then,* $1 \leq \kappa_{\mathrm{IC}} \leq 5$ *and IC holds with* $\alpha = \mu/\gamma$.

Proposition 1 can be easily extended to general choices of $W$. In particular, suppose that $W = \tau \mathbf{1}\mathbf{1}^\top + E$ for some $\tau > 0$, where $\mathbf{1}$ is the vector of ones. Then, Proposition 1 combined with a simple matrix perturbation bound reveals that

$$\alpha \geq 1/2 - \mathcal{O}(\|E\|_{\max}), \text{ and } 1 \leq \kappa_{\mathrm{IC}} \leq 5 + \mathcal{O}(\|E\|_{\max}).$$

In other words, IC holds and $\kappa_{\mathrm{IC}}$ remains bounded, provided that $\|E\|_{\max} = \mathcal{O}(1)$, that is, the elements of the weight matrix $W$ do not vary too much. Later in our numerical experiments, we will show that such choices of $W$ provide the best statistical results on both synthetically generated as well as gene expression datasets.

# 6   Parameter Tuning and Implementation

In this section, we explain different implementation aspects of our proposed method.

**Tuning $W$:** To obtain a solution for Elem-$q$, we first need to fine-tune the parameters $\mu, \gamma, \nu_k, W$ based on the available data samples. Recall that, for every pair $(k, l)$, the value of $W_{kl}^{-1}$ can be interpreted as the "distance" between precision matrices for distributions $k$ and $l$. Intuitively, $\Theta_k^\star$ and $\Theta_l^\star$ are close if their corresponding covariance matrices $\Sigma_k^\star$ and $\Sigma_l^\star$ are close. Therefore, to obtain an estimate of $W$, we first compute the distance between any pair of sample covariance matrices $D_{kl} = \left\| \widehat{\Sigma}_k - \widehat{\Sigma}_l \right\|_{\ell_2}$, and then assign $W_{kl} = 1/(1 + D_{kl})$ for every $k \neq l$.

**Tuning $\mu, \gamma$, and $\nu_k$:** Recall that the parameter $\mu$ controls the sparsity of the estimated precision matrices, whereas $\gamma$ penalizes their differences. Moreover, $\nu_k$ is the threshold for used in the proposed approximate backward mapping. In Theorems 1 and 2, we provide an explicit value for these parameters that depend on the parameters of the true solution, which are not known *a priori*. Without any prior knowledge on the true solution, these parameters can be selected by minimizing the *extended Bayesian Information Criterion* (BIC) [38]:

$$(\widehat{\mu}, \widehat{\gamma}, \widehat{\nu}) = \arg\min_{\mu, \gamma, \nu} \mathrm{BIC}(\mu, \gamma, \nu), \quad \text{where}$$

$$\mathrm{BIC}(\mu, \gamma, \nu) := \sum_{k=1}^{K} n_k [\mathrm{Tr}(\widehat{\Sigma}_k \widehat{\Theta}_k(\mu, \gamma, \nu)) - \log \det \widehat{\Theta}_k(\mu, \gamma, \nu)] + \log(n_k)\mathrm{df}^{(k)} + 4\mathrm{df}^{(k)} \log d, \tag{5}$$

In the above definition, $\widehat{\Theta}_k(\mu, \gamma, \nu)$ is the optimal solution of (Elem-$q$) with parameters $(\mu, \gamma, \nu)$. Moreover, $\mathrm{df}^{(k)}$ is defined as the number of nonzero elements in $\widehat{\Theta}_k(\mu, \gamma, \nu)$. Theorems 2 and 1 suggest that $\gamma = C_1\sqrt{\log d/n_{\min}}$, $\mu = C_2\sqrt{\log d/n_{\min}}$, and $\nu_k = C_3\sqrt{\log d/n_k}$, where $C_1, C_2$, and

10

$C_3$ are constants that depend on the true solution. Therefore, to pick the parameters $\mu$, $\gamma$, and $\nu_k$, we perform a grid search over the constants $C_1$, $C_2$, and $C_3$, and pick those that minimize $\mathrm{BIC}(\mu, \gamma, \nu)$.

**Algorithm:** Next, we explain a general algorithm for solving Elem-$q$. As mentioned before, Elem-$q$ decomposes over different coordinates $(i, j)$, where each subproblem can be written as Elem-$(i, j, q)$. This decomposition leads to a parallelizable algorithm, where each thread solves Elem-$(i, j, q)$, for a subset of the coordinates $(i, j)$. This approach is oulined in Algorithm 1.

---

**Algorithm 1** General algorithm for solving Elem-$q$

---

**Input:** Data samples $\{x_i^k\}$, parameters $(\mu, \gamma, \nu_k)$, and weight matrix $W$.;
**Output:** Solution $\{\widehat{\Theta}_k\}_{k=1}^K$ for Elem-$q$;
Compute the sample covariance matrix $\widehat{\Sigma}_k$ for every $k = 1, ..., K$;
Compute the approximate backward mapping $\widetilde{F}^*(\widehat{\Sigma}_k) = \left[\mathtt{ST}_{\nu_k}(\widehat{\Sigma}_k)\right]^{-1}$ for every $k = 1, ..., K$;
**for** every $(i, j)$ **do**
   Compute a sub-gradient $D_t \in \partial f_{\ell_1}(U_t)$;
   Obtain $\widehat{\Theta}_{ij}$ by solving Elem-$(i, j, q)$;
**end for**

---

Next, we analyze the computational cost of each step of our proposed algorithm. Given $n_k$ number of samples, the sample covariance matrix $\widehat{\Sigma}_k$ can be computed in $\mathcal{O}(n_k d^2)$ time and memory (Line 3). Moreover, given each sample covariance matrix, the approximate backward mapping can be obtained by an element-wise soft-thresholding followed by a matrix inversion, which can be done in $\mathcal{O}(d^3)$ time and memory (Line 4). Finally, for each $(i, j)$ and the choices of $q = 1, 2$, the subproblem Elem-$(i, j, q)$ can be reformulated as a linearly constrained convex quadratic problem. Suppose that $W$ has $\mathtt{nnz}$ number of nonzero elements. Then, each subproblem can be solved in $\mathcal{O}(\mathtt{nnz}^3)$ [13]. Moreover, assuming that the algorithm is parallelized over $M$ machines, the total complexity of solving all subproblems is $\mathcal{O}((d^2/M)\mathtt{nnz}^3)$. In the next section, we show that our proposed algorithm is extremely efficient in practice.

# 7 Numerical Experiments

In this section, we evaluate the performance of our proposed method on synthetically generated dataset, as well as the Glioblastoma spatial transcriptomics dataset. All experiments are implemented using MATLAB 2021b, and performed with a 3.2 GHz 8-Core AMD Ryzen 7 5800H CPU with 16 GB of RAM. We use the function $\mathtt{quadprog}$ in MATLAB to solve each subproblem.

## 7.1 Synthetically generated dataset.

First, we use synthetically generated dataset to compare the statistical performance of our proposed method with two other estimators: the fused graphical lasso (FGL) [25] and FASJEM [85]. FGL is an MLE-based approach augmented by a regularizer to promote spatial similarity among different distributions. On the other hand, FASJEM uses the same Elementary Estimator [88] framework as ours while having different regularization term. By comparing the estimated parameters with their true counterparts, we will show that our method outperforms both FGL and FASJEM in recovering the true precision matrices.

**Data generation:** Our data generation process is motivated by ideas proposed by Peng et al. [73] and Lyu et al. [62] to imitate the gene expression profiles from a synthetically-generated co-expression network. Our main goal is to generate the data synthetically from a known distribution, and then evaluate the performance of the estimated parameters by comparing them to the ground truth.

We simulate the true precision matrices for $K$ distinct clusters (populations) with varying level of similarity. Within each cluster, we assume that the graph representing the true precision matrix has a disjoint modular structure, with power law degree distribution for nodes within each module. Specifically, we split $d$ genes into $M$ modules, with $d/M$ genes per module generated based on Barabasi-Albert model [2]. Within each cluster, the modules are simulated independently and concatenated to produce a block-diagonal matrix, which is treated as the true precision matrix for the corresponding cluster.

In order to simulate the true precision matrices for all clusters, we first generate a random spanning tree over clusters. Starting at the root cluster, we generate $M$ modules, and in each module, we randomly generate a graph with $d/M$ vertices according to the Barabasi-Albert model. Based on the adjacency matrix of this graph, we select the edge weights uniformly from $[-1, -0.4] \cup [0.4, 1]$. To ensure the positive semi-definiteness of the constructed precision matrix, we use 1.1 times the sum of the absolute values of all off-diagonal elements in each row as the value of the diagonal elements in that row. Finally, we construct the precision matrix as a block diagonal matrix with $M$ modules on the diagonal blocks. We then traverse the spanning tree from the root cluster and, at every new cluster, construct the precision matrix by perturbing its parent cluster. We consider two types of perturbations: $(i)$ edge weight perturbation; and $(ii)$ edge reconnection. To perform Type $(i)$ perturbation, we sample a subset of the $M$ modules at the parent cluster, and add a uniform perturbation from the interval $[-0.04, 0.04]$ to the non-zero edges. For Type $(ii)$ perturbation, we replace one of the $M$ modules with a newly simulated one following a power-law degree distribution. Thus, at every cluster, the precision matrix is slightly perturbed relative to its parent, and the precision matrix differences accumulate, which means that the number of different edges of two precision matrices increases with their distance. Figure 1 illustrates the precision matrices for the two adjacent clusters. Having simulated the precision matrices, at every cluster $k$, we next collect $n_k$ samples from a zero-mean Gaussian distribution with the constructed precision matrix.

*Experiment 1: Varying number of samples.* In our first experiment, we fix $K = 5, d = 250$, and $M = 5$, and compare the performance of Elem-$q$ with FGL and FASJEM with varying number of samples $n_k$. We compare the estimation accuracy in terms of Recall $= \text{TP}/(\text{TP} + \text{FN})$, Precision $= \text{TP}/(\text{TP} + \text{FP})$, and F1-score $= 2(\text{Recall} \times \text{Precision})/(\text{Recall} + \text{Precision})$, where TP, FN, and FP correspond to the true positive, false negative, and false positive values, respectively. To fine-tune the weight matrix $W$ and the parameters $(\mu, \gamma, \nu_k)$, we use the distance measure and BIC approach delineated in Section 6. Moreover, we use the same BIC approach to fine-tune the parameters of FGL and FASJEM.

Figure 2 illustrates the performance of different estimation methods. It can be seen that Elem-$q$ with $q = 1, 2$ (denoted as *Elem*-1 and *Elem*-1) perform almost the same, and they both outperform FASJEM and FGL in terms of the Precision and F1 scores. On the other hand, the Recall score for FASJEM is artificially high due to the underestimation of the regularization parameters via BIC, which in turn leads to overly dense estimation of the precision matrices.

*Experiment 2: Varying dimension.* Next, we analyze the performance of our proposed method for different dimensions $d$. In particular, we consider a high-dimensional regime where $d$ is significantly larger than the number of available samples $n_k$. We fix $K = 5$ and set $n_k = d/2$. The parameters
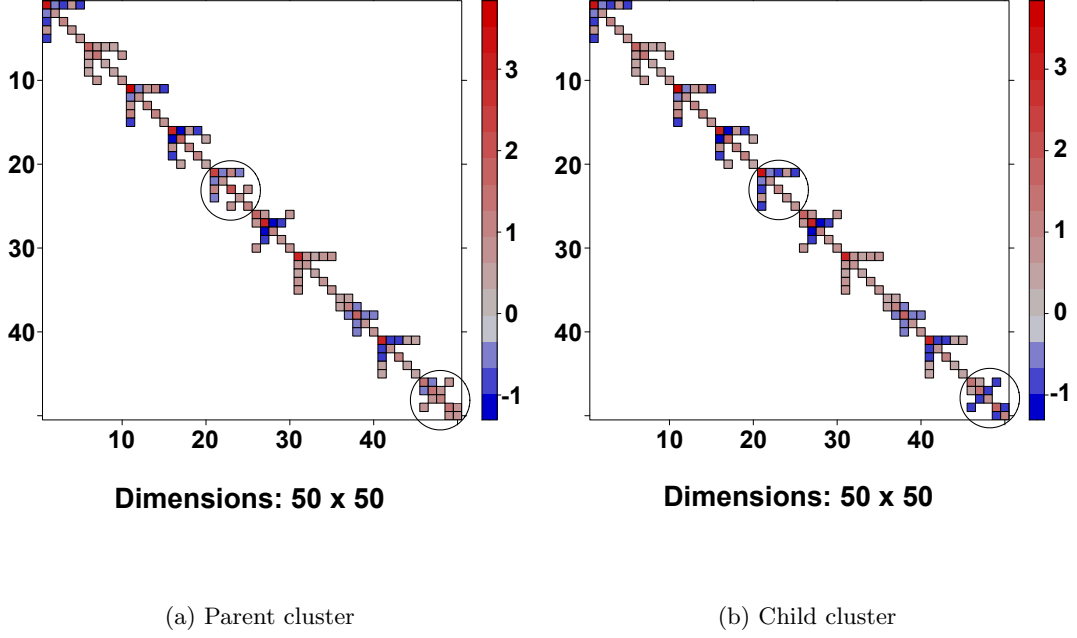
(a) Parent cluster

(b) Child cluster

Figure 1: The child cluster is obtained from the parent cluster by regenerating module 5 and perturbing the edge weights of module 10.
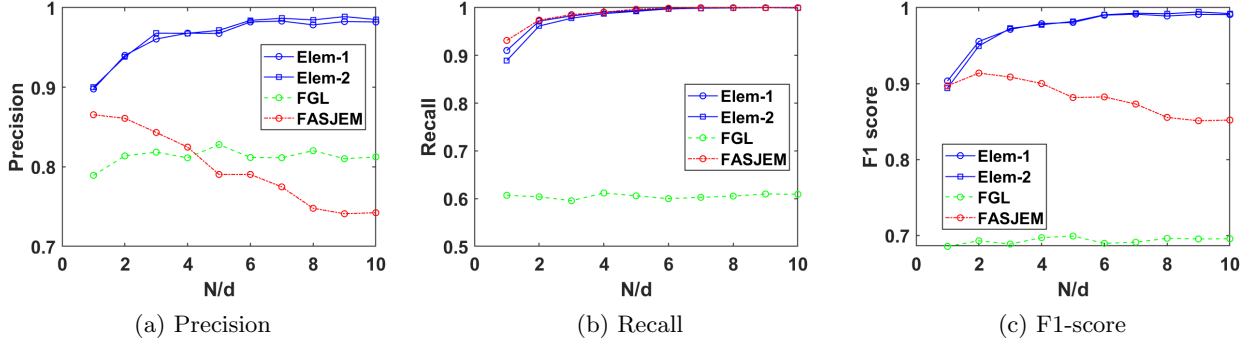


(a) Precision

(b) Recall

(c) F1-score

Figure 2: Precision, Recall, and F1-score for the estimated precision matrices for different methods with varying sample size. *Elem-1* and *Elem-2* perform similarly, and they both outperform *FGL* and *FASJEM*. The higher value of Recall for *FASJEM* is due to the underestimation of the regularization parameters that promote sparsity, which in turn leads to a large value of TP.

$\mu, \gamma, \nu_k$ and the weight matrix $W$ are tuned as before.

Figure 3 depicts the Precision, Recall, and F1-score, as well as the runtime of our proposed

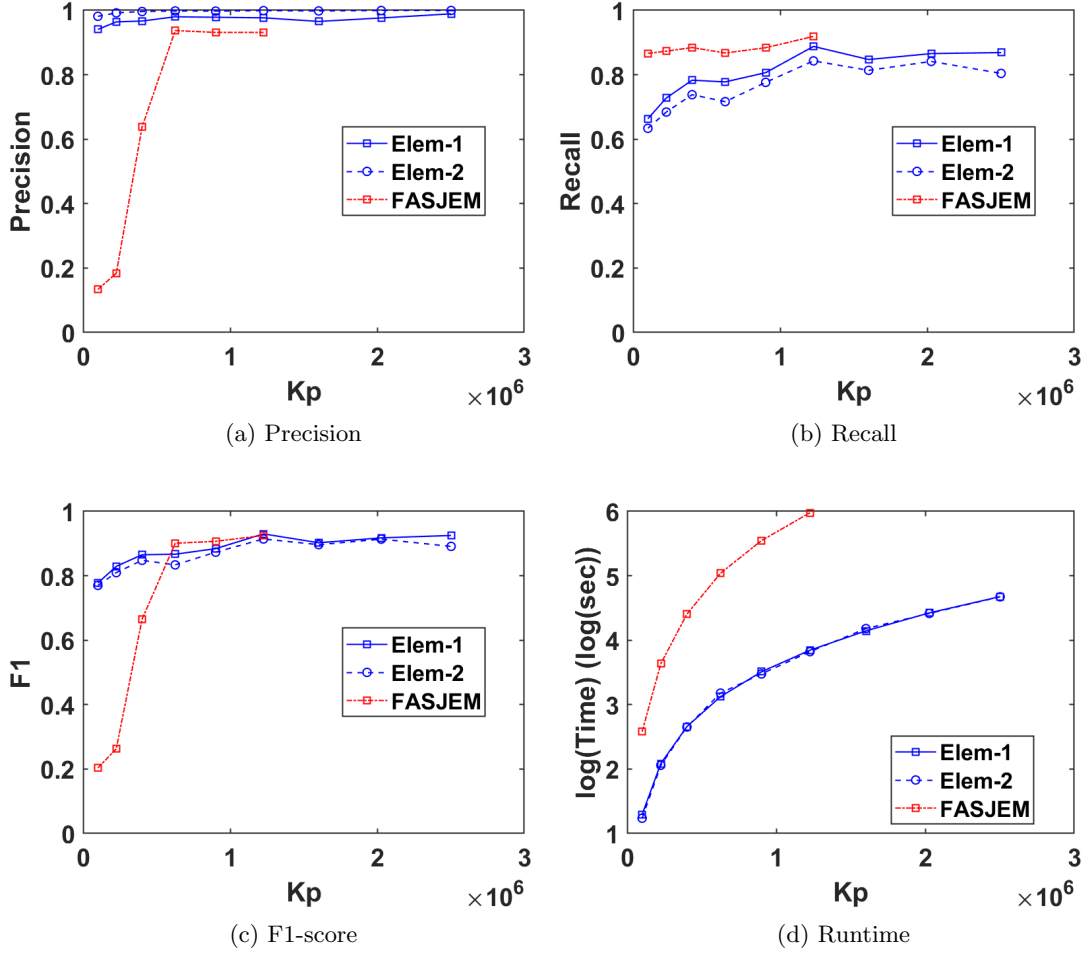(a) Precision  (b) Recall

(c) F1-score  (d) Runtime

Figure 3: Precision, Recall, F1-score for the estimated precision matrices, as well as the runtime of our proposed method with varying dimension. *Elem-1* and *Elem-2* outperform *FASJEM* in terms of Precision and F1-score, while *FASJEM* outperforming *Elem-1* and *Elem-2* in terms Recall. Similar to the previous experiment, the higher value of Recall for *FASJEM* is due to the underestimation of the regularization parameters which lead to overly dense precision matrices. Moreover, both *Elem-1* and *Elem-2* are drastically faster than *FASJEM*.

method and FASJEM with respect to $Kp = Kd(d+1)/2$ which ranges from $10^5$ to $2.5 \times 10^6$.[1] It can be seen that the runtime of our proposed method scales almost linearly with $p$, with the largest instance solved in less than 2 minutes. On the contrary, FASJEM has an undesirable dependency on $p$, with a runtime exceeding 10 minutes for medium-scale instances of the problem. The linear time of our algorithm with respect to $p$ is due to its decomposable nature of over different coordinates of the precision matrices.

*Experiment 3: Varying number of clusters.* Finally, we evaluate the performance of our method

---

[1] Due to the large scale of these instances, FGL did not converge within 10 minutes even for the smallest instance with $d = 200$. Therefore, it is omitted from our subsequent experiments.

(a) Precision         (b) Recall
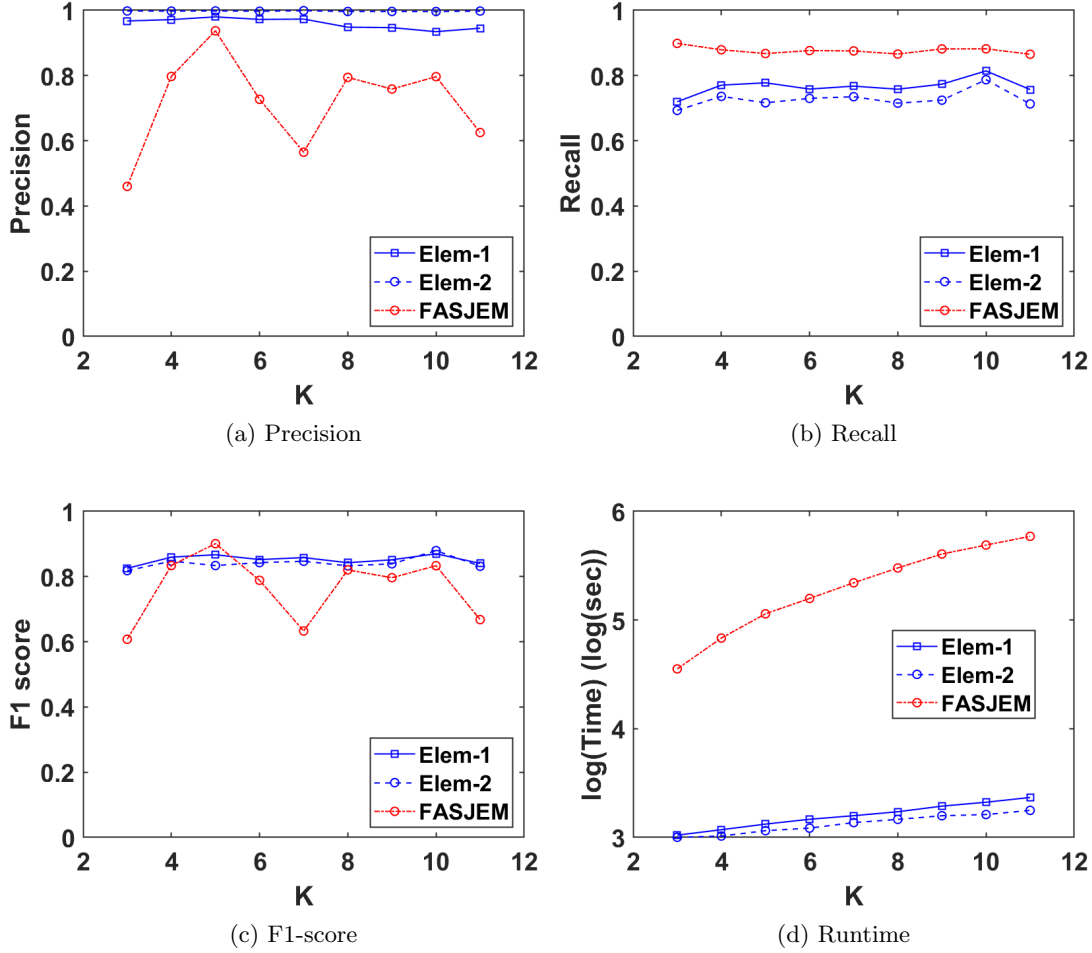
(c) F1-score         (d) Runtime

Figure 4: Precision, Recall, F1-score for the estimated precision matrices, as well as the runtime of our proposed method with varying number of clusters. *Elem-1* and *Elem-2* outperform *FASJEM* in terms of Precision and F1-score, while *FASJEM* outperforming *Elem-1* and *Elem-2* in terms Recall. Similar to the previous experiments, the higher value of Recall for *FASJEM* is due to the underestimation of the regularization parameters which lead to overly dense precision matrices. Moreover, both *Elem-1* and *Elem-2* are drastically faster than *FASJEM*.

with varying number of clusters $K$. We fix $d = 500$, $M = 10$ and $n_k = 250$, and use the same tuned parameters in the previous experiment. Figure 4 shows the Precision, Recall, and F1 score for our proposed method and FASJEM, as well as their runtime with respect to $K$. Similar to the previous experiments, both Elem-1 and Elem-2 outperform FASJEM in terms of the estimation accuracy. Moreover, it can be seen that in practice, the runtime of Elem-1 and Elem-2 scale almost linearly with $K$.
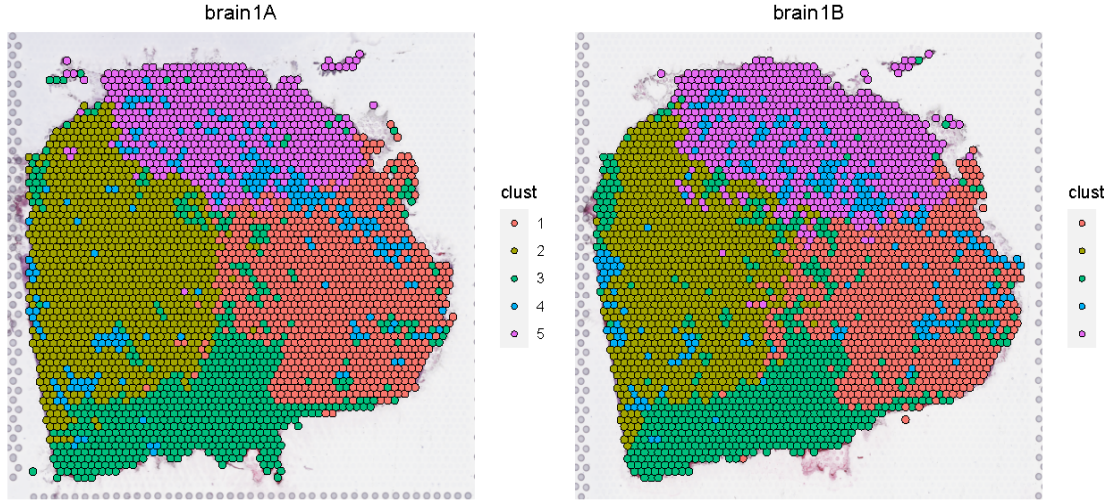
Figure 5: Adjacent tumor sections from a primary GBM patient sample are separated into five distinct clusters informed by their expression similarity and spatial proximity.

# 8 Application to Glioblastoma Spatial transcriptomics dataset

We collected gene expression profile using the Visium spatial transcriptomics (ST) platform from a primary GBM patient tumor showing high perfusion signal in diffusion MRI (relative cerebral blood volume parameter derived from dynamic susceptibility contrast MR perfusion). We sampled two adjacent tissue sections, giving us 6500 spots with transcriptomics data from this region. Since most routine clustering algorithms for spatial transcriptomics datasets are only based on expression-based proximity between cells, and completely ignore the spatial information, we first define a simple clustering algorithm that is also informed by spatial context.

We integrate data from adjacent tissue slices using the reciprocal PCA method in the R package for single cell data analysis (also known as *Seurat*) [48]. We use the dimension reduction algorithm PHATE [69] to obtain a 3D embedding of the integrated counts data. We then compute pairwise Euclidean distances between spots in the embedded space, and with their spatial coordinates. We perform upper quantile normalization of distance matrices based on their 75th quantile to ensure that both expression and spatial distances are in the same scale, and use their sum to define pairwise distances between spots. This dissimilarity matrix is used as input for PAM clustering. Optimal number of clusters (k = 5) is identified using the Calinski-Harabasz criterion [18], with the resulting clusters shown in Figure 5.

In order to understand biological characteristics of these clusters, and to aid in downstream interpretation of inferred networks, we performed spot deconvolution using the RCTD algorithm [16]. Since spots in the Visium microarray have a resolution of about $60\mu m$, they could be composed of multiple cell types. We thus used annotated single cell RNASeq dataset from [26] to identify cell type
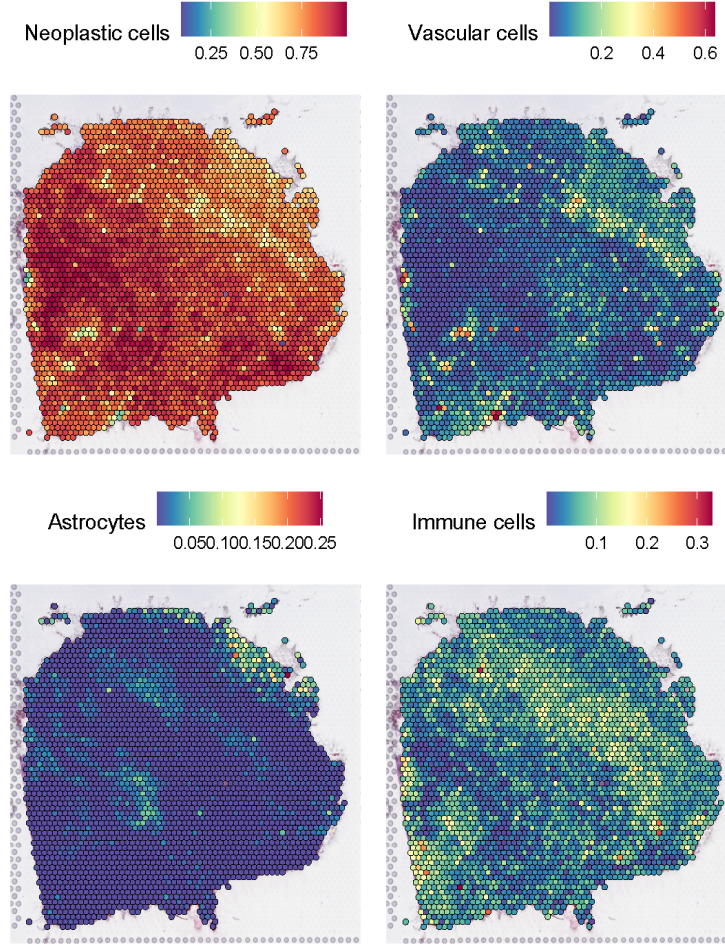
Figure 6: Output of RCTD spot deconvolution algorithm, visualized as fraction of spots composed of major contributing cell types. Cluster 4 is enriched for Vascular and immune cells, and cluster 5 has some nascent astrocytic populations.

compositional differences between the regions. We visualize in Figure 6 the proportion of each spot containing each of the major cell types. We see that the tissue is primarily composed of neoplastic cells, with some vascular niches and Astrocytic populations. We can see that Cluster 4 corresponds to a distinct vascular niche in the tumor with significant immune infiltration, and Cluster 5 has some non-tumor astrocytic cells. Thus the obtained clusters are biologically meaningful, and we can now seek to understand how gene network interactions vary in different microenvironments of this tumor.

Having defined biologically meaningful subdomains in the tumor section, we identify the top 2500 genes showing significant spatial trends in their expression, as determined using the SparkX algorithm [95]. Since the Visium data is highly sparse, we carry out a non-paranormal transformation of normalized spot-level counts data using the `huge` package in R [93]. Inter-cluster similarity constraints for network inference are imposed based on pairwise distances between the cluster medoids. Similar
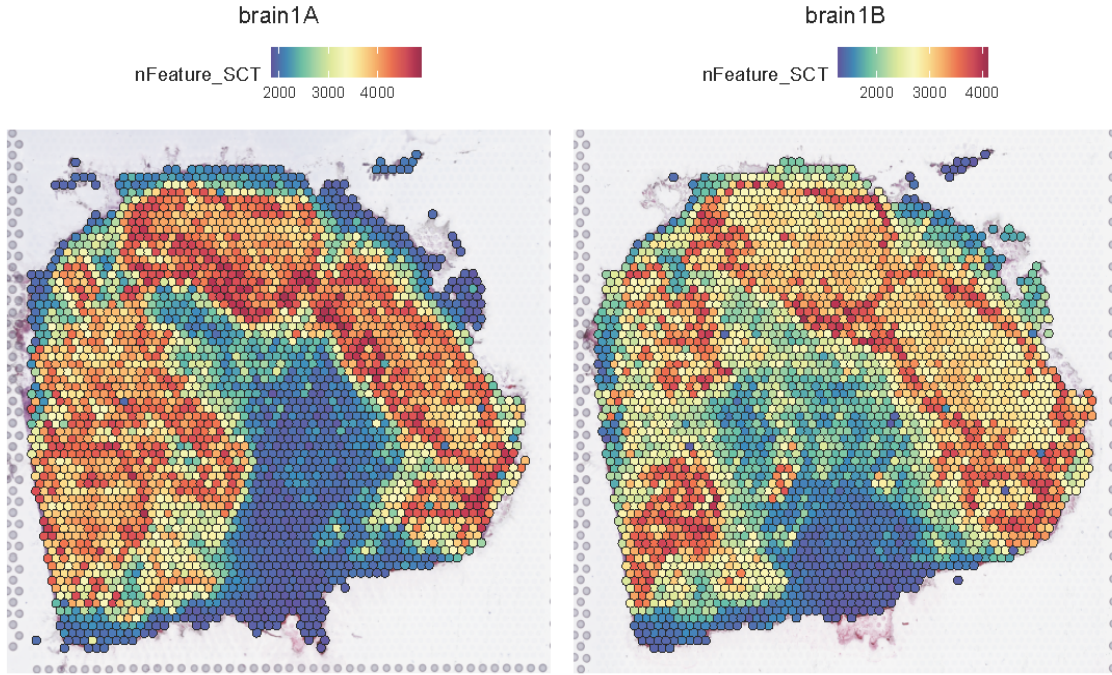
Figure 7: The number of detected reads and genes varies significantly across the tissue section. The regions associated with cluster 3 have the least number of detected genes, and could explain the very low number of edges in the corresponding graph.

to the previous case study, we use the BIC criterion to learn the optimal parameter values. Using these parameters, we find that of 2500 genes, 1180 have an edge in at least one cluster.

The number of inferred edges per cluster are respectively 4511, 13785, 446, 8400 and 4534. The significant variation in the number of edges in the clusters could be caused by the difference in numbers of detected read counts across the tissue section. This is shown in Figure 7, where we can see that Cluster 3 has significantly fewer detected genes than the other regions. While performing count imputation could help to some extent, we observed that the variation in the expressed gene counts persisted, showing similar spatial trends. This is therefore a biological and not technical artefact of the data.

To compare the extent of similarity between the inferred networks, we use the DeltaCon algorithm [55], a statistically principled and scalable inter-graph similarity function. The relative similarity values are shown in Figure 8. We can see that inferred networks from the different clusters are largely distinct, with Clusters 1 and 3 having the maximum pairwise similarity of 0.27. The network in Cluster 4 is maximally different from the other regions. This is as expected, given that this region is compositionally most unique. The gene network in Cluster 2 is also very different from other zones. As we saw in Figure 7, this region is transcriptionally most active, and the inferred graph
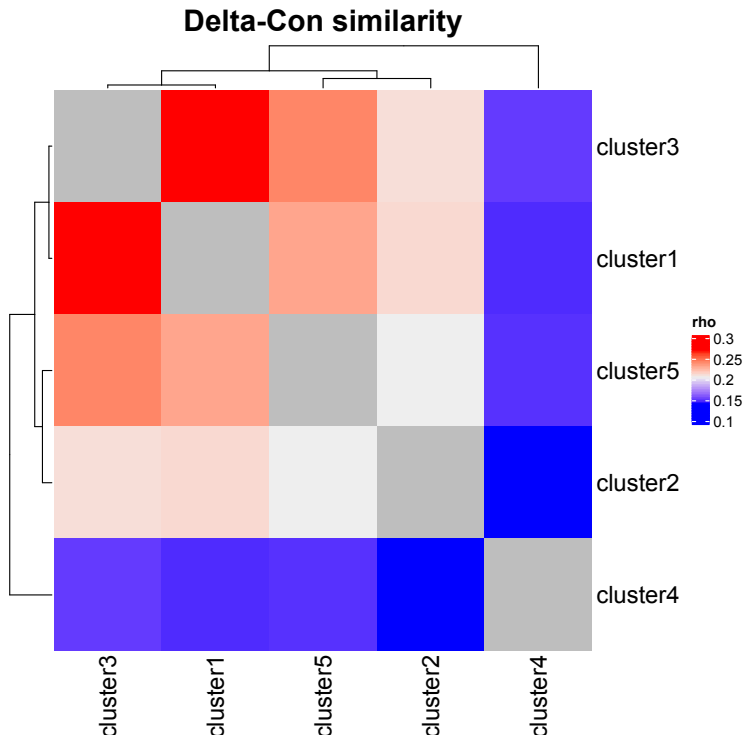
**Delta-Con similarity**

Figure 8: Inter-network similarity computed using the Delta-Con algorithm highlight extent of change in gene regulatory networks between different clusters.

has nearly twice as many interactions between genes as in the other regions.

Figure 9 shows the top 15 strongest edges in each cluster from the inferred network, with node sizes scaled by their respective degree and negative interactions shown as dotted lines. We can immediately see that each cluster has distinct underlying regulatory interactions driving their transcriptional states, even if they appear compositionally homogeneous.

Transcription factors are proteins that play a dominant role in regulating gene expression networks of cells and are particularly important in driving tumor growth and evolution [15]. By binding to regulatory regions of target genes, they are responsible for enhancing or suppressing gene expression and thereby controlling cell states. Regulatory interactions involving TFs are therefore of particular interest in understanding gene networks. We highlight these interactions in Figure 10. Since Cluster 2 has an order of magnitude more edges, we highlight only the top 100 edges.

Frequently highlighted TFs active across different regions include the AP1 family TFs FOS and JUN, which are known downstream effectors of the Mesenchymal state in Gliomas [64], and other master regulators such as CEBPD [86] and oligodendroglial lineage factors OLIG1 and OLIG2 [54]. We also see significant activity of SOX2 in Cluster 1, a known drivers of stemness features and radiation-resistance in Gliomas [80]. Cluster 3 shows significant activity of Lactotransferrin (LTF), which encodes an iron-binding protein with known innate immune and tumor-suppresive activity [24]. Interestingly, this gene has also been characterized as being an upstream master regulator of different GBM subtypes [14], warranting further exploration of this gene in driving tumorigenesis in GBM.
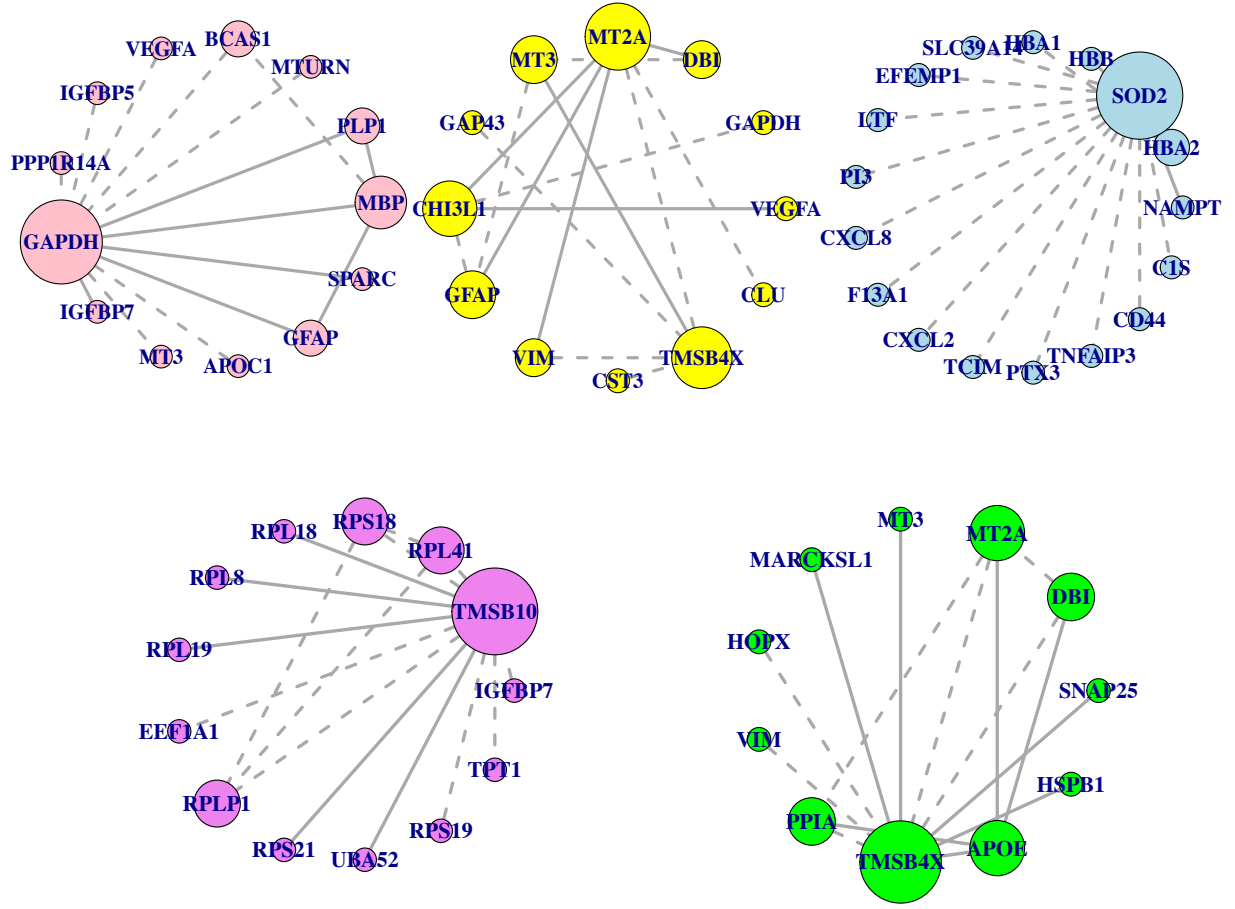
Figure 9: Top 15 strongest edges in each network are shown. Size of the nodes reflects their degree. Dotted lines indicate negative interactions. The nodes are colored by the cluster identity.

The TF network in Cluster 4, which represents the peri-vascular niche, is most different from the other regions, as expected given its unique microenvironment. This region shows prominent activity of HES4, a known downstream effector of the NOTCH signaling pathway that is known to inhibit cell differentiation and helps maintain the stemness features in Gliomas [9]. HES4 specifically regulates proliferative properties of neural stem cells, and reduces their differentiation. This is a very promising observation, given that the perivascular niche is known to harbor therapy-resistant glioma stem cells whose properties are critically driven through NOTCH signaling [51]. Cluster 5 has dominant activity of NME2, an nucleotide diphosphate kinase enzyme involved in cellular nucleotide metabolism and DNA repair [75]. The NME2 protein has also been identified to be a highly specific Tumor-associated antigen in IDH mutant Gliomas [28]. By studying the regulatory networks in each cluster separately, we are thus able to infer differential activity of different master regulators in different microenvironmental niches, which informs us of the varying cell states.

We observed that the majority of the estimated edges are unique to the respective clusters, with
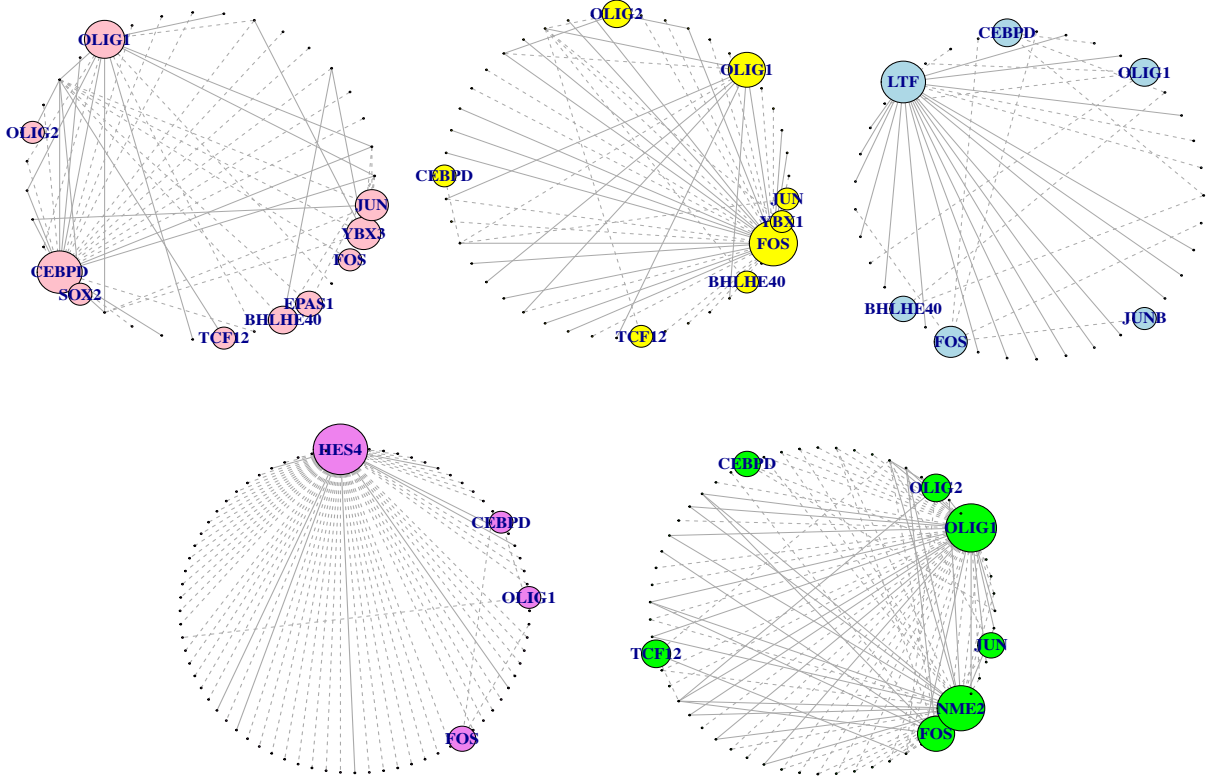
Figure 10: Transcription factor interactions in each cluster are highlighted. For each cluster, we show the top 100 strongest edges involving TFs. The graphs are visualized using the Davidson-Harel layout. Color scheme is same as in Figure 9.

Cluster 4 being most different with 84% unique edges. This network is characterized by a very high degree of connectivity between genes encoding ribosomal proteins. About a third of the nodes (60 / 184) having node degree over 100, and is qualitatively very different from the degree distribution we saw in the other networks. We next characterized the networks using different centrality measures such as their degree, betweenness, closeness, eigen and pagerank, each of which measures a different aspect of importance of nodes [29]. We reduce each network to its set of unique edges, and consider the top 10% of nodes by each centrality measure to be hub genes. We then compare multiplicity of these nodes across the different networks. As expected, majority of hub genes are specific to individual networks. However in spite of removing shared edges, we observe that the gene TMSB4X is identified as a shared hub gene in Clusters 1, 2 and 5, suggesting its potential importance in Glioma growth. We also note that TMSB4X has the highest degree of all genes in a base network that is shared across clusters. This gene encodes an actin-sequestering polypeptide, and is known to

promote stemness and invasive phenotype in Gliomas [22].

Next we performed a Gene Ontology Enrichment analysis with the cluster-specific hub genes shown in Figure 11. We can clearly see that the networks in each cluster are specific for different tasks. Cluster 1 shows an enrichment for neural differentiation related genes, which are known to be downregulated in GBM, as well as extracellular matrix related processes associated with invasive properties of the tumor. Cluster 2 specific hubs are associated with metabolic and biosynthetic processes. Cluster 3 hubs are associated with innate immune responses, in agreement with our observation that LTF is a major TF in this network. Cluster 4 has a large number of ribosomal genes with high connectivity. High levels of ribosomal protein activity has been shown to be associated with promoting stemness characteristics of Gliomas [79], and we see it to be a defining characteristic of the peri-vascular niche. Cluster 5 shows enrichment for neuronal processes like synaptic transmission, consistent with the presence of normal astrocytes in this region.

In summary, using our scalable framework of inferring gene regulatory networks across multiple spatially informed clusters, we are able to learn and characterize variations in tumor cell states with their microenvironmental niches and identify master regulators that are differentially active in each region. We are able to reinforce the role of known TFs such as CEBPD, FOS-JUN, OLIG and SOX family TFs, as well as identify less known drivers of context-dependent tumor adaptation. Using our method, we can demonstrate the significant strengths and show how it can be used to get a very deep understanding of tumor growth and adaptation from spatial gene expression datasets.

# 9 Conclusion

In this work, we study the inference of spatially-varying Gaussian Markov random fields (SV-GMRFs) and its application in gene regulatory networks in Spatially resolved transcriptomics. The existing methods for inferring GMRFs suffer from the so-called *curse of dimentionality*, which limit their applicability to small-scale and spatially-invariant networks. To address this challenge, we propose a simple and efficient inference framework for inferring SV-GMRFs that comes equipped with strong statistical guarantees. Contrary to the existing MLE-based methods, our proposed method is amenable to parallelization and is based on solving a series of decomposable convex quadratic programs. We show that our proposed method is extremely efficient in practice, and outperforms the existing state-of-the-art techniques—both computationally and statistically. We study the developed framework in the context of inferring gene networks underlying oncogenesis, using Glioblastoma (GBM) as a case study. We uncover the nature of spatial gene relationships across multiple subregions of the tumor. Given that the tumor cell states are dynamically regulated by their spatial context, our discovered context specific master regulators is an important step towards developing targeted therapies in the future.
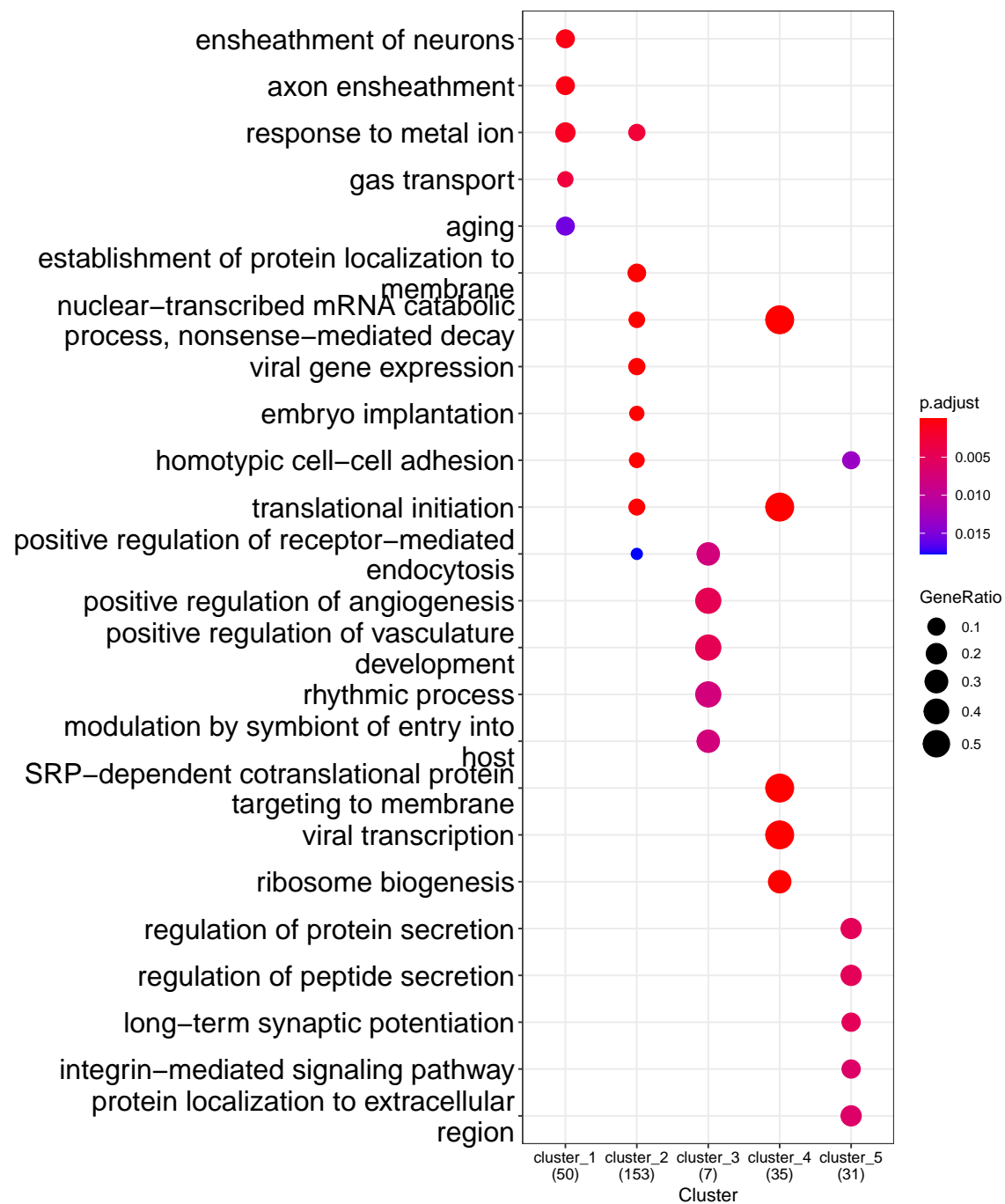
Figure 11: GO enrichment analysis for cluster-specific hub genes shows specific activity of different biological processes across the tumor section. Cluster 4 that represents a perivascular niche shows a high level of activity of ribosomal genes and translational activity.

# A  Proof of Theorem 1

The overarching idea behind the proof of Theorem 1 is as follows: we first derive a deterministic guarantee on the estimation error of the proposed method. We then prove our main theorem by extending this result using probabilistic concentration bound.

Our first result provides a set of deterministic conditions for the optimal solution to have small estimation error and correct sparsity pattern.

**Proposition 2** (Deterministic guarantee for smoothly-changing SV-MRFs). *For a given index* $(i, j)$, *suppose that the regularization parameters satisfy*

$$\gamma < \frac{1}{2K \|W\|_{\max}}, \tag{6}$$

$$6\|\tilde{F}^*_{ij} - \Theta^\star_{ij}\|_\infty + 6\gamma K \|W\|_{\max} D \leq \mu < \frac{3\Theta_{\min}}{2}. \tag{7}$$

*Then, the following statements hold:*

- *(Sparsistency) The solution* $\widehat{\Theta}_{ij}$ *is unique and satisfies* $\operatorname{supp}(\widehat{\Theta}_{ij}) = \operatorname{supp}(\Theta^\star_{ij})$.

- *(Estimation error) We have* $\|\widehat{\Theta}_{ij} - \Theta^\star_{ij}\|_\infty \leq 2\mu/3$

The proof of the above proposition is provided in Appendix D.1. Next, we show how this proposition can be used to complete the proof of Theorem 1. In order to invoke Proposition 2, we first need to prove that its conditions are satisfied. Recall that $\gamma = \sqrt{\log d/(K^2 \|W\|^2_{\max} n_{\min})}$ and $n_{\min} \geq c_1 \left( s(p)^{\frac{2}{1-p}} \log d \right) / \Theta^2_{\min} \geq c_1 \log d$ for some $c_1$. It is easy to see that, for $c_1 > 4$, we have $\gamma < 1/(2K \|W\|_{\max})$ and hence, the first condition (6) is satisfied. To show the validity of the second condition (7), we need the following intermediate lemma borrowed from [33].

**Lemma 1** (Theorem 3 of [33]). *Under the conditions of Theorem 1, the following inequality holds with probability of at least* $1 - Kd^{-10}$

$$\left\| \widetilde{F}^*(\widehat{\Sigma}_k) - \Theta^\star_k \right\|_{\max} \leq \frac{128\kappa_1\kappa_3}{\kappa_2} \sqrt{\frac{\log d}{n_k}}, \quad \text{for every } k.$$

*Proof.* The proof is a direct consequence of the proof of Theorem 3 in [33]. The details are omitted for brevity. □

*Proof of Theorem 1.* Based on Lemma 1 and our choice of $\gamma$, one can write

$$6\|\tilde{F}^*_{ij} - \Theta^\star_{ij}\|_\infty + 6\gamma K \|W\|_{\max} D \leq \left( \frac{768\kappa_1\kappa_3}{\kappa_2} + 6D \right) \sqrt{\frac{\log d}{n_{\min}}} := \mu$$

Now, it only remains to show that the defined $\mu$ indeed satisfies $\mu < 3\Theta_{\min}/2$. This can be readily verified by our choice of $n_k$:

$$n_k \geq c_2 \left( \frac{\kappa_1\kappa_3}{\kappa_2} + D \right)^2 \frac{\log d}{\Theta^2_{\min}}$$

for sufficiently large constant $c_2$. Therefore, the conditions of Proposition 2 are satisfied, and as a result, we have $\text{supp}(\widehat{\Theta}_{ij}) = \text{supp}(\Theta_{ij}^\star)$ for every $(i,j)$ and

$$\left\|\widehat{\Theta}_k - \Theta_k^\star\right\|_{\max} \leq 2\mu_3/3 \lesssim \left(\frac{\kappa_1\kappa_3}{\kappa_2} + D\right)\sqrt{\frac{\log d}{n_{\min}}}$$

which completes the proof of Theorem 1. $\qquad\square$

# B  Proof of Theorem 2

Similar to the proof of Theorem 1, first we provide a deterministic guarantee on the estimation error.

**Proposition 3** (Deterministic guarantee on sparsely-changing SV-MRFs)**.** *For a given $(i,j)$, suppose that the irrepresentability assumption is satisfied. Moreover, suppose that*

$$\frac{8\kappa_{\text{IC}}}{\alpha}\|\tilde{F}_{ij}^* - \Theta_{ij}^\star\|_\infty < \mu < \frac{\min\{2\Theta_{\min}, \Delta\Theta_{\min}\}}{4\sqrt{2\,\|W\|_{\max}\,D_0} + 5\sqrt{S_0}},$$

*Then, the following statements hold:*

- *(Sparsistency) The solution $\widehat{\Theta}_{ij}$ is unique and satisfies $\text{supp}(\widehat{\Theta}_{ij}) = \text{supp}(\Theta_{ij}^\star)$ and $\text{supp}(\widehat{\Theta}_{k;ij} - \widehat{\Theta}_{l;ij}) = \text{supp}(\Theta_{k;ij}^\star - \Theta_{l;ij}^\star)$ for every $l > k$.*

- *(Estimation error) We have*

$$\left\|\widehat{\Theta}_{ij} - \Theta_{ij}^\star\right\|_2 \leq 2\left(\sqrt{2\,\|W\|_{\max}\,D_0} + 1.25\sqrt{S_0}\right)\mu.$$

The proof of the above proposition is provided in Appendix D.2. Based on the above proposition, we proceed with the proof of Theorem 2.

*Proof of Theorem 2.* In light of Lemma 1, we have

$$\left\|\widetilde{F}^*(\widehat{\Sigma}_k) - \Theta_k^\star\right\|_{\max} \leq \frac{128\kappa_1\kappa_3}{\kappa_2}\sqrt{\frac{\log d}{n_k}},$$

for every $k$ and with probability of at least $1 - Kd^{-10}$. This implies that

$$\frac{8\kappa_{\text{IC}}}{\alpha}\|\tilde{F}_{ij}^* - \Theta_{ij}^\star\|_\infty \leq \frac{1024\kappa_{\text{IC}}\kappa_1\kappa_3}{\kappa_2\alpha}\sqrt{\frac{\log d}{n_{\min}}} := \mu$$

Now, in order to use Proposition 3, it suffices to have

$$\mu < \frac{\min\{2\Theta_{\min}, \Delta\Theta_{\min}\}}{4\sqrt{2\,\|W\|_{\max}\,D_0} + 5\sqrt{S_0}}$$

25

which is guaranteed to hold once

$$n_{\min} \geq c_3 \left( \frac{\kappa_{\mathrm{IC}} \kappa_1 \kappa_3}{\kappa_2 \alpha} \right) \left( \frac{\|W\|_{\max} D_0 + S_0}{\min\{\Theta_{\min}, \Delta\Theta_{\min}\}} \right) \cdot \log d$$

for sufficiently large constant $c_3$. Therefore, Proposition 3 holds and we have $\mathrm{supp}(\widehat{\Theta}_k) = \mathrm{supp}(\Theta_k^\star)$ for every $k$ and $\mathrm{supp}(\widehat{\Theta}_{k;ij} - \widehat{\Theta}_{l;ij}) = \mathrm{supp}(\Theta_{k;ij}^\star - \Theta_{l;ij}^\star)$ for every $i,j$ and $l > k$. Moreover, we have

$$\left\| \widehat{\Theta}_{ij} - \Theta_{ij}^\star \right\|_2 \leq 2 \left( \sqrt{2 \|W\|_{\max} D_0} + 1.25 \sqrt{S_0} \right) \mu$$

$$\lesssim \left( \sqrt{\|W\|_{\max} D_0} + \sqrt{S_0} \right) \frac{\kappa_{\mathrm{IC}} \kappa_1 \kappa_3}{\kappa_2 \alpha} \sqrt{\frac{\log d}{n_{\min}}}.$$

This completes the proof. $\qquad\square$

## C Proof of Proposition 1

Without loss of generality and to streamline the presentation, we assume that $W_{kl} = 1$ for every $k, l$. It is easy to observe that $B_{\mathcal{S}_B:}$ has full column rank, thus we have $(B_{\mathcal{S}_B:})^\dagger = \left( B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:} \right)^{-1} B_{\mathcal{S}_B:}^\top$. Let $(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})^{-1} = \det(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})^{-1} F$, where $F$ is the adjugate of $B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:}$ defined as $F_{ij} = (-1)^{i+j} M_{ij}$ and $M_{ij}$ is the minor of $B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:}$ formed by deleting its $i$-th row and $j$-th column. We split $\mathcal{S}_B$ into two sets $\mathcal{S}_1$ and $\mathcal{S}_2$, where $\mathcal{S}_1$ represents the support set of $GA\Theta_{ij}^\star$ and $\mathcal{S}_2 = \mathcal{S}_B \backslash \mathcal{S}_1$. We also split $\mathcal{S}_B^c$ into two sets $\mathcal{S}_1^c$ and $\mathcal{S}_2^c$ in a similar way. Our first goal is to show that IC holds with $\alpha = \mu/\gamma$ which satisfies $1/2 \leq \alpha \leq 1$. To this end, we show that

$$B_{\mathcal{S}_1^c} B_{\mathcal{S}_B:}^\dagger \, \mathrm{sign} \left( (B\Theta_{ij}^\star)_{\mathcal{S}_B:} \right) = 0, \tag{8}$$

$$\left\| B_{\mathcal{S}_2^c} B_{\mathcal{S}_B:}^\dagger \, \mathrm{sign} \left( (B\Theta_{ij}^\star)_{\mathcal{S}_B:} \right) \right\|_\infty \leq 1 - \mu/\gamma \tag{9}$$

To prove (8), it suffices to show that

$$\left[ (B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})^{-1} B_{\mathcal{S}_B:}^\top \mathrm{sign}(B_{\mathcal{S}_B:} \Theta_{ij}^\star) \right]_u = \left[ (B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})^{-1} B_{\mathcal{S}_B:}^\top \mathrm{sign}(B_{\mathcal{S}_B:} \Theta_{ij}^\star) \right]_v,$$

for every $u \neq v$ such that $\Theta_{u;ij}^\star = \Theta_{v;ij}^\star$. First, we write

$$\left[ (B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})^{-1} B_{\mathcal{S}_B:}^\top \mathrm{sign}(B_{\mathcal{S}_B:} \Theta_{ij}^\star) \right]_u = \left[ \left( B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:} \right)^{-1} \right]_{u:} B_{\mathcal{S}_B:}^\top \mathrm{sign} \left( B_{\mathcal{S}_B:} \Theta_{ij}^\star \right)$$

$$= \left[ \left( B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:} \right)^{-1} \right]_{u:} \left[ B_{\mathcal{S}_1}^\top \quad B_{\mathcal{S}_2}^\top \right] \mathrm{sign} \left( B_{\mathcal{S}_B:} \Theta_{ij}^\star \right)$$

$$= \frac{\frac{\gamma}{\mu} \sum_{k<l} (F_{uk} - F_{ul}) \mathrm{sign}(\Theta_{k;ij}^\star - \Theta_{l;ij}^\star) + \sum_k F_{uk} \mathrm{sign}(\Theta_{k;ij}^\star)}{\det(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})}$$

where the last equality is due to the definition of $F$. On the other hand, one can verify that

$$(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})_{kl} = \begin{cases} -\frac{\gamma^2}{\mu^2} \mathbb{I}\{\pi(k,l) \in \mathcal{S}_1\}, & k \neq l \\ \mathbb{I}\{\Theta_{k;ij}^\star \neq 0\} - \sum_{u \neq k} (B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})_{ku}, & k = l \end{cases} \tag{10}$$

where $\mathbb{I}(\cdot)$ is the indicator function. Before proceeding, we need the following intermediate lemma.

**Lemma 2.** *The following statements hold:*

- *$F_{uu} = F_{vv}$ for every $u, v$ such that $\Theta^\star_{u;ij} = \Theta^\star_{v;ij}$.*

- *$F_{uk} = F_{vk}$ for every $u, v, k$ such that $\Theta^\star_{u;ij} = \Theta^\star_{v;ij}$, $k \neq u$, and $k \neq v$.*

The proof of the above lemma can be found in Appendix D.3. For $u, v$ such that $\Theta^\star_{u;ij} = \Theta^\star_{v;ij}$, one can write

$$
\left[ (B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})^{-1} B_{\mathcal{S}_B:}^\top \mathrm{sign}(B_{\mathcal{S}_B:} \Theta^\star_{ij}) \right]_u - \left[ (B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})^{-1} B_{\mathcal{S}_B:}^\top \mathrm{sign}(B_{\mathcal{S}_B:} \Theta^\star_{ij}) \right]_v
$$

$$
= \frac{1}{\det(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})} \left( \frac{\gamma}{\mu} \sum_{k<l} \left( (F_{uk} - F_{vk}) - (F_{ul} - F_{vl}) \right) \right.
$$

$$
\left. \mathrm{sign}(\Theta^\star_{k;ij} - \Theta^\star_{l;ij}) + \sum_k (F_{uk} - F_{vk}) \mathrm{sign}(\Theta^\star_{k;ij}) \right)
$$

Therefore, it suffices to prove that $\sum_{k<l} \left( (F_{uk} - F_{vk}) - (F_{ul} - F_{vl}) \right) \mathrm{sign}(\Theta^\star_{k;ij} - \Theta^\star_{l;ij}) = 0$ and $\sum_k (F_{uk} - F_{vk}) \mathrm{sign}(\Theta^\star_{k;ij}) = 0$. Invoking Lemma 2, one can write

$$
\sum_{k<l} (F_{uk} - F_{vk}) \mathrm{sign}(\Theta^\star_{k;ij} - \Theta^\star_{l;ij}) = \sum_{u<l} (F_{uu} - F_{vu}) \mathrm{sign}(\Theta^\star_{u;ij} - \Theta^\star_{l;ij})
$$

$$
+ \sum_{v<l} (F_{uv} - F_{vv}) \mathrm{sign}(\Theta^\star_{v;ij} - \Theta^\star_{l;ij})
$$

$$
= (F_{uu} - F_{uv}) \left( \sum_{u<l} \mathrm{sign}(\Theta^\star_{u;ij} \right.
$$

$$
\left. - \Theta^\star_{l;ij}) - \sum_{v<l} \mathrm{sign}(\Theta^\star_{u;ij} - \Theta^\star_{l;ij}) \right)
$$

$$
= (F_{uu} - F_{uv}) \sum_{l=u+1}^{v} \mathrm{sign}(\Theta^\star_{u;ij} - \Theta^\star_{l;ij}).
$$

Thus, we have

$$
\sum_{k<l} \left( (F_{uk} - F_{vk}) - (F_{ul} - F_{vl}) \right) \mathrm{sign}(\Theta^\star_{k;ij} - \Theta^\star_{l;ij}) = (F_{uu} - F_{uv}) \left( \sum_{l=u+1}^{v} \mathrm{sign}(\Theta^\star_{u;ij} - \Theta^\star_{l;ij}) \right.
$$

$$
\left. + \sum_{k=u+1}^{v} \mathrm{sign}(\Theta^\star_{k;ij} - \Theta^\star_{u;ij}) \right) = 0.
$$

On the other hand, one can write

$$
\sum_k (F_{uk} - F_{vk}) \mathrm{sign}(\Theta^\star_{k;ij}) = (F_{uu} - F_{vu}) \mathrm{sign}(\Theta^\star_{u;ij}) + (F_{uv} - F_{vv}) \mathrm{sign}(\Theta^\star_{v;ij})
$$

$$
= (F_{uu} - F_{uv}) \mathrm{sign}(\Theta^\star_{u;ij}) - (F_{uu} - F_{uv}) \mathrm{sign}(\Theta^\star_{v;ij})
$$

$$
= 0,
$$

which completes the proof of (8).

Now, to prove (9), we derive the explicit form of $\left[\left(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:}\right)^{-1} B_{\mathcal{S}_B:}^\top \operatorname{sign}\left(B_{\mathcal{S}_B:}\Theta_{ij}^\star\right)\right]_u$, for any $u$ such that $\Theta_{u;ij}^\star = 0$. Recall that the determinant of a matrix remains unchanged after adding multiples of a column to another column. Adding all the other columns to the column $u$ of $B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:}$ changes this column to $[\mathbb{I}_{\{\Theta_{1;ij}^\star \neq 0\}}, \mathbb{I}_{\{\Theta_{2;ij}^\star \neq 0\}}, ..., \mathbb{I}_{\{\Theta_{K;ij}^\star \neq 0\}}]^\top$. Therefore, for any $\frac{\gamma}{\mu} > 0$, we have

$$\sum_{k \in S} -\frac{\gamma^2}{\mu^2} F_{ku} + |S| \frac{\gamma^2}{\mu^2} F_{uu} = \sum_{k \in S} F_{ku} = \det(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})$$

$$\implies F_{uu} = \frac{(1 + \frac{\gamma^2}{\mu^2}) \det(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})}{|S| \frac{\gamma^2}{\mu^2}}$$

On the other hand, for every $v \in S_2^c \setminus u$, one can write

$$\left[B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:}\right]_{v:} \left[(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})^{-1}\right]_{:u} = \frac{\sum_{k \in S} -\frac{\gamma^2}{\mu^2} F_{ku} + |S| \frac{\gamma^2}{\mu^2} F_{vu}}{\det(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})} = 0,$$

Recalling that $\sum_{k \in S} F_{ku} = \det(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})$, the above equality implies $F_{uv} = \frac{\det(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})}{|S|}$. Similarly, for $k \in S_2$, one can write

$$\left[B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:}\right]_{u:} \left[(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})^{-1}\right]_{:k} = \frac{\sum_{l \in S} -\frac{\gamma^2}{\mu^2} F_{lk} + |S| \frac{\gamma^2}{\mu^2} F_{uk}}{\det(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})} = 0. \tag{11}$$

which again implies $F_{uk} = \frac{\det(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})}{|S|}$. Therefore, for any $u \in S_2^c$, we have

$$F_{uk} = \begin{cases} \frac{\det(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})}{|S|}, & k \neq u \\ \frac{(1 + \frac{\gamma^2}{\mu^2}) \det(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})}{|S| \frac{\gamma^2}{\mu^2}}, & k = u \end{cases} \tag{12}$$

Now, based on the above formula, we can compute the explicit form of $\left[\left(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:}\right)^{-1} B_{\mathcal{S}_B:}^\top \operatorname{sign}\left(B_{\mathcal{S}_B:}\Theta_{ij}^\star\right)\right]_u$. First, note that

$$\sum_{k < l} (F_{uk} - F_{ul}) \operatorname{sign}(\Theta_{k;ij}^\star - \Theta_{l;ij}^\star) = \sum_{k < u} (F_{uk} - F_{uu}) \operatorname{sign}(\Theta_{k;ij}^\star - \Theta_{u;ij}^\star)$$

$$+ \sum_{u < l} (F_{uu} - F_{ul}) \operatorname{sign}(\Theta_{u;ij}^\star - \Theta_{l;ij}^\star)$$

$$= \sum_{k < u} (F_{uk} - F_{uu}) \operatorname{sign}(\Theta_{k;ij}^\star) + \sum_{u < l} (F_{ul} - F_{uu}) \operatorname{sign}(\Theta_{l;ij}^\star)$$

$$= \sum_{k \neq u} (F_{uk} - F_{uu}) \operatorname{sign}(\Theta_{k;ij}^\star)$$

$$= \sum_{k \in S} (F_{uk} - F_{uu}) \operatorname{sign}(\Theta_{k;ij}^\star)$$

28

Therefore

$$
\begin{aligned}
\left[\left(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:}\right)^{-1} B_{\mathcal{S}_B:}^\top \operatorname{sign}\left(B_{\mathcal{S}_B:}\Theta_{ij}^\star\right)\right]_u &= \frac{\frac{\gamma}{\mu} \sum_{k\in S}(F_{uk}-F_{uu})\operatorname{sign}(\Theta_{k;ij}^\star) + \sum_k F_{uk}\operatorname{sign}(\Theta_{k;ij}^\star)}{\det(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})} \\
&= \frac{(\frac{\gamma}{\mu}+1)\sum_{k\in S}F_{uk}\operatorname{sign}(\Theta_{k;ij}^\star) - \frac{\gamma}{\mu}\sum_{k\in S}F_{uu}\operatorname{sign}(\Theta_{k;ij}^\star)}{\det(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})} \\
&= \frac{(\frac{\gamma}{\mu}+1)\sum_{k\in S}\operatorname{sign}(\Theta_{k;ij}^\star)}{|S|} - \frac{\gamma}{\mu}\sum_{k\in S}\operatorname{sign}(\Theta_{k;ij}^\star)\frac{1+\frac{\gamma^2}{\mu^2}}{|S|\frac{\gamma^2}{\mu^2}} \\
&= \frac{\sum_{k\in S}\operatorname{sign}(\Theta_{k;ij}^\star)}{|S|}\left(1-\frac{\mu}{\gamma}\right) \leq 1-\frac{\mu}{\gamma}
\end{aligned}
\tag{13}
$$

where in the third equality, we used the explicit form of $F_{uk}$ in (12). Therefore, we have $\left\|B_{\mathcal{S}^c:}\left(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:}\right)^{-1}B_{\mathcal{S}_B:}^\top\operatorname{sign}\left(B_{\mathcal{S}_B:}\Theta_{ij}^\star\right)\right\|_\infty \leq |1-\mu/\gamma|$, which implies that $\alpha = \mu/\gamma$.

Next, we provide an upper and lower bound for $\kappa_{\mathrm{IC}}$. Recall that $\kappa_{\mathrm{IC}} := \left\|B_{\mathcal{S}^c:}\left(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:}\right)^{-1}B_{\mathcal{S}_B:}^\top\right\|_\infty +$ 1. Hence, we trivially have $\kappa_{\mathrm{IC}} \geq 1$. Therefore, it suffices to show that $\left\|B_{\mathcal{S}^c:}\left(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:}\right)^{-1}B_{\mathcal{S}_B:}^\top\right\|_\infty \leq$ 4. To this goal, we show that $\left\|B_{u:}\left(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:}\right)^{-1}B_{\mathcal{S}_B:}^\top\right\|_1 \leq 4$, for every $u \in \mathcal{S}_B^c$. We consider three cases:

*Case 1:* Suppose that $u = \pi(k,l)$ for some $(k,l)$ such that $\Theta_{k;ij}^\star = \Theta_{l;ij}^\star = 0$. One can write

$$
\begin{aligned}
B_{u:}\left(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:}\right)^{-1}B_{\mathcal{S}_B:}^\top &= \frac{\gamma}{\mu\det(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})}\left[F_{k1}-F_{l1},...,F_{kK}-F_{lK}\right]B_{\mathcal{S}_B:}^\top \\
&= \frac{\gamma}{\mu\det(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})} \\
&\quad\cdot[...,0,F_{kk}-F_{lk},0,...,0,F_{kl}-F_{ll},0,...]B_{\mathcal{S}_B:}^\top \\
&= \frac{\gamma}{\mu}\left[...,0,\frac{\mu^2}{\gamma^2|S|},0,...,0,\frac{-\mu^2}{\gamma^2|S|},0,...\right]B_{\mathcal{S}_B:}^\top
\end{aligned}
\tag{14}
$$

Thus $\|B_{u:}\left(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:}\right)^{-1}B_{\mathcal{S}_B:}^\top\|_1 = 2\frac{\gamma}{\mu}\sum_{u\in S}\frac{\mu^2}{\gamma^2|S|} \leq 4$.

*Case 2:* Suppose that $u = \pi(k,l)$ for some $(k,l)$ such that $\Theta_{k;ij}^\star = \Theta_{l;ij}^\star \neq 0$. Let $\bar{S} = \{r|\Theta_{r;ij}^\star \neq \Theta_{k;ij}^\star, r = 1,...,K\}$. It is easy to verify that column $k$ and $l$ of $B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:}$ are the same except for the $k$-th and $l$-th element. Subtracting column $k$ of $B_{\mathcal{S}^\top B:}$ from column $l$, one can write

$$
(|\bar{S}|\frac{\gamma^2}{\mu^2}+1)F_{kk} - (|\bar{S}|\frac{\gamma^2}{\mu^2}+1)F_{kl} = \det(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:}).
\tag{15}
$$

Therefore,

$$
F_{kk} - F_{kl} = \frac{\det(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})}{|\bar{S}|\frac{\gamma^2}{\mu^2}+1}
\tag{16}
$$

29

Combining (16) and Lemma 2, one can write

$$B_{u:} \left( B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:} \right)^{-1} B_{\mathcal{S}_B:}^\top = \frac{\gamma}{\mu \det(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})} \left[ F_{k1} - F_{l1}, ..., F_{kK} - F_{lK} \right] B_{\mathcal{S}_B:}^\top$$

$$= \frac{\gamma}{\mu} \left[ ..., 0, \frac{1}{|\bar{S}|\frac{\gamma^2}{\mu^2} + 1}, 0, , ..., 0, -\frac{1}{|\bar{S}|\frac{\gamma^2}{\mu^2} + 1}, 0, ... \right] B_{\mathcal{S}_B:}^\top \tag{17}$$

Therefore, again we have $\|B_{u:} \left( B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:} \right)^{-1} B_{\mathcal{S}_B:}^\top\|_1 \le 4$.

*Case 3:* Suppose that $u = \frac{K(K-1)}{2} + k$ for some $k$ and $\Theta_{k;ij}^\star = 0$. One can write

$$B_{u:} \left( B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:} \right)^{-1} B_{\mathcal{S}_B:}^\top = \frac{1}{\det(B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:})} \left[ F_{k1}, F_{k2}, ..., F_{kK} \right] B_{\mathcal{S}_B:}^\top$$

$$= \frac{\gamma}{\mu} \left[ ..., \frac{1}{|S|}, \frac{1 + \gamma^2/\mu^2}{|S|(\gamma^2/\mu^2)}, \frac{1}{|S|}, ... \right] B_{\mathcal{S}_B:}^\top \tag{18}$$

Thus $\|B_{u:} \left( B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:} \right)^{-1} B_{\mathcal{S}_B:}^\top\|_1 = 1 + \frac{\mu}{\gamma} \le 3$. The completes the proof. $\square$

# D  Additional Proofs

## D.1  Proof of Proposition 2

For the purpose of proof, we rewrite the optimization problem (2) with $q = 2$ as an instance of the Lasso problem [83].

$$\|\Theta_{ij} - \tilde{F}_{ij}^*\|_2^2 + \gamma\|GA\Theta_{ij}\|_2^2 = (\Theta_{ij} - \tilde{F}_{ij}^*)^\top (\Theta_{ij} - \tilde{F}_{ij}^*) + \gamma\Theta_{ij}^\top A^\top G^\top GA\Theta_{ij}$$

$$= \Theta_{ij}^\top (I + \gamma A^\top G^\top GA)\Theta_{ij} - 2\Theta_{ij}^\top \tilde{F}_{ij}^* + (\tilde{F}_{ij}^*)^\top \tilde{F}_{ij}^* \tag{19}$$

Since $I + \gamma A^\top G^\top GA$ is a strictly diagonally dominant symmetric matrix, it has a unique Cholesky decomposition $I + \gamma A^\top G^\top GA = C^\top C$. Therefore, one can write

$$\|\Theta_{ij} - \tilde{F}_{ij}^*\|_2^2 + \gamma\|GA\Theta_{ij}\|_2^2 = \|(C^\top)^{-1}\tilde{F}_{ij}^* - C\Theta_{ij}\|_2^2 - (\tilde{F}_{ij}^*)^\top C^{-1}(C^\top)^{-1}\tilde{F}_{ij}^* + (\tilde{F}_{ij}^*)^\top \tilde{F}_{ij}^* \tag{20}$$

Therefore, problem (Elem-$(i, j, q)$) is equivalent to:

$$\min \quad \|y - C\Theta_{ij}\|_2^2 + \mu\|\Theta_{ij}\|_1, \tag{21}$$

where $y = (C^\top)^{-1}\tilde{F}_{ij}^*$. Note that (21) is an instance of Lasso with the observation model $y = C\Theta_{ij}^\star + w$, where $w = (C^\top)^{-1}\tilde{F}_{ij}^* - C\Theta_{ij}^\star$ is the noise vector.

The above equivalence will allow us to invoke the exact recovery guarantee of Lasso in our setting. For any fixed $(i, j)$, let $S \subset \{1, 2, ..., K\}$ with $|S| = s$ be the support of $\Theta_{ij}^\star$ satisfying $\Theta_{k;ij}^\star \neq 0$ for all $k \in S$. Moreover, let $S^c = \{1, 2, ..., K\}\backslash S$. Without loss of generality and to streamline our presentation, we assume that $S = \{1, 2, \ldots, s\}$. At the crux of the exact recovery guarantees based on Lasso lies the classical notion of *mutual incoherency* [92, 19, 83].

**Assumption 6** (Mutual Incoherency). *There exists some $\alpha \in [0, 1)$ such that*

$$\max_{j \in S^c} \left\| \left( C_S^{\mathrm{T}} C_S \right)^{-1} C_S^{\mathrm{T}} C_j \right\|_1 \leq \alpha$$

Mutual incoherency entails that the effect of the columns of $C$ corresponding to "unimportant" (zero) elements of $\Theta_{ij}^\star$ on the remaining columns is small. Although mutual incoherency cannot be guaranteed for general choices of $C$ and $S$, our next lemma shows that it is indeed satisfied for our problem, provided that $\gamma$ is sufficiently small.

**Lemma 3.** *For $C^\top C = I + \gamma A^\top G^\top G A$ and any $S \subseteq \{1, 2, \ldots, K\}$, the mutual incoherency holds with $\alpha = 1/2$, provided that $\gamma < 1/(2K \|W\|_{\max})$.*

*Proof.* It is easy to verify that

$$
A^\top G^\top G A = \begin{bmatrix}
\sum_{i \neq 1} W_{1,i} & -W_{1,2} & \cdots & -W_{1,K} \\
-W_{2,1} & \sum_{i \neq 2} W_{2,i} & \cdots & -W_{2,K} \\
& & \cdots & \\
-W_{K,1} & -W_{K,2} & \cdots & \sum_{i \neq K} W_{K,i}
\end{bmatrix}
\tag{22}
$$

Since $S \cap S^c = \emptyset$, the elements of $C_S^\top C_{S^c}$ can only include the off-diagonal elements of $C^\top C$, which in turn correspond to the off-diagonal entries of $A^\top G^\top G A$ defined as (22). Therefore, we have

$$\|C_S^\top C_j\|_\infty \leq \gamma \|W\|_{\max}, \quad j \in S^c. \tag{23}$$

On the other hand, since $C_S^\top C_S$ is strictly diagonally dominant, its inverse matrix satisfies

$$
\begin{aligned}
\|(C_S^\top C_S)^{-1}\|_\infty &\leq \frac{1}{\min_i \left\{ \left| (C_S^\top C_S)_{i,i} \right| - \sum_{j \neq i} \left| (C_S^\top C_S)_{i,j} \right| \right\}} \\
&= \frac{1}{\min_i \left\{ 1 + \gamma \sum_{j \neq i} |W_{i,j}| - \gamma \sum_{j \in S \setminus i} |W_{i,j}| \right\}} \\
&= \frac{1}{\min_i \left\{ 1 + \gamma \sum_{j \in S^c \setminus i} |W_{i,j}| \right\}} \leq 1
\end{aligned}
\tag{24}
$$

where the first inequality is due to [81, Theorem 1]. Combined with (23), we have

$$
\begin{aligned}
\|(C_S^\top C_S)^{-1} C_S^\top C_j\|_1 &\leq K \|(C_S^\top C_S)^{-1}\|_\infty \|C_S^\top C_j\|_\infty \\
&\leq K \gamma \|W\|_{\max} \|(C_S^\top C_S)^{-1}\|_\infty \\
&\leq K \gamma \|W\|_{\max} < 1/2,
\end{aligned}
\tag{25}
$$

which completes the proof.

$\square$

Given the mutual incoherency condition, we follow the so-called *primal-dual witness (PDW)* approach introduced by Wainwright [83] to prove Proposition 2. To this goal, first we delineate the optimality conditions for (21). Given a convex function $f : \mathbb{R}^K \to \mathbb{R}$, we say that $z \in \mathbb{R}^K$ is a subgradient of $f$ at $\Theta_{ij}$, denoted by $z \in \partial f(\Theta_{ij})$, if we have

$$f(\Theta_{ij} + \delta) \geq f(\Theta_{ij}) + \langle z, \delta \rangle \text{ for all } \delta \in \mathbb{R}^K.$$

When $f(\Theta_{ij}) = \|\Theta_{ij}\|_1$, it can be seen that $z \in \partial \|\Theta_{ij}\|_1$ if and only if $z_k = \text{sign}(\Theta_{k;ij})$ for all $k = 1, 2, ..., K$. For (21), we say that a pair $(\widehat{\Theta}_{ij}, \hat{z})$ is *primal-dual optimal* if

$$\hat{z} \in \partial \|\Theta_{ij}\|_1, \quad 2C^\top(C\widehat{\Theta}_{ij} - y) + \mu\hat{z} = 0 \tag{26}$$

Evidently, if $(\widehat{\Theta}_{ij}, \hat{z})$ is primal-dual optimal, then $\widehat{\Theta}_{ij}$ is the minimizer of (21). Given this optimality condition, the Primal-dual witness is constructed as follows.

---

**PDW construction:** PDW construction has the following steps:

1. Set $\widehat{\Theta}_{S^c;ij} = 0$

2. Determine $\left( \widehat{\Theta}_{S;ij}, \widehat{z}_S \right) \in \mathbb{R}^s \times \mathbb{R}^s$ by solving the oracle subproblem

$$\widehat{\Theta}_{S;ij} \in \arg\min_{\Theta_{S;ij} \in \mathbb{R}^s} \{ \underbrace{\|y - C_S\Theta_{S;ij}\|_2^2}_{=:f(\Theta_{S;ij})} + \mu\|\Theta_{S;ij}\|_1, \tag{27}$$

   then choose $\widehat{z}_S \in \partial \left\| \widehat{\Theta}_{S;ij} \right\|_1$ such that $\nabla f(\Theta_{S;ij})|_{\Theta_{S;ij}=\widehat{\Theta}_{S;ij}} + \mu\widehat{z}_S = 0$

3. Solve for $\widehat{z}_{S^c} \in \mathbb{R}^{K-s}$ via the zero-subgradient equation (26), and check whether or not the strict dual feasibility condition $\|\widehat{z}_{S^c}\|_\infty < 1$ holds.

---

Note that the vector $\widehat{\Theta}_{S^c;ij}$ is determined in Step 1, whereas the remaining three subvectors $\widehat{\Theta}_{S;ij}$, $\widehat{z}_S$, and $\widehat{z}_{S^c}$ are determined in Steps 2 and 3. By construction, $\widehat{\Theta}_{S^c;ij} = 0$ and the subvectors $\widehat{\Theta}_{S;ij}$, $\widehat{z}_S$ and $\widehat{z}_{S^c}$ satisfy the zero-subgradient condition (26). Therefore, we have

$$\begin{bmatrix} C_S^\top C_S & C_S^\top C_{S^c} \\ C_{S^c}^\top C_S & C_{S^c}^\top C_{S^c} \end{bmatrix} \begin{bmatrix} \widehat{\Theta}_{S;ij} - \Theta_{S;ij}^\star \\ 0 \end{bmatrix} + \frac{\mu}{2} \begin{bmatrix} \widehat{z}_S \\ \widehat{z}_{S^c} \end{bmatrix} = \begin{bmatrix} C_S^\top w \\ C_{S^c}^\top w \end{bmatrix} \tag{28}$$

We say that "PDW construction succeeds" if the vector $\widehat{z}_{S^c}$ constructed in step 3 satisfies the strict dual feasibility condition. The following result shows that the success of PDW construction implies that $(\widehat{\Theta}_{S;ij}, 0) \in \mathbb{R}^K$ is the unique solution of the Lasso problem.

**Lemma 4** (Lemma 7.23 of [83])**.** *The success of the PDW construction implies that the vector* $(\widehat{\Theta}_{S;ij}, 0) \in \mathbb{R}^K$ *is the unique optimal solution of the Lasso problem (21).*

*Proof of Proposition 2.* To apply Lemma 4, it suffices to show that the vector $\widehat{z}_{S^c} \in \mathbb{R}^{K-s}$ constructed in Step 3 satisfies the strict dual feasibility condition. Using the zero-subgradient condition (28), we have

$$\widehat{z}_{S^c} = -\frac{2}{\mu} C_{S^c}^\top C_S (\widehat{\Theta}_{S;ij} - \Theta_{S;ij}^\star) + \frac{2}{\mu} C_{S^c}^\top w. \tag{29}$$

On the other hand, (28) implies that

$$\widehat{\Theta}_{S;ij} - \Theta_{S;ij}^\star = (C_S^\top C_S)^{-1} C_S^\top w - \frac{\mu}{2} (C_S^\top C_S)^{-1} \widehat{z}_S \tag{30}$$

Substituting this expression back into (29) yields

$$\widehat{z}_{S^c} = C_{S^c}^\top C_S (C_S^\top C_S)^{-1} \widehat{z}_S + C_{S^c}^\top [I - C_S (C_S^\top C_S)^{-1} C_S^\top] \frac{2w}{\mu} \tag{31}$$

Due to Lemma 3, we have $\|C_{S^c}^\top C_S (C_S^\top C_S)^{-1} \widehat{z}_S\|_\infty \le 1/2$. Therefore, if the following inequalities are satisfied

$$\left\| C_{S^c}^\top [I - C_S (C_S^\top C_S)^{-1} C_S^\top] \frac{2w}{\mu} \right\|_\infty \le \frac{1}{2}$$

$$\implies 4 \| C_{S^c}^\top w - C_{S^c}^\top C_S \left( C_S^\top C_S \right)^{-1} C_S^\top w \|_\infty \le \mu,$$

then we conclude that $\widehat{z}_{S^c} < 1$, which establishes the strict dual feasibility condition. On the other hand, one can write

$$4 \left\| C_{S^c}^\top w - C_{S^c}^\top C_S \left( C_S^\top C_S \right)^{-1} C_S^\top w \right\|_\infty \le 4 \| C_{S^c}^\top w \|_\infty + 4 \left\| C_{S^c}^\top C_S \left( C_S^\top C_S \right)^{-1} C_S^\top w \right\|_\infty$$

$$\le 4 \| C^\top w \|_\infty + 4K\gamma \|W\|_{\max} \|C^\top w\|_\infty$$

$$= 4(1 + K\gamma \|W\|_{\max}) \|C^\top w\|_\infty \le 6 \|C^\top w\|_\infty$$

where the second inequality follows from Lemma 3. We also have $C^\top w = C^\top y - C^\top C \Theta_{ij}^\star = \tilde{F}_{ij}^* - \Theta_{ij}^\star - \gamma A^\top G^\top G A \Theta_{ij}^\star$. Therefore, upon choosing

$$\mu \ge 6 \|\Theta_{ij}^\star - \tilde{F}_{ij}^*\|_\infty + 6\gamma K \|W\|_{\max} D$$

$$\ge 6 \|\tilde{F}_{ij}^* - \Theta_{ij}^\star - \gamma A^\top G^\top G A \Theta_{ij}^\star\|_\infty,$$

we establish the strict dual feasibility condition. This implies that $\widehat{\Theta}_{ij} = (\widehat{\Theta}_{S;ij}, 0)$ is the unique solution of (21). Therefore, we have $\operatorname{supp}(\widehat{\Theta}_{ij}) \subseteq \operatorname{supp}(\Theta_{ij}^\star)$ and

$$\begin{aligned} \|\widehat{\Theta}_{ij} - \Theta_{ij}^\star\|_\infty &= \|\widehat{\Theta}_{S;ij} - \Theta_{S;ij}^\star\|_\infty \\ &\le \|(C_S^\top C_S)^{-1} C_S^\top w\|_\infty + \frac{\mu}{2} \|(C_S^\top C_S)^{-1}\|_\infty \\ &\le \|(C_S^\top C_S)^{-1}\|_\infty \left( \|C^\top w\|_\infty + \frac{\mu}{2} \right) \\ &\le \frac{2\mu}{3 + 3\gamma \min_i \{\sum_{j \in S^c \setminus i} |W_{i,j}|\}} \le \frac{2\mu}{3} \end{aligned} \tag{32}$$

Now, it suffices to show that $\operatorname{supp}(\Theta_{ij}^\star) \subseteq \operatorname{supp}(\widehat{\Theta}_{ij})$. Suppose that $\Theta_{k;ij}^\star \ne 0$ for some $k$. Then, one can write

$$|\widehat{\Theta}_{k;ij}| \ge |\Theta_{k;ij}^\star| - |\widehat{\Theta}_{k;ij} - \Theta_{k;ij}^\star|$$

$$\ge \Theta_{\min;ij} - \frac{2\mu}{3} > 0.$$

Therefore, we have $\text{supp}(\Theta^\star_{ij}) \subseteq \text{supp}(\widehat{\Theta}_{ij})$ which shows that $\text{supp}(\widehat{\Theta}_{ij}) = \text{supp}(\Theta^\star_{ij})$. This completes the proof. $\qquad\square$

## D.2 Proof of Proposition 3

To prove Proposition 3, we first provide the following lemma adapted from [58].

**Lemma 5** (Corollary 4.2 of [58]). *Suppose that the irrepresentability assumption holds and*

$$\|\tilde{F}^*_{ij} - \Theta^\star_{ij}\|_\infty < \frac{\alpha}{8\kappa_{\mathrm{IC}}}\mu.$$

*Then, the optimal solution to* (3) *is unique and satisfies the following properties:*

- *(Estimation error):* $\left\|\widehat{\Theta}_{ij} - \Theta^\star_{ij}\right\|_2 \leq 2\left(\kappa_\rho + \frac{\alpha}{4}\frac{\kappa_\varrho}{\kappa_{\mathrm{IC}}}\right)\mu$

- *(Sparsistency):* $(B\widehat{\Theta}_{ij})_{\mathcal{S}_B:^c} = 0,$

*where*

$$
\begin{aligned}
\kappa_\rho &= \sup_\theta\left\{\|B\theta\|_1 \mid \|\theta\|_2 = 1, (B\theta)_{\mathcal{S}_B:^c} = 0\right\} \\
\kappa_\varrho &= \sup_\theta\left\{\|\theta\|_1 \mid \|\theta\|_2 = 1, (B\theta)_{\mathcal{S}_B:^c} = 0\right\} \\
\kappa_{\mathrm{IC}} &= \sup_{\|z\|_\infty \leq 1}\left\|B_{\mathcal{S}_B:^c}B^\dagger_{\mathcal{S}_B:}z_{\mathcal{S}_B:}\right\|_\infty + \|z_{\mathcal{S}_B:^c}\|_\infty
\end{aligned}
\tag{33}
$$

Based on the above lemma, we proceed with the proof of Proposition 3

*Proof of Theorem 3.* Based on our assumptions, Lemma 5 can be invoked to show that $\text{supp}(\widehat{\Theta}_{ij}) \subseteq \text{supp}(\Theta^\star_{ij})$ and $\text{supp}(\widehat{\Theta}_{k;ij} - \widehat{\Theta}_{l;ij}) \subseteq \text{supp}(\Theta^\star_{k;ij} - \Theta^\star_{l;ij})$. Now, it remains to prove the upper bound on the estimation error, as well as $\text{supp}(\Theta^\star_{ij}) \subseteq \text{supp}(\widehat{\Theta}_{ij})$ and $\text{supp}(\Theta^\star_{k;ij} - \Theta^\star_{l;ij}) \subseteq \text{supp}(\widehat{\Theta}_{k;ij} - \widehat{\Theta}_{l;ij})$. First, it is easy to verify that $\kappa_\rho \leq \sqrt{2\|W\|_{\max}D_0} + \sqrt{S_0}$ and $\kappa_\varrho \leq \sqrt{S_0}$. Moreover, by setting $z_i = 1$ for some $i \in \mathcal{S}_B^c$, one can easily verify that $\kappa_{IC} \geq 1$. Therefore, according to Lemma 5, we have

$$
\begin{aligned}
\left\|\widehat{\Theta}_{ij} - \Theta^\star_{ij}\right\|_2 &\leq 2\left(\kappa_\rho + \frac{\alpha}{4}\frac{\kappa_\varrho}{\kappa_{\mathrm{IC}}}\right)\mu \\
&\leq 2\left(\sqrt{2\|W\|_{\max}D_0} + \sqrt{S_0} + \frac{\sqrt{S_0}}{4}\right)\mu \\
&\leq 2\left(\sqrt{2\|W\|_{\max}D_0} + 1.25\sqrt{S_0}\right)\mu,
\end{aligned}
$$

where in the second inequality, we used $\kappa_\rho \leq \sqrt{2\|W\|_{\max}D_0} + \sqrt{S_0}$, $\kappa_\varrho \leq \sqrt{S_0}$, $\alpha \leq 1$, and $\kappa_{IC} \geq 1$.

Now, it suffices to show that $\text{supp}(\Theta^\star_{ij}) \subseteq \text{supp}(\widehat{\Theta}_{ij})$ and $\text{supp}(\Theta^\star_{k;ij} - \Theta^\star_{l;ij}) \subseteq \text{supp}(\widehat{\Theta}_{k;ij} - \widehat{\Theta}_{l;ij})$. Suppose that $\Theta^\star_{k;ij} \neq 0$ for some $k$. Then, one can write

$$
\begin{aligned}
|\widehat{\Theta}_{k;ij}| &\geq |\Theta^\star_{k;ij}| - |\widehat{\Theta}_{k;ij} - \Theta^\star_{k;ij}| \\
&\geq \Theta_{\min;ij} - 2\left(\sqrt{2\|W\|_{\max}D_0} + 1.25\sqrt{S_0}\right)\mu > 0,
\end{aligned}
$$

34

where the last inequality follows from the assumption $\mu < \Theta_{\min;ij} / \left( 2 \left( \sqrt{2 \|W\|_{\max} D_0} + 1.25\sqrt{S_0} \right) \right)$. This implies that $\widehat{\Theta}_{k;ij} \neq 0$, and hence, $\mathrm{supp}(\Theta_{ij}^\star) \subseteq \mathrm{supp}(\widehat{\Theta}_{ij})$. Similarly, suppose $\Theta_{k;ij}^\star - \Theta_{l;ij}^\star \neq 0$ for some $(k, l)$. One can write

$$
\begin{aligned}
|\widehat{\Theta}_{k;ij} - \widehat{\Theta}_{l;ij}| &\geq |\Theta_{k;ij}^\star - \Theta_{l;ij}^\star| - |\widehat{\Theta}_{k;ij} - \Theta_{k;ij}^\star| - |\widehat{\Theta}_{l;ij} - \Theta_{l;ij}^\star| \\
&\geq \Delta\Theta_{\min;ij} - 4\left( \sqrt{2\|W\|_{\max} D_0} + 1.25\sqrt{S_0} \right)\mu \\
&> 0,
\end{aligned}
$$

where the first inequality is due to triangle inequality and the last inequality follows from $\mu < \Delta\Theta_{\min;ij} / \left( 4 \left( \sqrt{2\|W\|_{\max} D_0} + 1.25\sqrt{S_0} \right) \right)$. This in turn implies that $\mathrm{supp}(\Theta_{k;ij}^\star - \Theta_{l;ij}^\star) \subseteq \mathrm{supp}(\widehat{\Theta}_{k;ij} - \widehat{\Theta}_{l;ij})$, which completes the proof. $\square$

### D.3 Proof of Lemma 2

Using the standard properties of adjucate matrices, one can obtain $F_{vv}$ from $F_{uu}$ after the following steps:

1. Move column $v$ of $M_{uu}$ to position $u$, so that it becomes the $u$-th column of $M_{uu}$.

2. Move row $v$ of $M_{uu}$ to position $u$, so that it becomes the $u$-th row of $M_{uu}$.

To be more specific, the column and row indices of $M_{uu}$ change from $\{1, 2, ..., u-1, u+1, ..., v-1, v, v+1, ..., K\}$ to $\{1, 2, ..., u-1, v, u+1, ..., v-1, v+1, ..., K\}$. Since the column/row indices of $M_{vv}$ are $\{1, 2, ..., u-1, u, u+1, ..., v-1, v+1, ..., K\}$, we only need to show that $M_{uu}$ and $M_{vv}$ are the same at the $u$-th column and $u$-th row. Moreover, due to symmetry, it suffices to show that $M_{uu}$ and $M_{vv}$ are the same at the $u$-th column. Now, the $u$-th column of $M_{uu}$ is $\left( B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:} \right)_{\{1,...,u-1,v,u+1,...,v-1,v+1,...,K\},v}$ and the $u$-th column of $M_{vv}$ is $\left( B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:} \right)_{\{1,...,K\}\backslash v, u}$. From the structure of $B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:}$ and the fact that $\Theta_{u;ij}^\star = \Theta_{v;ij}^\star$, we have $\left( B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:} \right)_{\{1,...u,...,v,...,K\},u} = \left( B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:} \right)_{\{1,...v,...u,...,K\},v}$. Thus one can write

$$
\begin{aligned}
\left( B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:} \right)_{\{1,...,u-1,v,u+1,...,v-1,v+1,...,K\},v} &= \left( B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:} \right)_{\{1,...u-1,u,u+1,...,v-1,v+1,...,K\},u} \\
&= \left( B_{\mathcal{S}_B:}^\top B_{\mathcal{S}_B:} \right)_{\{1,...,K\}\backslash v, u}
\end{aligned}
\tag{34}
$$

Therefore, $M_{vv}$ can be obtained from $M_{uu}$ using the above two steps. This implies that $M_{uu}(-1)^{2|v-u-1|} = M_{vv}$ which in turn leads to $M_{uu} = M_{vv}$. On the other hand, due to the definition of $F_{uu}$ and $F_{vv}$, we have $F_{uu} = (-1)^{2u} M_{uu} = (-1)^{2v} M_{vv} = F_{vv}$. The second statement can be proved in a similar fashion. The details are omitted for brevity. $\square$

## References

[1] Wajd N Al-Holou, Hanxiao Wang, Visweswaran Ravikumar, Morgan Oneka, Roel Verhaak, Hoon Kim, Drew Pratt, Sandra Camelo-Piragua, Corey Speers, Todd Hollon, et al. Subclonal evolution and expansion of thy1-positive cells is associated with recurrence in glioblastoma. *bioRxiv*, 2021.

[2] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

[3] Michaela Asp, Joseph Bergenståhle, and Joakim Lundeberg. Spatially resolved transcriptomes—next generation tools for tissue exploration. *BioEssays*, 42(10):1900221, 2020.

[4] Onureena Banerjee, Laurent El Ghaoui, and Alexandre d'Aspremont. Model selection through sparse maximum likelihood estimation, 2007.

[5] Albert-Laszlo Barabasi and Zoltan N Oltvai. Network biology: understanding the cell's functional organization. *Nature reviews genetics*, 5(2):101–113, 2004.

[6] Cornelia Bargmann, William Newsome, A Anderson, E Brown, Karl Deisseroth, J Donoghue, Peter MacLeish, E Marder, R Normann, and J Sanes. Brain 2025: a scientific vision. *Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Working Group Report to the Advisory Committee to the Director, NIH*, 2014.

[7] Dalia Barkley, Anjali Rao, Maayan Pour, Gustavo S França, and Itai Yanai. Cancer cell states and emergent properties of the dynamic tumor system. *Genome Research*, 31(10):1719–1727, 2021.

[8] Danielle S Bassett and Edward T Bullmore. Human brain networks in health and disease. *Current Opinion in Neurology*, 22(4):340, 2009.

[9] Riccardo Bazzoni and Angela Bentivegna. Role of notch signaling pathway in glioblastoma pathogenesis. *Cancers*, 11(3):292, 2019.

[10] Winston R Becker, Stephanie A Nevins, Derek C Chen, Roxanne Chiu, Aaron Horning, Rozelle Laquindanum, Meredith Mills, Hassan Chaib, Uri Ladabaum, Teri Longacre, et al. Single-cell analyses reveal a continuum of cell state and composition changes in the malignant transformation of polyps to colorectal cancer. *bioRxiv*, 2021.

[11] Michele Benzi and Valeria Simoncini. Decay bounds for functions of hermitian matrices with banded or kronecker structure. *SIAM Journal on Matrix Analysis and Applications*, 36(3): 1263–1282, 2015. doi: 10.1137/151006159.

[12] Razouk Benzi, Michele and Nader. Decay bounds and algorithms for approximating functions of sparse matrices. *ETNA. Electronic Transactions on Numerical Analysis [electronic only]*, 28: 16–39, 2007.

[13] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[14] Serdar Bozdag, Aiguo Li, Mehmet Baysan, and Howard A Fine. Master regulators, regulatory networks, and pathways of glioblastoma subtypes. *Cancer informatics*, 13:CIN–S14027, 2014.

[15] John H Bushweller. Targeting transcription factors in cancer—from undruggable to reality. *Nature Reviews Cancer*, 19(11):611–624, 2019.

[16] Dylan M Cable, Evan Murray, Luli S Zou, Aleksandrina Goeva, Evan Z Macosko, Fei Chen, and Rafael A Irizarry. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nature Biotechnology*, pages 1–10, 2021.

[17] Tony Cai, Weidong Liu, and Xi Luo. A constrained l1 minimization approach to sparse precision matrix estimation, 2011.

[18] Tadeusz Caliński and Jerzy Harabasz. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27, 1974.

[19] Emmanuel Candes and Justin Romberg. Sparsity and incoherence in compressive sampling. *Inverse problems*, 23(3):969, 2007.

[20] Carlos Carmona-Fontaine, Vanni Bucci, Leila Akkari, Maxime Deforet, Johanna A Joyce, and Joao B Xavier. Emergence of spatial structure in the tumor microenvironment due to the warburg effect. *Proceedings of the National Academy of Sciences*, 110(48):19402–19407, 2013.

[21] Carlos Carmona-Fontaine, Maxime Deforet, Leila Akkari, Craig B Thompson, Johanna A Joyce, and Joao B Xavier. Metabolic origins of spatial organization in the tumor microenvironment. *Proceedings of the National Academy of Sciences*, 114(11):2934–2939, 2017.

[22] Quan Cheng, Jing Li, Fan Fan, Hui Cao, Zi-Yu Dai, Ze-Yu Wang, and Song-Shan Feng. Identification and analysis of glioblastoma biomarkers based on single cell sequencing. *Frontiers in bioengineering and biotechnology*, 8:167, 2020.

[23] Tammy MK Cheng, Sakshi Gulati, Rudi Agius, and Paul A Bates. Understanding cancer mechanisms through network dynamics. *Briefings in functional genomics*, 11(6):543–560, 2012.

[24] Antimo Cutone, Luigi Rosa, Giusi Ianiro, Maria Stefania Lepanto, Maria Carmela Bonaccorsi di Patti, Piera Valenti, and Giovanni Musci. Lactoferrin's anti-cancer properties: Safety, selectivity, and wide range of action. *Biomolecules*, 10(3):456, 2020.

[25] Patrick Danaher, Pei Wang, and Daniela M. Witten. The joint graphical lasso for inverse covariance estimation across multiple classes, 2012.

[26] Spyros Darmanis, Steven A Sloan, Derek Croote, Marco Mignardi, Sophia Chernikova, Peyman Samghababi, Ye Zhang, Norma Neff, Mark Kowarsky, Christine Caneda, et al. Single-cell rna-seq analysis of infiltrating neoplastic cells at the migrating front of human glioblastoma. *Cell reports*, 21(5):1399–1410, 2017.

[27] Stephen Demko, William F. Moss, and Philip W. Smith. Decay rates for inverses of band matrices. *Mathematics of Computation*, 43(168):491–499, 1984. ISSN 00255718, 10886842.

[28] Steffen Dettling, Slava Stamova, Rolf Warta, Martina Schnölzer, Carmen Rapp, Anchana Rathinasamy, David Reuss, Kolja Pocha, Saskia Roesch, Christine Jungk, et al. Identification of crkii, cfl1, cntn1, nme2, and tkt as novel and frequent t-cell targets in human idh-mutant glioma. *Clinical Cancer Research*, 24(12):2951–2962, 2018.

[29] Donglei Du. Social network analysis: Centrality measures. *University of New Brunswick*, 2019.

[30] W Du and O Elemento. Cancer systems biology: embracing complexity to develop better anticancer therapeutic strategies. *Oncogene*, 34(25):3215–3225, 2015.

[31] Yuhui Du, Zening Fu, and Vince D Calhoun. Classification and prediction of brain disorders using functional connectivity: promising but challenging. *Frontiers in neuroscience*, 12:525, 2018.

[32] Salar Fattahi and Andres Gomez. Scalable inference of sparsely-changing markov random fields with strong statistical guarantees. *CoRR*, abs/2102.03585, 2021.

[33] Salar Fattahi and Andres Gomez. Scalable inference of sparsely-changing gaussian markov random fields. *Advances in Neural Information Processing Systems*, 34, 2021.

[34] Salar Fattahi and Somayeh Sojoudi. Closed-form solution and sparsity path for inverse covariance estimation problem. In *2018 Annual American Control Conference (ACC)*, pages 410–417. IEEE, 2018.

[35] Salar Fattahi and Somayeh Sojoudi. Graphical lasso and thresholding: Equivalence and closed-form solutions. *The Journal of Machine Learning Research*, 20(1):364–407, 2019.

[36] Salar Fattahi, Richard Y Zhang, and Somayeh Sojoudi. Sparse inverse covariance estimation for chordal structures. In *2018 European Control Conference (ECC)*, pages 837–844. IEEE, 2018.

[37] Salar Fattahi, Richard Y Zhang, and Somayeh Sojoudi. Linear-time algorithm for learning large-scale sparse graphical models. *IEEE access*, 7:12658–12672, 2019.

[38] Rina Foygel and Mathias Drton. Extended bayesian information criteria for gaussian graphical models, 2010.

[39] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the lasso, 2007.

[40] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

[41] Kristjan Greenewald, Seyoung Park, Shuheng Zhou, and Alexander Giessing. Time-dependent spatially varying graphical models, with application to brain fmri data analysis, 2017.

[42] Sarah Maddox Groves, Abbie Ireland, Qi Liu, Alan J Simmons, Ken Lau, Wade T Iams, Darren Tyson, Christine M Lovly, Trudy G Oliver, and Vito Quaranta. Cancer hallmarks define a continuum of plastic cell states between small cell lung cancer archetypes. *BioRxiv*, 2021.

[43] Jian Guo, Elizaveta Levina, George Michailidis, and Ji Zhu. Joint estimation of multiple graphical models. *Biometrika*, 98:1–15, 03 2011. doi: 10.1093/biomet/asq060.

[44] David Hallac, Youngsuk Park, Stephen Boyd, and Jure Leskovec. Network inference via the time-varying graphical lasso, 2017.

[45] Dolores Hambardzumyan and Gabriele Bergers. Glioblastoma: defining tumor niches. *Trends in cancer*, 1(4):252–265, 2015.

[46] Douglas Hanahan and Robert A Weinberg. Hallmarks of cancer: the next generation. *cell*, 144 (5):646–674, 2011.

[47] Farina Hanif, Kanza Muzaffar, Kahkashan Perveen, Saima M Malhi, and Shabana U Simjee. Glioblastoma multiforme: a review of its epidemiology and pathogenesis through clinical presentation and treatment. *Asian Pacific journal of cancer prevention: APJCP*, 18(1):3, 2017.

[48] Yuhan Hao, Stephanie Hao, Erica Andersen-Nissen, William M Mauck III, Shiwei Zheng, Andrew Butler, Maddie J Lee, Aaron J Wilk, Charlotte Darby, Michael Zager, et al. Integrated analysis of multimodal single-cell data. *Cell*, 184(13):3573–3587, 2021.

[49] Jean Hausser and Uri Alon. Tumour heterogeneity and the evolutionary trade-offs of cancer. *Nature Reviews Cancer*, 20(4):247–257, 2020.

[50] Shuai Huang, Jing Li, Liang Sun, Jieping Ye, Adam Fleisher, Teresa Wu, Kewei Chen, and Eric Reiman. Learning brain connectivity of alzheimer's disease by sparse inverse covariance estimation. *NeuroImage*, 50(3):935–949, 2010.

[51] Erik Jung, Matthias Osswald, Miriam Ratliff, Helin Dogan, Ruifan Xie, Sophie Weil, Dirk C Hoffmann, Felix T Kurz, Tobias Kessler, Sabine Heiland, et al. Tumor cell plasticity, heterogeneity, and resistance in crucial microenvironmental niches in glioma. *Nature communications*, 12(1):1–16, 2021.

[52] D. Kershaw. Inequalities on the elements of the inverse of a certain tridiagonal matrix. *Mathematics of Computation*, 24:155–158, 1970.

[53] Junghi Kim and Wei Pan. Highly adaptive tests for group differences in brain functional connectivity. *NeuroImage: Clinical*, 9:625–639, 2015.

[54] Jennifer Kosty, Fanghui Lu, Robert Kupp, Shwetal Mehta, and Q Richard Lu. Harnessing olig2 function in tumorigenicity and plasticity to target malignant gliomas. *Cell Cycle*, 16(18): 1654–1660, 2017.

[55] Danai Koutra, Joshua T Vogelstein, and Christos Faloutsos. Deltacon: A principled massive-graph similarity function. In *Proceedings of the 2013 SIAM international conference on data mining*, pages 162–170. SIAM, 2013.

[56] Saran Kumar, Husni Sharife, Tirzah Kreisel, Maxim Mogilevsky, Libat Bar-Lev, Myriam Grunewald, Elina Aizenshtein, Rotem Karni, Iddo Paldor, Tomer Shlomi, et al. Intra-tumoral metabolic zonation and resultant phenotypic diversification are dictated by blood vessel proximity. *Cell metabolism*, 30(1):201–211, 2019.

[57] Justin D Lathia, John M Heddleston, Monica Venere, and Jeremy N Rich. Deadly teamwork: neural cancer stem cells and the tumor microenvironment. *Cell stem cell*, 8(5):482–485, 2011.

[58] Jason D. Lee, Yuekai Sun, and Jonathan E. Taylor. On model selection consistency of regularized m-estimators, 2013. URL https://arxiv.org/abs/1305.7477.

[59] Wonyul Lee and Yufeng Liu. Joint estimation of multiple precision matrices with common structures. *Journal of Machine Learning Research*, 16(31):1035–1062, 2015. URL http://jmlr.org/papers/v16/lee15a.html.

[60] Ed Lein, Lars E Borm, and Sten Linnarsson. The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. *Science*, 358(6359):64–69, 2017.

[61] Xiao Liu and Jeff H Duyn. Time-varying functional network information extracted from brief instances of spontaneous brain activity. *Proceedings of the National Academy of Sciences*, 110 (11):4392–4397, 2013.

[62] Yafei Lyu, Lingzhou Xue, Feipeng Zhang, Hillary Koch, Laura Saba, Katerina Kechris, and Qunhua Li. Condition-adaptive fused graphical lasso (cfgl): An adaptive procedure for inferring condition-specific gene co-expression network. *PLoS computational biology*, 14(9):e1006436, 2018.

[63] Jing Ma and George Michailidis. Joint structural estimation of multiple graphical models. *Journal of Machine Learning Research*, 17(166):1–48, 2016. URL http://jmlr.org/papers/v17/15-656.html.

[64] Carolina Marques, Thomas Unterkircher, Paula Kroon, Barbara Oldrini, Annalisa Izzo, Yuliia Dramaretska, Roberto Ferrarese, Eva Kling, Oliver Schnell, Sven Nelander, et al. Nf1 regulates mesenchymal glioblastoma plasticity and aggressiveness through the ap-1 transcription factor fosl1. *Elife*, 10:e64846, 2021.

[65] Vivien Marx. Method of the year: spatially resolved transcriptomics. *Nature methods*, 18(1): 9–14, 2021.

[66] Vinod Menon. Large-scale brain networks and psychopathology: a unifying triple network model. *Trends in Cognitive Sciences*, 15(10):483–506, 2011.

[67] Lauren MF Merlo, John W Pepper, Brian J Reid, and Carlo C Maley. Cancer as an evolutionary and ecological process. *Nature reviews cancer*, 6(12):924–935, 2006.

[68] Karthik Mohan, Palma London, Maryam Fazel, Daniela Witten, and Su-In Lee. Node-based learning of multiple gaussian graphical models. *The Journal of Machine Learning Research*, 15 (1):445–488, 2014.

[69] Kevin R Moon, David van Dijk, Zheng Wang, Scott Gigante, Daniel B Burkhardt, William S Chen, Kristina Yim, Antonia van den Elzen, Matthew J Hirn, Ronald R Coifman, et al. Visualizing structure and transitions in high-dimensional biological data. *Nature biotechnology*, 37(12):1482–1492, 2019.

[70] Lambda Moses and Lior Pachter. Museum of spatial transcriptomics. *Nature Methods*, 19(5): 534–546, 2022.

[71] Manjari Narayan, Genevera I Allen, and Steffie Tomson. Two sample inference for populations of graphical models with applications to functional connectivity. *arXiv preprint arXiv:1502.03853*, 2015.

[72] Satoru Osuka, Erwin G Van Meir, et al. Overcoming therapeutic resistance in glioblastoma: the way forward. *The Journal of clinical investigation*, 127(2):415–426, 2017.

[73] Jie Peng, Pei Wang, Nengfeng Zhou, and Ji Zhu. Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association*, 104(486):735–746, 2009.

[74] Florian A Potra and Stephen J Wright. Interior-point methods. *Journal of computational and applied mathematics*, 124(1-2):281–302, 2000.

[75] Gemma S Puts, M Kathryn Leonard, Nidhi V Pamidimukkala, Devin E Snyder, and David M Kaetzel. Nuclear functions of nme proteins. *Laboratory Investigation*, 98(2):211–218, 2018.

[76] Anjali Rao, Dalia Barkley, Gustavo S França, and Itai Yanai. Exploring tissue architecture using spatial transcriptomics. *Nature*, 596(7871):211–220, 2021.

[77] Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069, 2010.

[78] Takumi Saegusa and Ali Shojaie. Joint estimation of precision matrices in heterogeneous populations. *Electronic Journal of Statistics*, 10(1):1341 – 1392, 2016. doi: 10.1214/16-EJS1137. URL https://doi.org/10.1214/16-EJS1137.

[79] Yuki Shirakawa, Takuichiro Hide, Michiko Yamaoka, Yuki Ito, Naofumi Ito, Kunimasa Ohta, Naoki Shinojima, Akitake Mukasa, Hideyuki Saito, and Hirofumi Jono. Ribosomal protein s6 promotes stem-like characters in glioma cells. *Cancer science*, 111(6):2041–2051, 2020.

[80] Milena Stevanovic, Natasa Kovacevic-Grujicic, Marija Mojsin, Milena Milivojevic, and Danijela Drakulic. Sox transcription factors and glioma stem cells: Choosing between stemness and differentiation. *World Journal of Stem Cells*, 13(10):1417, 2021.

[81] James M Varah. A lower bound for the smallest singular value of a matrix. *Linear Algebra and its applications*, 11(1):3–5, 1975.

[82] Martin J Wainwright. Sharp thresholds for high-dimensional and noisy sparsity recovery using $\ell_1$-constrained quadratic programming (lasso). *IEEE transactions on information theory*, 55 (5):2183–2202, 2009.

[83] Martin J. Wainwright. *Sparse linear models in high dimensions*, page 194–235. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019. doi: 10.1017/9781108627771.007.

[84] Martin J Wainwright, Michael I Jordan, et al. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305, 2008.

[85] Beilun Wang, Ji Gao, and Yanjun Qi. A Fast and Scalable Joint Estimator for Learning Multiple Related Sparse Gaussian Graphical Models. In Aarti Singh and Jerry Zhu, editors, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 1168–1177. PMLR, 20–22 Apr 2017. URL https://proceedings.mlr.press/v54/wang17e.html.

[86] Shao-Ming Wang, Wen-Chi Lin, Hong-Yi Lin, Yen-Lin Chen, Chiung-Yuan Ko, and Ju-Ming Wang. Ccaat/enhancer-binding protein delta mediates glioma stem-like cell enrichment and atp-binding cassette transporter abca1 activation for temozolomide resistance in glioblastoma. *Cell death discovery*, 7(1):1–11, 2021.

[87] Yahaya A Yabo, Simone P Niclou, and Anna Golebiewska. Cancer cell heterogeneity and plasticity: A paradigm shift in glioblastoma. *Neuro-oncology*, 2021.

[88] Eunho Yang, Aurelie C Lozano, and Pradeep K Ravikumar. Elementary estimators for graphical models. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.

[89] Ming Yuan and Yi Lin. Model selection and estimation in the Gaussian graphical model. *Biometrika*, 94(1):19–35, 03 2007. ISSN 0006-3444. doi: 10.1093/biomet/asm018. URL https://doi.org/10.1093/biomet/asm018.

[90] Bai Zhang, Ye Tian, and Zhen Zhang. Network biology in medicine and beyond. *Circulation: Cardiovascular Genetics*, 7(4):536–547, 2014.

[91] Richard Y Zhang, Salar Fattahi, and Somayeh Sojoudi. Large-scale sparse inverse covariance estimation via thresholding and max-det matrix completion. *International Conference on Machine Learning*, 2018.

[92] Peng Zhao and Bin Yu. On model selection consistency of lasso. *The Journal of Machine Learning Research*, 7:2541–2563, 2006.

[93] Tuo Zhao, Han Liu, Kathryn Roeder, John Lafferty, and Larry Wasserman. The huge package for high-dimensional undirected graph estimation in r. *The Journal of Machine Learning Research*, 13:1059–1062, 2012.

[94] Shuheng Zhou, John Lafferty, and Larry Wasserman. Time varying undirected graphs, 2008.

[95] Jiaqiang Zhu, Shiquan Sun, and Xiang Zhou. Spark-x: non-parametric modeling enables scalable and robust detection of spatial expression patterns for large spatial transcriptomic studies. *Genome Biology*, 22(1):1–25, 2021.