Taylor & Francis
Taylor & Francis Group

Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/uasa20

Graphical Model Inference with Erosely Measured Data

Lili Zheng & Genevera I. Allen

To cite this article: Lili Zheng & Genevera I. Allen (2024) Graphical Model Inference with Erosely Measured Data, Journal of the American Statistical Association, 119:547, 2282-2293, DOI: 10.1080/01621459.2023.2256503

To link to this article: https://doi.org/10.1080/01621459.2023.2256503

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2024, VOL. 119, NO. 547, 2282–2293: Theory and Methods
https://doi.org/10.1080/01621459.2023.2256503

Graphical Model Inference with Erosely Measured Data

Lili Zheng[a] and Genevera I. Allen[a,b,c,d,e]

[a]Department of Electrical and Computer Engineering, Rice University, Houston, TX; [b]Department of Computer Science, Rice University, Houston, TX; [c]Department of Statistics, Rice University, Houston, TX; [d]Department of Pediatrics-Neurology, Baylor College of Medicine, Houston, TX; [e]Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital

ABSTRACT

In this article, we investigate the Gaussian graphical model inference problem in a novel setting that we call *erose* measurements, referring to irregularly measured or observed data. For graphs, this results in different node pairs having vastly different sample sizes which frequently arises in data integration, genomics, neuroscience, and sensor networks. Existing works characterize the graph selection performance using the minimum pairwise sample size, which provides little insights for erosely measured data, and no existing inference method is applicable. We aim to fill in this gap by proposing the first inference method that characterizes the different uncertainty levels over the graph caused by the erose measurements, named GI-JOE (Graph Inference when Joint Observations are Erose). Specifically, we develop an edge-wise inference method and an affiliated FDR control procedure, where the variance of each edge depends on the sample sizes associated with corresponding neighbors. We prove statistical validity under erose measurements, thanks to careful localized edge-wise analysis and disentangling the dependencies across the graph. Finally, through simulation studies and a real neuroscience data example, we demonstrate the advantages of our inference methods for graph selection from erosely measured data. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received October 2022
Accepted August 2023

KEYWORDS
FDR control; Graph selection; Missing data; Graph structure inference; Uneven measurements

1. Introduction

Graphical models have been powerful tools for understanding connection and interaction patterns hidden in large-scale data (Koller and Friedman 2009), by exploiting the conditional dependence relationships among a large number of variables. For instance, graphical models have been applied to learn the connectivity among tens of thousands of neurons (Vinci et al. 2018), gene expression networks (Dobra et al. 2004; Allen and Liu 2012), sensor networks (Dasarathy et al. 2016; Dasarathy 2019), among many others. The last decade has witnessed a plethora of new statistical methods and theory proposed for various types of models in this area, including the Gaussian graphical models (Meinshausen and Bühlmann 2006; Yuan and Lin 2007; Friedman, Hastie, and Tibshirani 2008; Cai, Liu, and Luo 2011; Ravikumar et al. 2011), graphical models for exponential families and mixed variables (Yang et al. 2014, 2015; Chen, Witten, and Shojaie 2015), Gaussian copula models (Liu, Lafferty, and Wasserman 2009; Dobra and Lenkoski 2011; Liu et al. 2012), etc.

Despite the abundant literature in this area, most existing methods and theory for graphical models assume even measurements over the graph, where either all variables are measured simultaneously, or they are missing with similar probabilities. However, many real large-scale datasets usually take the form of *erose measurements*, which are irregular over the graph, and different pairs of variables may have *drastically different sample sizes*. Such datasets frequently arise in genetics, neuroscience, sensor networks, among many others, due to various technological limits.

1.1. Problem Setting and Motivating Applications

Consider the following sparse Gaussian graphical model: $x \sim \mathcal{N}(0, \Sigma^*)$, $\Theta^* = (\Sigma^*)^{-1}$, where $\Theta^* \in \mathbb{R}^{p\times p}$ is the sparse precision matrix. The graph structure is dictated by the nonzero patterns in Θ^*: $\mathcal{G} = (V, E)$, $V = \{1, \ldots, p\}$, $E = \{(i, j) : \Theta^*_{ij} \neq 0\}$, where the unknown edge set E is of primary interest. Suppose that we only have access to the following observations: $\{x_{i,V_i} : V_i \subseteq [p]\}_{i=1}^n$, where V_i is the observed index set of data point i. Then the joint observation set for node pair (j, k) is $O_{jk} = \{i : j, k \in V_i\}$ of size $n_{jk} = |O_{jk}|$. There are a number of applications where n_{jk} can be drastically different.

Heterogeneous missingness: In a variety of biological experiments, some variables could be missing or have erroneous zero reads (dropouts) much more than others, for example, the expression levels of certain genes (Gong et al. 2018; Huang et al. 2018; Gan, Vinci, and Allen 2020), or the abundance of some microbes (Williams et al. 2019). Figure 1 shows the observational patterns and pairwise sample sizes of two real single-cell RNA sequencing (scRNA-seq) datasets, which is far from uniform.

CONTACT Lili Zheng lz67@rice.edu Department of Electrical and Computer Engineering, Rice University, Houston, TX.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

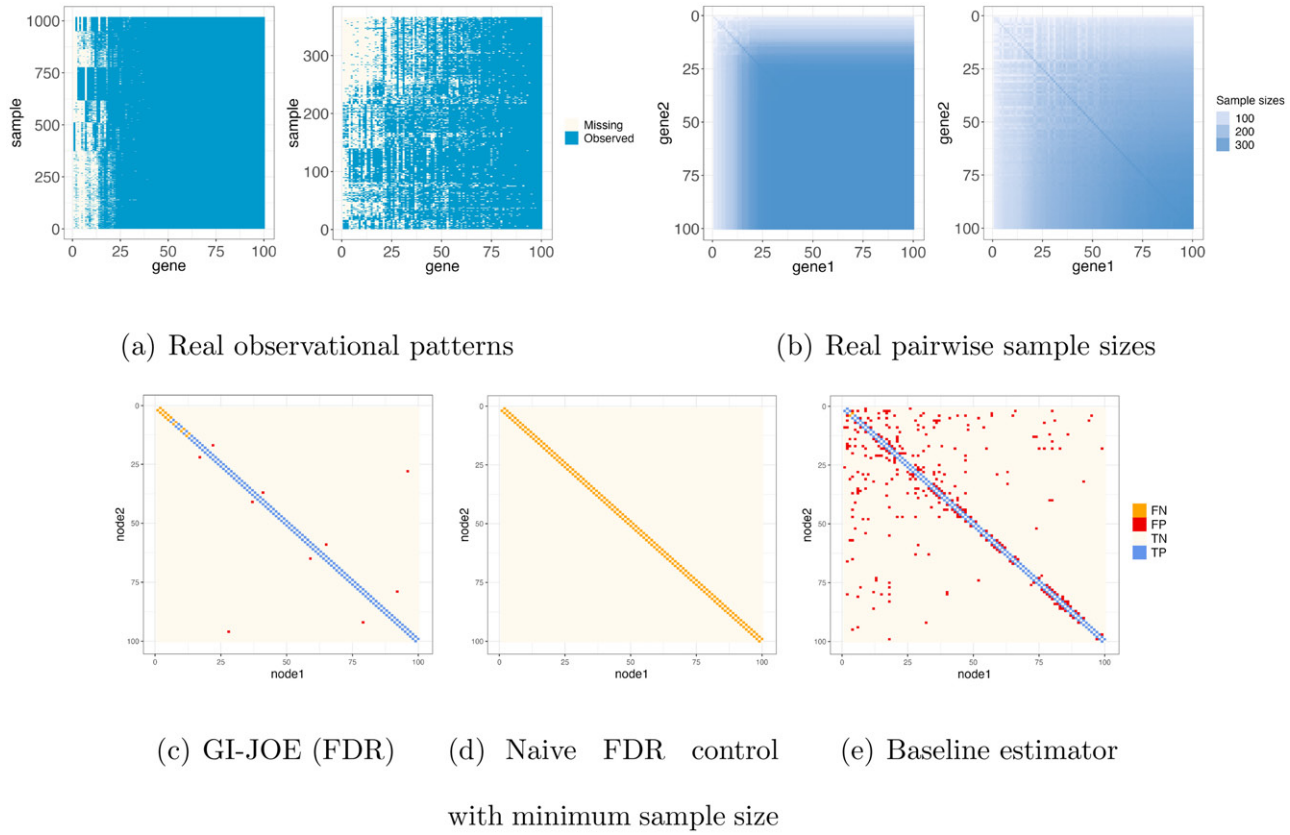


Figure 1. Two erose measurement patterns in real scRNA-seq datasets (Darmanis et al. 2015; Chu et al. 2016) are presented in (a), (b), including the top 100 genes with the highest variances. The pairwise sample sizes range from 0 to 1018 (*chu* data, left) and from 12 to 366 (*darmanis*, right). (c)–(e) present the graph selection and inference results for a chain graph, when the data has the *darmanis* measurement pattern. (c) is selected by our GI-JOE (FDR) approach and is the most accurate; (d) is obtained by an ad hoc implementation of the debiased graphical lasso (Jankova and Van De Geer 2015) that plugs in the minimum pairwise sample size, which is too conservative and identifies no edge at all; (e) is the estimated graph by a baseline approach (Kolar and Xing 2012), which plugs in a covariance estimate into the graphical lasso, and the many false positives suggest that the graph selection problem with such dataset is nontrivial.

Data integration / size-constrained measurements: Non-simultaneous and uneven measurements also frequently arise from data integration and size-constrained measurements. For instance, to better understand the neuronal circuits from neuronal functional activities, one promising strategy is to estimate a large neuronal network (Vinci et al. 2018; Chang and Allen 2021) from in vivo calcium imaging datasets. However, to ensure a sufficient temporal resolution of the recording, the spatial resolution is limited, putting a constraint on the number of neurons simultaneouly measured (Bae et al. 2021; Zheng, Rewolinski, and Allen 2022), and neuron pairs that are further from each other are less likely to be measured together. In genome-wide association studies (GWAS), it is also desirable to integrate genomic data across multiple sources due to the limited sample sizes of each dataset, while these different sources might have different genomic coverage (Cai, Cai, and Zhang 2016). Similar measurement constraints also arise in sensor networks where it is extremely expensive to synchronize a large number of sensors (Dasarathy et al. 2016; Dasarathy 2019).

1.2. Limitations of Existing Works for Erose Measurements

To learn graphical models from erosely measured data, one might want to leverage the current literature on graphical models with missing data (Städler and Bühlmann 2012; Kolar and Xing 2012; Wang et al. 2014; Park, Wang, and Lim 2021). However, most of these works assume the variables are missing independently with the same missing probability. While Park, Wang, and Lim (2021) allows for arbitrary missing probabilities and dependency in their problem formulation, their theoretical guarantees still hinge on the minimum observational probability. Using the minimum pairwise sample size over the whole graph to characterize the performance of the graph learning result can be too coarse and provides little insights to erosely measured datasets. Interestingly, one recent work (Zheng and Allen 2022) provides a localized theoretical guarantee for neighborhood selection consistency, requiring only sample size conditions imposed upon the corresponding neighbors instead of all node pairs. Such theoretical results suggest that the estimation accuracy should vary over the graph when measurements are erose, and a coarse characterization based on the minimum sample size would only provide insights for the worst part of the graph estimate.

Inspired by this intuition, here arises one natural question: *can we develop a statistical inference method that quantifies the different uncertainty levels over the graph arising from the erose measurements?* Over the last decade, significant efforts have been devoted to the statistical inference in high-dimensional settings, including techniques such as the debiased Lasso (Van de Geer et al. 2014; Zhang and Zhang 2014; Javanmard and Montanari 2014), post-selection inference approaches (Lee et al. 2016; Tibshirani et al. 2016), knockoff methods (Barber and Candès 2015;

Candès et al. 2018), and various other FDR control methods (Liu 2013; Javanmard and Javadi 2019). These techniques have been applied in regression or classification problems, as well as in graphical models. However, these prior works mainly consider simultaneous measurements across all variables (Liu 2013; Jankova and Van De Geer 2015; Ren et al. 2015; Gu et al. 2015; Janková and van de Geer 2017; Yu, Gupta, and Kolar 2020), which, in the context of graphical models, would result in the same sample size across the entire graph; or they consider the missing data setting where all variables are missing independently with the same missing probability (Belloni, Chernozhukov, and Kaul 2017), still leading to approximately the same sample sizes. To the best of our knowledge, there is no applicable statistical inference method for the general observational patterns and erose measurements that we are considering. If practitioners want to apply these existing inference methods with erosely measured data, they have to come up with one single sample size quantity n to determine the uncertainty levels for each edge. To ensure the validity of the test, one ad hoc way might be to plug in the minimum pairwise sample size, which can be extremely conservative and has no power (see Figure 1(d)).

The rest of the article is organized as follows. We first review the set-ups and neighborhood selection results from Zheng and Allen (2022) in Section 2, which serves as an inspiration and basis of our graph inference method under erose measurements; Our key contribution, the GI-JOE approach, is introduced in Sections 3 and 4. In particular, Section 3 is devoted to the edge-wise inference method, and for any node pair, we characterize its Type I error and power based on the sample sizes involving the node pair's neighbors. Section 4 focuses on the FDR control procedure, also shown to be theoretically valid under appropriate conditions. The synthetic and real data experiments are included in Sections 5. We conclude with discussion of some open questions in Section 6.

Notations: For any matrix $A \in \mathbb{R}^{p_1 \times p_2}$, let $\|A\|_\infty = \max_{j,k} |A_{j,k}|$, $\|A\| = \sup_{\|u\|_2=1} \|Au\|_2$ be its spectral norm, and $|||A|||_\infty = \max_{j=1,\ldots,p_1} \sum_{k=1}^{p} |A_{j,k}|$ be the matrix-operator ℓ_∞ to ℓ_∞ norm. For any tensor $\mathcal{T} \in \mathbb{R}^{p_1 \times p_2 \times p_3 \times p_4}$ and matrix $A \in \mathbb{R}^{p_1, q_1}$ define the tensor-matrix/vector product $\mathcal{T} \times_1 A \in \mathbb{R}^{q_1 \times p_2 \times p_3 \times p_4}$ as follows: $(\mathcal{T} \times_1 A)_{i_1,i_2,i_3,i_4} = \sum_{j_1=1}^{p_1} A_{j_1,i_1} \mathcal{T}_{j_1,i_2,i_3,i_4}$. Similarly we extend this definition of tensor-matrix product to other modes.

2. Graph Selection with Erose Measurements

In this section, we review the set-up and neighborhood selection theory in Zheng and Allen (2022), as it underpins our own inference procedure and theory in Section 3. In particular, we follow Zheng and Allen (2022) and study a variant of the neighborhood lasso method instead of other graph estimation methods (Yuan and Lin 2007; Cai, Liu, and Luo 2011), since its form makes it easier to disentangle the effects of different parts of the graph on each other.

The neighborhood lasso algorithm proposed in Zheng and Allen (2022) consists of two steps: estimating the true covariance Σ^* and plugging the estimate into a neighborhood lasso estimator. An unbiased estimate $\widehat{\Sigma}$ is defined as follows: given observations $\{x_{i,V_i}\}_{i=1}^n$, for each entry (j,k), $\widehat{\Sigma}_{j,k} = \frac{1}{n_{j,k}} \sum_{i:j,k \in V_i} x_{i,j} x_{i,k}$. However, $\widehat{\Sigma}$ is not guaranteed to be positive semidefinite, resulting in both optimization and statistical issues in neighborhood lasso. To ensure convexity and preserve the entry-wise error bounds for $\widehat{\Sigma}_{j,k} - \Sigma^*_{j,k}$, an additional projection step upon the positive semidefinite cone is considered:

$$\widetilde{\Sigma} = \arg\min_{\Sigma \succeq 0} \max_{j,k} \sqrt{n_{j,k}} |\Sigma_{j,k} - \widehat{\Sigma}_{j,k}|, \tag{1}$$

where $n_{j,k}$ is the pairwise sample size associated with node pair (j,k), defined as in Section 1.1. The projection problem (1) can be solved by the ADMM, and we include the detailed optimization steps in Section A of the supplementary material.

Given the covariance estimate $\widetilde{\Sigma}$, for any target node a of which we want to estimate the neighborhood, consider the following neighborhood regression problem:

$$\widehat{\theta}^{(a)} = \arg\min_{\theta \in \mathbb{R}^p, \theta_a = 0} \frac{1}{2} \theta^\top \widetilde{\Sigma} \theta - \widetilde{\Sigma}_{a,:} \theta + \sum_{j=1}^{p} \lambda_j^{(a)} |\theta_j|, \tag{2}$$

where $\lambda^{(a)} = (\lambda_1^{(a)}, \ldots, \lambda_p^{(a)})^\top \in \mathbb{R}^p$ is a vector of tuning parameters, with each entry $\lambda_j^{(a)}$ corresponding to a potential edge connecting node j and a. The solution $\widehat{\theta}^{(a)}$ serves as an estimate for $\theta^{(a)*} = \arg\min_{\theta \in \mathbb{R}^p, \theta_a = 0} \frac{1}{2} \theta^\top \Sigma^* \theta - \Sigma^*_{a,:} \theta$, which satisfies $\theta^{(a)*}_{\setminus a} = (\Sigma^*_{\setminus a, \setminus a})^{-1} \Sigma^*_{\setminus a, a} = \frac{1}{\Theta^*_{a,a}} \Theta^*_{\setminus a, a}$, and hence the support set of $\theta^{(a)*}$ equals the true neighborhood of node a: $\mathcal{N}_a = \{j \neq a : \Theta^*_{a,j} \neq 0\}$. Then one can estimate $\mathcal{N}_a$ by the support of $\widehat{\theta}^{(a)}$: $\widehat{\mathcal{N}}_a = \{j \neq a : \widehat{\theta}^{(a)} \neq 0\}$. It was shown in Zheng and Allen (2022) that the neighborhood selection consistency is guaranteed with sample size conditions involving the neighbors of node a. Here, we present a similar theoretical result, with only a slight modification on the tuning parameter choice. Let $\gamma_a = \frac{\max_{j \in \mathcal{N}_a^c} \min_k n_{j,k}}{\min_{j \in \mathcal{N}_a} \min_k n_{j,k}}$ be the sample size ratio between a's non-neighbors and neighbors.

Theorem 1 (Neighborhood Selection Consistency, Similar to Zheng and Allen 2022). Consider the Gaussian graphical model with erose measurement setting described in Section 1.1 and the estimator $\widehat{\theta}^{(a)}$ defined in (2). Suppose Assumption B.1 in the Supplementary Material (the mutual incoherence condition) holds, and the tuning parameters $\lambda_j^{(a)}$'s in (2) satisfy $\lambda_j^{(a)} \asymp \|\Sigma^*\|_\infty \frac{\|\Theta^*_{:,a}\|_1}{\Theta^*_{a,a}} \sqrt{\frac{\log p}{\min_k n_{j,k}}}$. If $\gamma_a \leq C$ for some $C > 0$ depending on the incoherence parameter,

$$\min_{j \in \mathcal{N}_a} \min_k n_{j,k} \geq C(\Sigma^*) \|\Sigma^*\|_\infty^2 \left[d_a^2 + (\theta^{(a)}_{\min})^{-2} \right] \log p, \tag{3}$$

where the constant $C(\Sigma^*)$ depends on Σ^*, then $\widehat{\mathcal{N}}_a = \{j : \widehat{\theta}_j^{(a)} \neq 0\} = \mathcal{N}_a$ with probability at least $1 - p^{-c}$ for some absolute constants $c > 0$.

The complete version of Theorem 1, additional ℓ_1 and ℓ_2 error bounds for $\widehat{\theta}^{(a)} - \theta^{(a)*}$, a pictorial illustration of the sample size condition (3), and the proofs can be found in Section B and G of the supplementary material. The localized characterization

of the graph estimation performance in Theorem 1 inspires us to develop an inference method that quantifies the uneven uncertainty levels over the graph.

3. Edge-wise Inference: Quantifying Uncertainties from Erose Measurements

In this section, we propose our GI-JOE method for edge-wise inference with erose data. The key idea follows the debiased lasso (Van de Geer et al. 2014), while the main challenge and innovation is characterizing the uncertainty level associated with each edge-wise statistic. We first introduce our edge-wise debiased statistic $\widetilde{\theta}_b^{(a)}$ in Section 3, and characterize its asymptotic distribution in Section 3.2; We further propose a consistent estimator of its variance and establish statistical validity of edge-wise inference in Section 3.3.

3.1. Debiased Neighborhood Lasso

First we introduce the key idea and intuition behind how we construct our debiased test statistic. Recall that our neighborhood regression estimator $\widehat{\theta}^{(a)}$ was defined as in (2), and by the Karush–Kuhn–Tucker (KKT) condition, we know that it satisfies

$$\widetilde{\Sigma}_{\backslash a,\backslash a}\widehat{\theta}_{\backslash a}^{(a)} - \widetilde{\Sigma}_{\backslash a,a} + (\lambda^{(a)} \circ \widehat{Z})_{\backslash a} = 0,$$

where $\circ$ represents element-wise multiplication, and $\widehat{Z} \in \mathbb{R}^p$ satisfies that $\|\widehat{Z}\|_\infty \le 1$ and $\widehat{Z}_j = \mathrm{sgn}(\widehat{\theta}_j^{(a)})$ if $\widehat{\theta}_j^{(a)} \neq 0$. Noting the fact that $\Sigma^*\theta^{(a)*} - \Sigma^*_{:,a} = 0$, and the relationship between $\theta^{(a)*}$ and $\Theta^*_{:,a}$, we can use some rearrangements to obtain the following:

$$\begin{aligned}
&\widetilde{\Sigma}_{\backslash a,\backslash a}(\widehat{\theta}^{(a)} - \theta^{(a)*})_{\backslash a} + (\lambda^{(a)} \circ \widehat{Z})_{\backslash a} \\
&\quad = (\Theta^*_{a,a})^{-1}(\widetilde{\Sigma} - \Sigma^*)_{\backslash a,:}\Theta^*_{:,a}, \\
&\widetilde{\Sigma}_{\backslash a,\backslash a}(\widehat{\theta}^{(a)} - \theta^{(a)*})_{\backslash a} + \widetilde{\Sigma}_{\backslash a,a} - \widetilde{\Sigma}_{\backslash a,\backslash a}\widehat{\theta}_{\backslash a}^{(a)} \\
&\quad = (\Theta^*_{a,a})^{-1}(\widetilde{\Sigma} - \Sigma^*)_{\backslash a,:}\Theta^*_{:,a}.
\end{aligned}$$

The derivation above follows similar arguments for the debiased lasso in Van de Geer et al. (2014), and ideally, we would hope the RHS of the equation above has (asymptotically) normal distribution and can serve as a basis for our inference. However, since $\widetilde{\Sigma}$ is the solution of a weighted ℓ_∞ projection onto the positive semi-definite cone, the CLT is not directly applicable. Instead, we want to change it to a function of $\widehat{\Sigma}$, whose entries can be written as independent sums. As will be shown in our proofs, substituting $\widetilde{\Sigma}$ by $\widehat{\Sigma}$ in the debiasing terms above can help us achieve this goal: $\widetilde{\Sigma}_{\backslash a,\backslash a}(\widehat{\theta}^{(a)} - \theta^{(a)*})_{\backslash a} + \widetilde{\Sigma}_{\backslash a,a} - \widehat{\Sigma}_{\backslash a,\backslash a}\widehat{\theta}_{\backslash a}^{(a)} \approx (\Theta^*_{a,a})^{-1}(\widehat{\Sigma} - \Sigma^*)_{\backslash a,:}\Theta^*_{:,a}$. Furthermore, to invert the factor $\widetilde{\Sigma}_{\backslash a,\backslash a}$, we need a good approximation of $(\Sigma^*_{\backslash a,\backslash a})^{-1} \in \mathbb{R}^{(p-1)\times(p-1)}$. Define the debiasing matrix $\Theta^{(a)*} \in \mathbb{R}^{p\times p}$, which satisfies $\Theta^{(a)*}_{a,:} = 0$, $\Theta^{(a)*}_{:,a} = 0$, and $\Theta^{(a)*}_{\backslash a,\backslash a} = (\Sigma^*_{\backslash a,\backslash a})^{-1}$. Then suppose we have a good estimate $\Theta^{(a)}$, one would be able to show

$$\begin{aligned}
&\widehat{\theta}^{(a)} - \theta^{(a)*} + \Theta^{(a)}(\widehat{\Sigma}_{:,a} - \widehat{\Sigma}_{:,\backslash a}\widehat{\theta}_{\backslash a}^{(a)}) \\
&\quad \approx (\Theta^*_{a,a})^{-1}\Theta^{(a)*}(\widehat{\Sigma} - \Sigma^*)\Theta^*_{:,a}. \qquad (4)
\end{aligned}$$

Given a node pair (a, b) for $a \neq b$, this motivates us to consider an edge-wise test statistic of the form $\widehat{\theta}_b^{(a)} + \Theta^{(a)}_{b,:}(\widehat{\Sigma}_{:,a} - \widehat{\Sigma}_{:,\backslash a}\widehat{\theta}_{\backslash a}^{(a)})$, where $\Theta^{(a)}_{b,:}$ is an appropriate estimate for $\Theta^{(a)*}_{b,:}$.

Throughout the rest of this section, suppose that we are interested in testing whether there is an edge between node $a \neq b$. Now we introduce our estimates for $\Theta^{(a)*}_{b,:}$. Denote by $\mathcal{N}_b^{(a)}$ the support set of $\Theta^{(a)*}_{b,:}$ and $\overline{\mathcal{N}}_b^{(a)} = \mathcal{N}_b^{(a)} \cup j$. By block matrix inverse formula, $\Theta^{(a)*}_{b,:} = \Theta^*_{b,:} - (\Theta^*_{a,a})^{-1}\Theta^*_{b,a}\Theta^*_{a,:}$ and hence is also sparse with $d_b^{(a)} := |\mathcal{N}_b^{(a)}| \le d_a + d_b$. Therefore, we can estimate $\Theta^{(a)*}_{b,:}$ by performing another neighborhood regression. Let

$$\begin{aligned}
\widehat{\theta}^{(a,b)} &= \underset{\theta\in\mathbb{R}^p,\theta_a=\theta_b=0}{\arg\min} \frac{1}{2}\theta^\top\widetilde{\Sigma}\theta - \widetilde{\Sigma}_{b,:}\theta + \sum_{k=1}^p \lambda_k^{(a,b)}|\theta_k|, \qquad (5) \\
\widehat{\theta}_b^{(a,b)} &= 1,\ \widehat{\theta}_{\backslash b}^{(a,b)} = -\widehat{\theta}_{\backslash b}^{(a,b)},
\end{aligned}$$

where $\lambda_k^{(a,b)}$'s are tuning parameters depending on the pairwise sample sizes $\min_{i\in[p]} n_{i,b}$. Then $\widehat{\theta}^{(a,b)}$ serves as an estimate of $(\Theta^{(a)*}_{b,b})^{-1}\Theta^{(a)*}_{b,:}$. To estimate $\Theta^{(a)*}_{b,b}$, we note the fact that $\Theta^{(a)*}_{b,b} = [\Sigma^*_{b,:}(\Theta^{(a)*}_{b,b})^{-1}\Theta^{(a)*}_{:,b}]^{-1} = [(\Theta^{(a)*}_{b,b})^{-2}\Theta^{(a)*}_{b,:}\Sigma^*\Theta^{(a)*}_{:,b}]^{-1}$. Hence, either of the following two estimators can serve appropriately for estimating $\Theta^{(a)*}_{b,:}$:

$$\begin{aligned}
\widehat{\Theta}^{(a)}_{b,b} &= (\widetilde{\Sigma}_{b,:}\widehat{\theta}^{(a,b)})^{-1},\ \widehat{\Theta}^{(a)}_{b,:} = \widehat{\Theta}^{(a)}_{b,b}\widehat{\theta}^{(a,b)}, \\
\widetilde{\Theta}^{(a)}_{b,b} &= (\widehat{\theta}^{(a,b)\top}\widetilde{\Sigma}\widehat{\theta}^{(a,b)})^{-1},\ \widetilde{\Theta}^{(a)}_{b,:} = \widetilde{\Theta}^{(a)}_{b,b}\widehat{\theta}^{(a,b)}.
\end{aligned} \qquad (6)$$

As we will show in Theorem G.2 in the supplementary material, both estimators are consistent and lead to sufficiently good statistical error bounds. Based on some empirical investigation (details presented in Section F), we propose to use $\widehat{\Theta}^{(a)}_{b,:}$ for the debiasing step, but would revisit $\widetilde{\Theta}^{(a)}_{b,:}$ for variance estimation in Section 3.3. Then the debiased neighborhood lasso estimator for node pair (a, b) is

$$\widetilde{\theta}_b^{(a)} = \widehat{\theta}_b^{(a)} - \widehat{\Theta}^{(a)}_{b,:}(\widehat{\Sigma}\widehat{\theta}^{(a)} - \widehat{\Sigma}_{:,a}). \qquad (7)$$

3.2. Normal Approximation of Debiased Edge-wise Statistic

Although the edge-wise statistic $\widetilde{\theta}_b^{(a)}$ defined in (7) is similar to the debiased lasso in the literature, its asymptotical normality is not readily present due to the erose measurement setting we are concerned with. In the following, we present a novel characterization of $\widetilde{\theta}_b^{(a)}$ that consists of a bias term and an asymptotically normal error term, each term depending on one pairwise sample size quantity, respectively. Before presenting the main theorem, we first define and discuss these two key sample size quantities.

Given the target node pair (a, b), define two sets of node pairs involving a, b's neighbors: $S_1(a, b) = \{(j, k) : j \text{ or } k \in \mathcal{N}_a \cup \overline{\mathcal{N}}_b^{(a)}\}$, $S_2(a, b) = \{(j, k) : \Theta^{(a)*}_{j,b}\Theta^*_{k,a} + \Theta^{(a)*}_{k,b}\Theta^*_{j,a} \neq 0\}$, where $\overline{\mathcal{N}}_b^{(a)}$ and matrix $\Theta^{(a)*}$ are defined in the beginning of Section 3. Here the order of a and b matters since we first apply neighborhood lasso for node a and then debias its entry

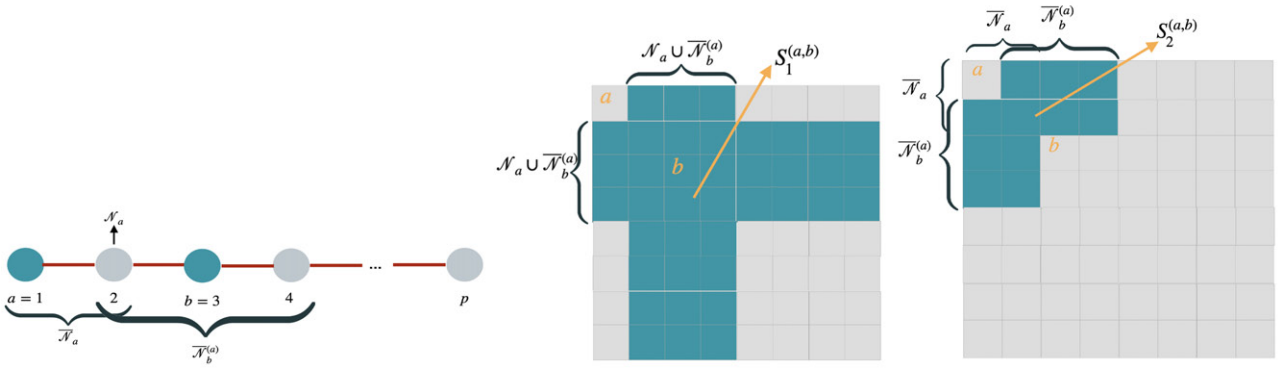


Figure 2. An illustration of the set $S_1(a,b)$ and $S_2(a,b)$ in a chain graph, when $a = 1$ and $b = 3$. The minimum sample size in $S_1(a,b)$ determines the bias for estimating edge (a,b), while the minimum sample size in $S_2(a,b)$ determines the variance for estimating edge (a,b).

$\widehat{\theta}_b^{(a)}$. Proposition 1 characterizes the index set $S_2(a,b)$ and $\overline{\mathcal{N}}_b^{(a)}$ through their relationships with $\overline{\mathcal{N}}_a$ and $\overline{\mathcal{N}}_b$. Figure 2 also gives a pictorial illustration of $S_1(a,b)$ and $S_2(a,b)$ for a chain graph. The two key sample size quantities are then defined as the minimum pairwise sample sizes within these two sets: $n_1^{(a,b)} = \min_{(j,k)\in S_1(a,b)} n_{j,k}$, $n_2^{(a,b)} = \min_{(j,k)\in S_2(a,b)} n_{j,k}$, which will be shown to determine the bias and variance of the edge-wise statistic. The intuition behind these two node pair sets can be traced back to our main idea for constructing the debiased edge-wise statistic in Section 3. As shown in (4), our debiased test statistic for (a,b) can be well approximated by $\langle \widehat{\Sigma} - \Sigma^*, \frac{\Theta^*_{:,a}\Theta^{(a)*}_{b,:}}{\Theta^*_{a,a}} \rangle = \langle \widehat{\Sigma} - \Sigma^*, \frac{\Theta^*_{:,a}\Theta^{(a)*}_{b,:} + \Theta^{(a)*}_{:,b}\Theta^*_{a,:}}{2\Theta^*_{a,a}} \rangle$, which can be intuitively understood as a first order Taylor's expansion of the estimation error around Σ^*. The approximation errors constitute our bias term, which mainly depends on how well we estimate $\theta^{(a)*}$ and $\Theta^{(a)*}_{b,:}$ using neighborhood regression. Similar to the neighborhood selection theory presented in Section 2, we can show that the estimation error for $\theta^{(a)*}$ and $\Theta^{(a)*}_{b,:}$ depend on the pairwise sample sizes $n_{j,k}$ for j or k in the neighborhood sets $\mathcal{N}_a$ and $\overline{\mathcal{N}}_b^{(a)}$, respectively, and hence this leads to our definition of set $S_1(a,b)$. On the other hand, since the matrix $\Theta^*_{:,a}\Theta^{(a)*}_{b,:} + \Theta^{(a)*}_{:,b}\Theta^*_{a,:}$ has support set $S_2(a,b)$, the variance of $\langle \widehat{\Sigma} - \Sigma^*, \frac{\Theta^*_{:,a}\Theta^{(a)*}_{b,:} + \Theta^{(a)*}_{:,b}\Theta^*_{a,:}}{2\Theta^*_{a,a}} \rangle$ is then dominated by the minimum sample sizes in $S_2(a,b)$.

Proposition 1. For any given support set $\overline{E} \subseteq [p] \times [p]$, the following holds except when $\Theta^*_{\overline{E}} \in \mathbb{R}^{|\overline{E}|}$ falls in a measure zero set: (i) $S_2(a,b) = (\overline{\mathcal{N}}_a \times \overline{\mathcal{N}}_b^{(a)}) \cup (\overline{\mathcal{N}}_b^{(a)} \times \overline{\mathcal{N}}_a)$. (ii) If $b \in \mathcal{N}_a$, $\overline{\mathcal{N}}_b^{(a)} = \overline{\mathcal{N}}_a \cup \overline{\mathcal{N}}_b$; otherwise, $\overline{\mathcal{N}}_b^{(a)} = \overline{\mathcal{N}}_b$.

Similar to the support recovery guarantee in Theorem 1, here we also define the sample size ratio for node b here by $\gamma_b^{(a)} = \frac{\max_{j\in\overline{\mathcal{N}}_b^{(a)c}} \min_k n_{j,k}}{\min_{j\in\mathcal{N}_b^{(a)}} \min_k n_{j,k}}$. T he following covariance parameters are also useful: let $\mathcal{T}^*, \mathcal{T}^{(n)*} \in \mathbb{R}^{p\times p\times p\times p}$ satisfy

$$\mathcal{T}^*_{j,k,j',k'} = \mathrm{Cov}(X_jX_k, X_{j'}X_{k'}) = \Sigma^*_{j,j'}\Sigma^*_{k,k'} + \Sigma^*_{j,k'}\Sigma^*_{k,j'}$$

for $1 \le j,k,j',k' \le p$, and $(\mathcal{T}^{(n)*})_{j,k,j',k'} = \mathcal{T}^*_{j,k,j',k'}\frac{n_{j,k,j',k'}}{n_{j,k}n_{j',k'}}$, where $n_{j,k,j',k'} = |\{i : j,k,j',k' \in V_i\}|$ is the number of joint measurements for j,k,j',k'.

Assumption 1 (Sample size condition for accurate estimation).

$$n_1^{(a,b)} \ge C\frac{\|\Sigma^*\|_\infty^2}{\lambda_{\min}^2(\Sigma^*)}(\kappa_{\Sigma^*}^2 + \gamma_a + \gamma_b^{(a)})(d_a + d_b + 1)^2 \log p,$$

Assumption 1 is similar to the sample size condition in Theorem 1, while the only difference lies that here $\min_k n_{j,k}$ needs to be large as long as $j \in \mathcal{N}_a \cup \overline{\mathcal{N}}_b^{(a)}$ instead of $\mathcal{N}_a$ only, so that both $\widehat{\theta}^{(a)}$ and $\widehat{\Theta}^{(a)}_{b,:}$ are accurate estimators for $\theta^{(a)*}$ and $\Theta^{(a)*}_{b,:}$.

Assumption 2 (Sample size condition for normal approximation). $C_\varepsilon(\Sigma^*)(d_a + d_b + 1)^{2+\varepsilon} = o(n_2^{(a,b)})$ for some constant $\varepsilon > 0$, where $C_\varepsilon(\Sigma^*) = \left(\frac{C(1+2/\epsilon)\|\Sigma^*\|_\infty}{\lambda_{\min}(\Sigma^*)}\right)^{2+\epsilon}$.

Due to the erose measurements, establishing the Lyapunov condition is much more complicated than the same sample size setting. Assumption 2 is a technical assumption we need in this step so that the CLT can be applied to derive asymptotic normality results.

Assumption 3 (Sample size condition for controlling bias).

$$n_1^{(a,b)} \gg C^2(\Theta^*; a, b)(\kappa_{\Theta^*}^4 + \gamma_a + \gamma_b^{(a)})[(d_a + d_b + 1)\log p]^2 \frac{n_2^{(a,b)}}{n_1^{(a,b)}}, \tag{8}$$

where $C(\Theta^*; a, b) = \frac{C\kappa_{\Theta^*}^3 \|\Theta^*_{:,a}\|_1 \|\Theta^{(a)*}_{:,b}\|_1}{\min_{(j,k)\in S_2(a,b)} \left|\Theta^{(a)*}_{b,j}\Theta^*_{a,k} + \Theta^{(a)*}_{b,k}\Theta^*_{a,j}\right|}$.

Remark 1. Rearranging (8), we can also write the this sample size condition as

$$n_1^{(a,b)} \gg C(\Theta^*; a, b)(\kappa_{\Theta^*}^2 + \sqrt{\gamma_a} + \sqrt{\gamma_b^{(a)}}) \times [(d_a + d_b + 1)\log p]\sqrt{n_2^{(a,b)}}.$$

Remark 2. One may be confused when seeing $n_1^{(a,b)}$ both on the left- and right-hand side of (8). In fact, we present it this way in

order to connect it to the same sample size setting where $n_2^{(a,b)} = n_1^{(a,b)} = n$, and thus (8) becomes $n \gg C^2(\Theta^*; a, b)\kappa_{\Theta^*}^4[(d_a + d_b + 1)\log p]^2$. This is similar to the requirement in prior results on debiased lasso and debiased graphical lasso (Van de Geer et al. 2014; Zhang and Zhang 2014; Jankova and Van De Geer 2015) with the same sample sizes, which requires $n \gg d^2 \log^2 p$. The additional price we paid for uneven sample sizes is reflected in the sample size ratios γ_a, $\gamma_b^{(a)}$ and $\frac{n_2^{(a,b)}}{n_1^{(a,b)}}$.

Remark 3 (Effect of γ_a, $\gamma_b^{(a)}$ and $\frac{n_2^{(a,b)}}{n_1^{(a,b)}}$). γ_a, $\gamma_b^{(a)}$ are the sample size ratios between the most well measured non-neighbor and the worst measured neighbor of a and b. These two quantities have a negative effect on our theory, as when the sample sizes for the non-neighbors are all much larger than the neighbors, the neighbors would suffer from much stronger regularization than non-neighbors. While for the sample size ratio $\frac{n_2^{(a,b)}}{n_1^{(a,b)}}$, note that when $n_2^{(a,b)}$ grows too much more quickly than $n_1^{(a,b)}$, the bias term would dominate the variance term and then the normal approximation of $\widetilde{\theta}_b^{(a)}$ would not hold.

The following theorem establishes the asymptotic normality of $\widetilde{\theta}_b^{(a)} + \frac{\Theta_{a,b}^*}{\Theta_{a,a}^*}$ under these three sample size assumptions, and Corollary 1 presents its direct consequence when all pairwise sample sizes are equal ($n_{i,j} = n$), with simplified sample size assumptions that is comparable to prior literature (Van de Geer et al. 2014; Jankova and Van De Geer 2015).

Theorem 2 (Asymptotic Normality). Consider the Gaussian graphical model with erose measurement setting described in Section 1.1 and the debiased edge-wise statistic $\widetilde{\theta}_b^{(a)}$ defined in (7). Suppose that $\lambda^{(a)}$ in (2) is chosen as in Theorem 1, and $\lambda^{(a,b)}$ in (5) satisfies $\lambda_k^{(a,b)} \asymp \|\Sigma^*\|_\infty \frac{\|\Theta_{b,:}^{(a)*}\|_1}{\Theta_{b,b}^{(a)*}} \sqrt{\frac{\log p}{\min_{j\in[p]} n_{j,k}}}$. Then we have the following decomposition:

$$\widetilde{\theta}_b^{(a)} = -\frac{\Theta_{a,b}^*}{\Theta_{a,a}^*} + B + E. \tag{9}$$

If Assumption 1 holds, then with probability at least $1 - Cp^{-c}$, $|B| \le C(\Theta^*, \gamma_a, \gamma_b^{(a)})\frac{(d_a+d_b+1)\log p}{n_1^{(a,b)}}$, where $C(\Theta^*, \gamma_a, \gamma_b^{(a)}) = C\kappa_{\Sigma^*}(\kappa_{\Sigma^*}^2 + \sqrt{\gamma_a} + \sqrt{\gamma_b^{(a)}})\|\Sigma^*\|_\infty^2 \|\Theta_{:,a}^*\|_1 \|\Theta_{:,b}^{(a)*}\|_1$. If Assumption 2 holds, $\sigma_n^{-1}(a,b)E \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$ with $\sigma_n^2(a,b) = \frac{1}{\Theta_{aa}^{*2}} \mathcal{T}^{(n)*} \times_1 \Theta_{:,b}^{(a)*} \times_2 \Theta_{:,a}^* \times_3 \Theta_{:,b}^{(a)*} \times_4 \Theta_{:,a}^*$. Furthermore, if Assumptions 1–3 hold,

$$\sigma_n^{-1}(a,b)\left(\widetilde{\theta}_b^{(a)} + \frac{\Theta_{a,b}^*}{\Theta_{aa}^*}\right) \xrightarrow{\mathrm{d}} \mathcal{N}(0,1).$$

The proof of Theorem 2 is deferred to Section G of the supplementary material.

Remark 4 (Bias-Variance decomposition). As suggested by (9), the error of the debiased lasso estimator can be decomposed into a bias term (B) and a variance term (E), where B depends on the minimum pairwise sample size $n_1^{(a,b)}$ between any nodes and the neighbors of nodes a, b, while E depends only on the sample size $n_2^{(a,b)}$ for nodes within the neighborhoods of a, b (See Figure 2). When $C(\Theta^*, \gamma_a, \gamma_b^{(a)})$ is viewed as a constant, then $|B| \asymp \frac{(d_a+d_b+1)\log p}{n_1^{(a,b)}}$, and the term E scales as the asymptotic standard deviation $\sigma_n(a,b)$, which is further characterized by Proposition 2.

Proposition 2 (Variance characterization). The variance term $\sigma_n^2(a,b)$ satisfies

$$\begin{aligned}
\sigma_n(a,b) &\le \frac{\sqrt{2}\lambda_{\max}(\Sigma^*)\|\Theta_{:,b}^{(a)*}\|_2\|\Theta_{:,a}^*\|_2}{\Theta_{a,a}^*}(n_2^{(a,b)})^{-\frac{1}{2}} \\
&\le \sqrt{2}\kappa_{\Sigma^*}^2(n_2^{(a,b)})^{-\frac{1}{2}}, \\
\sigma_n(a,b) &\ge \frac{\sqrt{2}\lambda_{\min}(\Sigma^*)\min_{(j,k)\in S_2(a,b)}\left|\Theta_{b,j}^{(a)*}\Theta_{a,k}^* + \Theta_{b,k}^{(a)*}\Theta_{a,j}^*\right|}{2\Theta_{a,a}^*} \\
&\quad \times (n_2^{(a,b)})^{-\frac{1}{2}}.
\end{aligned}$$

When $C_1 \le \frac{\lambda_{\min}(\Sigma^*)\min_{(j,k)\in S_2(a,b)}\left|\Theta_{b,j}^{(a)*}\Theta_{a,k}^* + \Theta_{b,k}^{(a)*}\Theta_{a,j}^*\right|}{\Theta_{a,a}^*} \le \kappa_{\Sigma^*}^2 \le C_2$, Proposition 2 suggests that $\sigma_n(a,b) \asymp (n_2^{(a,b)})^{-1/2}$.

Corollary 1 (Normal Approximation with the Same Sample Size). Consider the same model, edge-wise statistic and tuning parameters as in Theorem 2. When the pairwise sample sizes are all equal: $n_{j,k} = n$, then if for some $\epsilon > 0, n \gg C^2(\Theta^*; a, b)\kappa_{\Theta^*}^4(d_a+d_b+1)^2\log^2 p + C_\epsilon(\Sigma^*)(d_a+d_b+1)^{2+\epsilon}$, we have $\sigma_n^{-1}(a,b)\left(\widetilde{\theta}_b^{(a)} + \frac{\Theta_{a,b}^*}{\Theta_{aa}^*}\right) \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$, where $C(\Theta^*; a, b)$ and $C_\epsilon(\Sigma^*)$ are as defined in Assumptions 2 and 3, depending only on Σ^* and Θ^*. In addition, if the sample size of all quadruples $n_{j,k,j',k'} = n$, $\sigma_n^2(a,b) = \frac{1}{n}\frac{\Theta_{a,a}^*\Theta_{b,b}^* - (\Theta_{a,b}^*)^2}{(\Theta_{a,a}^*)^2}$.

Remark 5. Corollary 1 is a direct consequence of Theorem 2. If $d_a + d_b + 1 \le (\log p)^c$ for some $c > 0$, the sample size condition is the same as the prior literature on debiased lasso and debiased graphical lasso (Van de Geer et al. 2014; Jankova and Van De Geer 2015). Note that Corollary 1 does not require all variables are measured simultaneously, and hence we can also apply it to the settings where only pairwise measurements or general size-constrained measurements are available (Dasarathy 2019).

3.3. Variance Estimation and Edge-Wise Inference

In this section, we propose our GI-JOE method for edge-wise statistical inference. That is, we test the null hypothesis: $H_0 : \Theta_{a,b}^* = 0$ against $H_1 : \Theta_{a,b}^* \neq 0$. With the aid of Theorem 2, we still need to estimate the unknown variance $\sigma_n^2(a,b)$ so that we can construct a test statistic with known distribution under H_0.

Recall the definition of $\sigma_n^2(a,b)$ in Theorem 2, and the fact that $\mathcal{T}_{j,k,j',k'}^* = \Sigma_{j,j'}^*\Sigma_{k,k'}^* + \Sigma_{j,k'}^*\Sigma_{k,j'}^*$, $(\mathcal{T}^{(n)*})_{j,k,j',k'} = \mathcal{T}_{j,k,j',k'}^* \frac{n_{j,k,j',k'}}{n_{j,k}n_{j',k'}}$, here we define an estimator $\widehat{\sigma}_n^2(a,b)$ as follows:

Algorithm 1: GI-JOE: edge-wise inference

Input: Dataset $\{x_{i,V_i} : V_i \subset [p]\}_{i=1}^n$, pairwise sample sizes $\{n_{j,k}\}_{j,k=1}^p$, node pair (a, b) for testing with $a \neq b$, significant level α

1. Compute the entry-wise estimate of the covariance matrix $\widehat{\Sigma} \in \mathbb{R}^{p\times p}$: $\widehat{\Sigma}_{j,k} = \frac{1}{n_{j,k}} \sum_{j,k\in V_i} x_{i,j}x_{i,k}$
2. Project $\widehat{\Sigma}$ onto the positive semi-definite cone: compute $\widetilde{\Sigma}$ as in (1).
3. Perform neighborhood regression for node a: compute $\widehat{\theta}^{(a)}$ as in (2)
4. Estimate the debiasing matrix by performing neighborhood regression for node b upon nodes $[p]\backslash\{a, b\}$: compute $\widehat{\Theta}_{b,:}^{(a)}$ as in (5) and (6).
5. Debias the neighborhood lasso estimate: $\widetilde{\theta}_b^{(a)} = \widehat{\theta}_b^{(a)} - \widehat{\Theta}_{b,:}^{(a)}(\widehat{\Sigma}\widehat{\theta}^{(a)} - \widehat{\Sigma}_{:,a})$.
6. Estimate the variance: $\widehat{\sigma}_n^2 = \widehat{\mathcal{T}}^{(n)} \times_1 \widetilde{\Theta}_{b,:}^{(a)} \times_2 \widehat{\widehat{\theta}}^{(a)} \times_3 \widetilde{\Theta}_{b,:}^{(a)} \times_4 \widehat{\widehat{\theta}}^{(a)}$, where
$$(\widehat{\mathcal{T}}^{(n)})_{j,k,j',k'} = (\widetilde{\Sigma}_{j,j'}\widetilde{\Sigma}_{k,k'} + \widetilde{\Sigma}_{j,k'}\widetilde{\Sigma}_{k,j'})\frac{n_{j,k,j',k'}}{n_{j,k}n_{j',k'}},$$
$\widehat{\widehat{\theta}}^{(a)}$ is defined as $\widehat{\widehat{\theta}}_a^{(a)} = 1$ and $\widehat{\widehat{\theta}}_{\backslash a}^{(a)} = -\widehat{\theta}_{\backslash a}^{(a)}$, and $\widetilde{\Theta}_{b,:}^{(a)}$ is computed as in (5) and (6).
7. Compute p-value $p_{a,b} = 2(1 - \Phi(\frac{\widetilde{\theta}_b^{(a)}}{\widehat{\sigma}_n(a,b)}))$ where $\Phi(\cdot)$ is the distribution function of standard Gaussian; confidence interval $\widehat{\mathbb{C}}_\alpha^{a,b} = [\widetilde{\theta}_b^{(a)} - z_{\alpha/2}\widehat{\sigma}_n(a, b), \widetilde{\theta}_b^{(a)} + z_{\alpha/2}\widehat{\sigma}_n(a, b)]$

Output: p-value $p_{a,b}$, confidence interval $\widehat{\mathbb{C}}_\alpha^{a,b}$ for $\theta_{a,b}^{(a)*} - \frac{\Theta_{a,b}^*}{\Theta_{a,a}^*}$.

$$\widehat{\sigma}_n^2(a, b) = \widehat{\mathcal{T}}^{(n)} \times_1 \widetilde{\Theta}_{b,:}^{(a)} \times_2 \widehat{\widehat{\theta}}^{(a)} \times_3 \widetilde{\Theta}_{b,:}^{(a)} \times_4 \widehat{\widehat{\theta}}^{(a)}, \tag{10}$$

where $\widehat{\mathcal{T}}^{(n)}$ is an estimator for $\mathcal{T}^{(n)*}$: $(\widehat{\mathcal{T}}^{(n)})_{j,k,j',k'} = (\widetilde{\Sigma}_{j,j'}\widetilde{\Sigma}_{k,k'} + \widetilde{\Sigma}_{j,k'}\widetilde{\Sigma}_{k,j'})\frac{n_{j,k,j',k'}}{n_{j,k}n_{j',k'}}$; $\widehat{\widehat{\theta}}^{(a)} \in \mathbb{R}^p$ serves as an estimate for $\frac{\Theta_{:,a}^*}{\Theta_{a,a}^*}$ and it satisfies $\widehat{\widehat{\theta}}_a^{(a)} = 1$ and $\widehat{\widehat{\theta}}_{\backslash a}^{(a)} = -\widehat{\theta}_{\backslash a}^{(a)}$; $\widetilde{\Theta}_{b,:}^{(a)}$ is defined in Section 3.1 and serves an estimate for $\Theta_{b,:}^{(a)*}$.

Assumption 4 (Sample size condition for variance estimation).

$$n_1^{(a,b)} \gg \frac{C^4(\Theta^*; a, b)}{\kappa_{\Theta^*}^4}(d_a + d_b + 1)^2 \log p \left(\frac{n_2^{(a,b)}}{n_1^{(a,b)}}\right)^2,$$

where $C(\Theta^*; a, b)$ is defined as in Assumption 3.

Proposition 3 (Estimation consistency of variance). Under Assumptions 1, 3, and 4, if the tuning parameters are as specified in Theorem 2, then (10) satisfies $\frac{\widehat{\sigma}_n^{-1}(a,b)}{\sigma_n^{-1}(a,b)} \xrightarrow{p} 1$.

Theorem 3 (Normal approximation with unknown variance). With appropriately chosen tuning parameters as in Theorem 2, if Assumptions 1–4 hold, $\widehat{\sigma}_n^{-1}(a, b)(\widetilde{\theta}_b^{(a)} - \theta_b^{(a)*}) \xrightarrow{\mathrm{d}} \mathcal{N}(0, 1)$ for $\widehat{\sigma}_n^2(a, b)$ defined in (10) and $\widetilde{\theta}_b^{(a)}$ defined in (7).

The proof of Theorem 3 can be found in Section G of the supplementary material. Theorem 3 suggests us to construct the test statistic $\widehat{z}(a, b) = \widehat{\sigma}_n^{-1}(a, b)\widetilde{\theta}_b^{(a)}$, and for a desired Type I error α, wereject $H_0 : \Theta_{a,b}^* = 0$ if $|\widehat{z}(a, b)| \geq z_{\alpha/2}$, where $z_{\alpha/2}$ is the $1 - \frac{\alpha}{2}$ quantile of standard Gaussian distribution. Our full GI-JOE (edge-wise inference) procedure is summarized in Algorithm 1, with its Type I error and power characterized by the following theorem.

Theorem 4 (Type I error and Power Analysis). Consider the Gaussian graphical model with erose measurement setting described in Section 1.1 and let $p_{a,b}$ be the p-value given by Algorithm 1 for node pair (a, b). If all conditions in Theorem 3 hold so that as $n, p \to \infty$, p, d_a, d_b, $n_1^{(a,b)}$, $n_2^{(a,b)}$ scale as in Assumptions 3 and 4, then the following holds:

1. Under the null hypothesis $H_0 : \Theta_{a,b}^* = 0$, $\lim_{n,p\to\infty} \mathbb{P}(p_{a,b} \leq \alpha) = \alpha$;
2. Under the alternative hypothesis $H_1 : \frac{\Theta_{a,b}^*}{\Theta_{a,a}^*} = \delta_n$,
   (a) if $\lim_{n,p\to\infty} \frac{\delta_n}{\sigma_n(a,b)} = 0$, $\lim_{n,p\to\infty} \mathbb{P}(p_{a,b} \leq \alpha) = \alpha$;
   (b) if $\lim_{n,p\to\infty} \frac{\delta_n}{\sigma_n(a,b)} = \delta$ for $\delta \neq 0$, $\lim_{n,p\to\infty} \mathbb{P}(p_{a,b} \leq \alpha) \geq \Phi(|\delta| - z_{\alpha/2})$, where $\Phi(\cdot)$ is the distribution function of standard Gaussian $\mathcal{N}(0, 1)$;
   (c) if $\lim_{n,p\to\infty} \frac{\delta_n}{\sigma_n(a,b)} = +\infty$, $\lim_{n,p\to\infty} \mathbb{P}(p_{a,b} \leq \alpha) = 1$.

The proof of Theorem 4 is deferred to Section G of the supplementary material. Theorem 4 suggests that when all conditions of Theorem 3 hold, the Type I error of this test is asymptotically α. Furthermore, as long as the signal strength $\frac{\Theta_{a,b}^*}{\Theta_{a,a}^*}$ shrinks no faster than $\sigma_n(a, b) \asymp (n_2^{(a,b)})^{-1/2}$, we can still reject the null with constant or high probability. Other than hypothesis testing, we can also construct asymptotically valid confidence intervals for each entry of the precision matrix $(\Theta_{a,b}^*)$, under similar assumptions to Theorem 3. More details can be found in Section F of the supplementary material.

4. FDR Control for Graph Inference with Erose Measurements

In many application scenarios, the inference of the full graph may be of more interest than the inference of one particular edge. Confronted with a multiple testing problem, we can simply apply Holm's correction upon the p-values of all $\frac{p(p-1)}{2}$ node pairs (a, b) for $a < b$, and hence control the family-wise error rate. However, as this approach can be too conservative, here we also propose an FDR control procedure. We leverage the ideas from Javanmard and Javadi (2019) and Liu (2013) which consider the FDR control for the debiased lasso and Gaussian graphical models. In particular, for any $0 \leq \rho \leq 1$, let $R(\rho) = \sum_{i<j} \mathbb{1}_{\{p_{i,j}\leq\rho\}}$ be the number of significant edges when ρ is the threshold for p-values. Also define $t_p = (2\log(p(p-1)/2) - 2\log\log(p(p-1)/2))^{\frac{1}{2}}$, and if there exists $2(1 - \Phi(t_p)) \leq \rho \leq 1$ such that $\frac{p(p-1)\rho}{2R(\rho)\vee 1} \leq \alpha$, the nominal level, then we would define $\rho_0 = \sup_{2(1-\Phi(t_p))\leq\rho\leq 1}\left\{\rho : \frac{p(p-1)\rho}{2(R(\rho)\vee 1)} \leq \alpha\right\}$; otherwise, $\rho_0 = 2(1 - \Phi(\sqrt{2\log(p(p-1)/2)}))$. The significant edge set is

then defined as $\widetilde{E} = \{(j,k) : p_{j,k} \le \rho_0\}$. This is similar to the Benjamini-Hochberg procedure with only an extra truncation step. The full procedure is summarized in Algorithm 2 in the supplementary material.

In the following, we provide theoretical guarantees for our GI-JOE (FDR) approach. Define $\epsilon_i(a,b) = \sum_{j,k}(x_{i,j}x_{i,k} - \Sigma^*_{j,k})\frac{\delta^{(i)}_{j,k}}{n_{j,k}}\frac{\Theta^{(a)*}_{j,b}\Theta^*_{k,a}+\Theta^{(a)*}_{k,b}\Theta^*_{j,a}}{2\Theta^*_{a,a}}$ as the error for estimating edge (a,b), contributed by the ith sample; $\xi_i(a,b)$ is the normalized error: $\xi_i(a,b) = \frac{\epsilon_i(a,b)}{\sigma_n(a,b)}$. One technical quantity that is useful in our proofs is $\alpha(\Theta^*, \{V_i\}_{i=1}^n) = \sup_{(a,b),(a',b')} \frac{\|\xi_i(a,b)\|^2_{\psi_1} \vee \|\xi_i(a',b')\|^2_{\psi_1}}{\lambda_{\min}(\mathrm{Cov}((\xi_i(a,b),\xi_i(a',b'))))}$. We want $\alpha(\Theta^*, \{V_i\}_{i=1}^n)$ to be not too large, similar to assuming $(\xi_i(a,b),\xi_i(a',b'))$ to be far from degenerate. Also define the covariance between the test statistics of different edges: $\sigma_n^2(a,b,a',b') = \frac{1}{\Theta^*_{a,a}\Theta^*_{a',a'}}\Gamma^{(n)*} \times_1 \Theta^{(a)*}_{:,b} \times_2 \Theta^*_{:,a} \times_3 \Theta^{(a')*}_{:,b'} \times_4 \Theta^*_{:,a'}$, and the correlation $\rho_n(a,b,a',b') = \frac{\sigma_n^2(a,b,a',b')}{\sigma_n(a,b)\sigma_n(a',b')}$.

Assumption 5. For any edge $(a,b) \in [p] \times [p]$, $n_2^{(a,b)} \ge \frac{C\|\Sigma^*\|^6_\infty \alpha^2(\Theta^*,\{V_i\}_{i=1}^n)}{\lambda^6_{\min}(\Sigma^*)}(d+1)^6(\log p)^6$, and $n_1^{(a,b)} \gg Cd^2(\log p)^5 \log\log p \left(\frac{n_2^{(a,b)}}{n_1^{(a,b)}}\right)^2$, where $d = \max_{a\in[p]} d_a$.

Assumption 5 is stronger than the sample size requirements in Assumptions 3 and 4 for edge-wise inference, since the proof for asymptotically valid FDR control needs stronger normal approximation guarantees, especially at the tail.

Remark 6. When all variables are simultaneously measured with sample size n, Assumption 5 reduces to $n \ge C(d+1)^6(\log p)^6$. This is much weaker than the condition in the literature of the graphical model FDR control (Liu 2013): $p \le n^r$ for some constant $r > 0$. They used this assumption to show the tail probability of the test statistics can be well approximated by the Gaussian tail, while we use a different proof that exploits the sub-exponential properties of $\widehat{\Sigma}_{j,k}$ (second moments of Gaussian variables are sub-exponential).

Assumption 6. For any $0 < \rho < 1$, $\gamma > 0$, let $\mathcal{A}_1(\rho) = \{(a,b,a',b') \in [p]\times[p] : \Theta^*_{a,b} = \Theta^*_{a',b'} = 0, |\rho_n(a,b,a',b')| > \rho\}$, and $\mathcal{A}_2(\rho,\gamma) = \{(a,b,a',b') \in [p]\times[p] : \Theta^*_{a,b} = \Theta^*_{a',b'} = 0, (\log p)^{-2-\gamma} < |\rho_n(a,b,a',b')| \le \rho\}$. There exist $0 < \rho_0 < 1$ and $\gamma > 0$ such that, $|\mathcal{A}_1(\rho_0)| \le Cp^2$, $|\mathcal{A}_2(\rho_0,\gamma)| \ll p^{\frac{4}{1+\rho_0}}(\log p)^{\frac{2\rho_0}{1+\rho_0}-\frac{1}{2}}(\log\log p)^{-\frac{1}{2}}$.

Assumption 6 enforces that most edge-wise test statistics are only weakly correlated, and similar assumptions have also appeared in Liu (2013) and Javanmard and Javadi (2019). To provide more intuition and to justify this Assumption, we prove that in the simultaneous measurement setting, this assumption holds when each node has a constant number of strongly connected neighbors; we further empirically validate this Assumption for many sparse graphs and erose measurement patterns. More details are included in Section C of the supplementary material. The following theorem suggests that our procedure can successfully control the false discovery proportion both in expectation and in probability when Assumptions 5 and 6 hold.

Theorem 5 (Validity of FDR control). Consider the GI-JOE (FDR) procedure described in Section 4 and suppose the significant edge set under a given nominal level α is given by $\widetilde{E}$. Let $\mathrm{FDP} = \frac{\sum_{(a,b)\in\mathcal{H}_0}\mathbb{1}_{\{(a,b)\in\widetilde{E}\}}}{|\widetilde{E}|\vee 1}$, where $\mathcal{H}_0 = \{(i,j) \in [p]\times[p] : \Theta^*_{i,j} = 0\}$; Also let $\mathrm{FDR} = \mathbb{E}\mathrm{FDP}$. If Assumptions 5 and 6 hold, then we have $\limsup_{n,p\to\infty} \mathrm{FDR} \le \alpha$, and for any $\epsilon > 0$, $\lim_{n,p\to\infty}\mathbb{P}(\mathrm{FDP} > \alpha + \epsilon) = 0$.

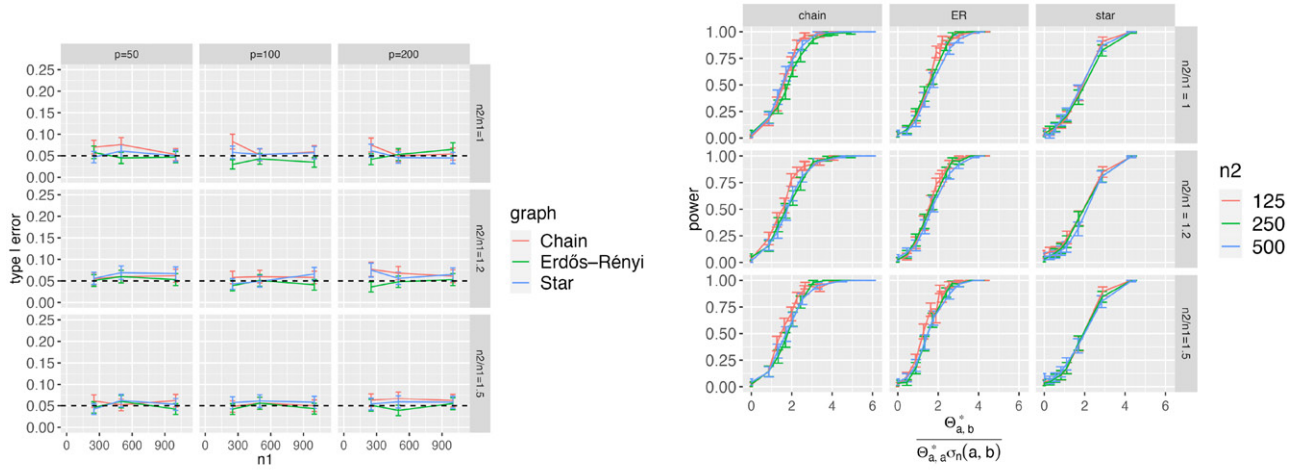
The proof of Theorem 5 can be found in Section G of the supplementary material. This is the first theoretical guarantee for FDR control with erosely measured data. Although we still require sufficient pairwise sample sizes for all pairs of nodes, we allow $n_{j,k}$ to be of different order. As shown in Section G of the supplementary material, our GI-JOE FDR approach indeed controls the FPR empirically in a wide range of erose measurement settings and graph structures, even when the sample size condition might be violated.

5. Empirical Studies

In this section, we present empirical studies to validate our GI-JOE approach for both edge-wise inference and full graph inference. We first verify the validity of our edge-wise inference procedure in Section 5.1; then we compare the graph selection performance using both our GI-JOE approaches and various baseline estimation and inference methods in Section 5.2. A real data example is included in Section 5.3.

5.1. Simulations for Edge-Wise Inference: Validating Theory

Here, we investigate the Type I error and power of GI-JOE for testing one node pair. To study the effect of different pairwise sample sizes, here we consider the pairwise measurement scenario where each sample only consists of two variables. When the given node pair for inference is (a,b), we set the pairwise sample size as follows: $n_{j,k} = \sum_{i=1}^n \mathbb{1}_{\{j,k\in V_i\}} = n_1$ if $(j,k) \in S_1(a,b)\backslash S_2(a,b)$, $n_{j,k} = n_2$ if $(j,k) \in S_2(a,b)$, and $n_{j,k} = 50$ otherwise. The precision matrix $\Theta^* = \Sigma^{*-1}$ is generated with three graph structures: a chain graph, a three-star graph, and an Erdős–Rényi graph with connection probability $\frac{3}{p-1}$. Then we study the Type I error for testing an unconnected node pair, and the power for testing an edge with different signal strengths. More details on the set-ups and the implementation of our GI-JOE approach can be found in Section E.1 of the supplementary material. Figure 3 summarizes the Type I error rate and power averaged over 200 replicates when the confidence level is set as 0.95, under three graph structures. In the Type I error plot (a), we consider three network sizes $p = 50, 100, 200$, and a range of $n_1^{(a,b)}$, $n_2^{(a,b)}$. We can see that the Type I error rates are close to 0.05 with moderately large, although differing, pairwise sample sizes. In the power plot, The dimension p is fixed as 200, sample sizes $n_2^{(a,b)} = n2 \in \{125, 250, 500\}$, $n_1^{(a,b)} = n_1 \in n_2/\{1, 1.2, 1.5\}$. The dependence of power on the SNR $\frac{\Theta^*_{a,b}}{\Theta^*_{a,a}\sigma_n(a,b)}$ is similar across different sample sizes, supporting Theorem 4.



(a) Type I Error (b) Power

Figure 3. Type I error rates versus sample size, averaged over 1000 replicates; Power versus signal to noise ratio, averaged over 200 replicates. Target Type I error rate is set as $\alpha = 0.05$, and the error bars represent the 95% confidence interval. The x-axis in (b) is the signal-to-noise ratio $\frac{\Theta^*_{a,b}}{\Theta^*_{a,a}\sigma_n(a,b)}$, which determines the asymptotic power of our test (see Theorem 4).

5.2. Graph Selection Study and Comparison with Baselines

Now we study the graph selection performance of our GI-JOE (Holm) and GI-JOE (FDR) approaches under different erose measurements, compared with some baselines. In particular, we consider four estimation methods and four inference methods. The estimation methods include standard plug-in methods (Kolar and Xing 2012; Park, Wang, and Lim 2021) with neighborhood lasso (Nlasso), graphical lasso (Glasso), CLIME, and the variant plug-in method described in Section 2 (Nlasso-JOE). The only difference between Nlasso-JOE and Nlasso is the former uses a different tuning parameter for each node that depends on its own sample sizes, as explained in Section 2. The inference methods include GI-JOE (Holm), GI-JOE (FDR), and also ad hoc implementations of the debiased graphical lasso (Jankova and Van De Geer 2015). Specifically, since there are no applicable inference methods designed for erose measurements, the only baseline we can implement is applying existing inference methods for simultaneous measurement settings and plug in the minimum pairwise sample size for computing the variance of each edge. This is not a method one would ever use in practice, since it is too conservative and has no theoretical guarantees, but we still present the results of such baselines, just to prove the concept that considering the different sample sizes over the graph is important. Although Jankova and Van De Geer (2015) is only concerned with edge-wise inference, we still add a Holm's correction and FDR control procedure on top of its edge-wise p-values for a fair comparison, and we denote these two procedures by DB-Glasso (Holm) and DB-Glasso (FDR). Some additional implementation details of these methods can be found in Section D.2 of the supplementary material.

The comparative studies here include both synthetic and real data-inspired simulations, and we first present the synthetic setup. For graph structures, we consider the chain graph, 10-star graph, and Erdős–Rényi graph, with dimension $p = 200$. We experiment with three synthetic erose measurement patterns, but due to space limit, here we only present the results for two measurement patterns and defer more complete results to the supplementary material. In measurement 1, each node is independently missing with low, moderate, or high probabilities; measurement 2 is the size-constrained measurements scenario, where each sample consists of randomly selected 20 nodes, and each node is sampled with probability weight positively depending on its degree. For each measurement scenario, we consider two different total sample sizes. Table 1 summarizes the F1 scores of our GI-JOE method and some other baseline estimation and inference methods, averaged over 20 independent runs (standard deviations included in parentheses). For both Nlasso and Nlasso-JOE, we present the results based on the AND rule; The results for OR rule and more detailed results on true positive rate, true negative rate, and true discovery rate can be found in Section E of the supplementary material. In summary, among all different measurement scenarios, sample sizes, and graph structures, GI-JOE (FDR) is either the best or comparable to the best among all inference and estimation methods, in terms of the F1 score. The Nlasso-JOE is usually the best among the estimation methods, suggesting that using different tuning parameters that accommodate for the different pairwise sample sizes may be a simple yet effective trick.

While for the real data-inspired simulations, either the graph structure is adopted from real neuroscience data or the measurement patterns are adopted from real gene expression datasets. Due to space limit, here we only present the latter (real erose measurements) but leave the real graph simulation results in the supplementary material. For the real measurement patterns, we take two publicly available single-cell RNA sequencing datasets (Darmanis et al. 2015; Chu et al. 2016), and focus on the top 200 genes with highest variances. The *chu* and *darmanis* measurement patterns have pairwise sample sizes ranging from 0 to 1018 and from 5 to 366. The graphs for data generation are

Table 1. F1 scores of the graphs selected by estimation methods (first three) and inference methods (last four) under two measurement scenarios.

Method	Chain graph				10-Star graph			
	Measurement 1		Measurement 2		Measurement 1		Measurement 2	
	n=600	n=800	n=20,000	n=30,000	n=600	n=800	n=20,000	n=30,000
Nlasso	0.62(0.18)	0.70(0.02)	0.41(0.01)	0.34(0.01)	0.35(0.11)	0.40(0.10)	0.28(0.01)	0.29(0.01)
Glasso	0.37(0.01)	0.35(0.01)	0.31(0.01)	0.47(0.21)	0.49(0.02)	0.47(0.01)	0.35(0.01)	0.39(0.25)
CLIME	0.40(0.04)	0.32(0.01)	0.27(0.00)	0.44(0.01)	0.25(0.01)	0.39(0.03)	0.28(0.03)	0.43(0.02)
Nlasso-JOE	**0.64**(0.12)	**0.91**(0.05)	**0.68**(0.20)	**0.88**(0.01)	**0.82**(0.07)	**0.86**(0.02)	**0.80**(0.02)	**0.90**(0.01)
DB-Glasso (Holm)	0.01(0.01)	0.05(0.02)	0.07(0.02)	0.42(0.03)	0.00(0.01)	0.00(0.01)	0.00(0.01)	0.04(0.01)
DB-Glasso (FDR)	0.00(0.01)	0.01(0.01)	0.01(0.01)	0.06(0.02)	0.00(0.01)	0.01(0.01)	0.01(0.01)	0.06(0.02)
GI-JOE (Holm)	0.78(0.01)	0.81(0.01)	0.37(0.12)	0.67(0.02)	0.32(0.04)	0.50(0.01)	0.96(0.01)	0.97(0.01)
GI-JOE (FDR)	**0.81**(0.01)	**0.86**(0.01)	**0.74**(0.08)	**0.92**(0.01)	**0.51**(0.03)	**0.61**(0.01)	**0.99**(0.01)	**0.98**(0.01)

NOTE: The average sample size $\frac{1}{p^2}\sum_{j,k} n_{j,k}$ ranges from 50 to 350. The highest F1 scores are in bold.

Table 2. F1 scores of estimation methods (first three) and inference methods (last four) with the ground truth graphs being a scale-free graph and a small-world graph with 200 nodes, under two real measurement patterns from single-cell RNA sequencing datasets (the chu data (Chu et al. 2016) and darmanis data (Darmanis et al. 2015)).

Method	*chu* measurement		*darmanis* measurement	
	Scale-free graph	Small-world graph	Scale-free graph	Small-world graph
Nlasso	0.54(0.25)	0.57(0.28)	0.31(0.02)	0.34(0.11)
Glasso	**0.61**(0.19)	0.70(0.17)	**0.59**(0.03)	0.49(0.10)
CLIME	0.41(0.07)	0.40(0.02)	0.26(0.04)	0.35(0.12)
Nlasso-JOE	**0.61**(0.30)	**0.92**(0.01)	0.51(0.03)	**0.81**(0.01)
DB-Glasso (Holm)	0.00(0.00)	0.00(0.00)	0.00(0.00)	0.00(0.00)
DB-Glasso (FDR)	0.00(0.00)	0.00(0.00)	0.00(0.00)	0.00(0.00)
GI-JOE (Holm)	**0.96**(0.01)	**0.93**(0.01)	0.44(0.04)	0.55(0.03)
GI-JOE (FDR)	0.94(0.03)	0.93(0.01)	**0.75**(0.03)	**0.76**(0.02)

NOTE: Our GI-JOE methods always have the highest F1-scores, and GI-JOE (FDR) is better for the darmanis measurement (average sample size 250) while GI-JOE (Holm) is better for the chu measurement (average sample size 850). The highest F1 scores are in bold.

scale-free graphs and small-world graphs with 200 nodes. The F1-scores summarized in Table 2 also suggest the efficacy of the GI-JOE (FDR) approach. Visualizations of the real graph and measurement patterns, and specifics of the simulated graphs can be found in Section F of the supplementary material.

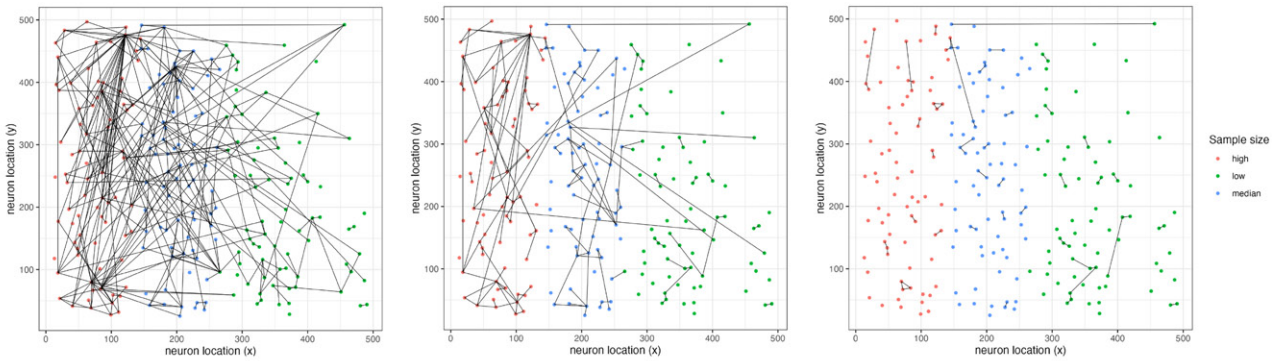
5.3. Real Data Example: Application to Calcium Imaging Data

The two-photon calcium imaging technology can record in vivo functional activities of thousands of neurons (Stringer et al. 2019), and such datasets can be used to understand the neuronal circuits with the help of graphical model techniques (Vinci, Dasarathy, and Allen 2019; Wang and Allen 2021). In this section, we investigate the potential of our GI-JOE approach on a real calcium imaging dataset from the Allen Institute (Lein et al. 2007), which contains the functional recordings of $p = 227$ neurons in a mouse's visual cortex, when different visual stimuli or no stimulus were presented to the mouse. Here, we focus on the raw fluorescence traces in one spontaneous session with no external stimulus. This session includes $n = 8931$ samples of the trace data associated with all 227 neurons. We manually mask the data for some neurons to create erose measurements, and then validate our methods by comparing the tested graph based on masked data with the tested graph based on the full dataset. In particular, the recorded neurons all lie on the same vertical plane in a mouse's visual cortex (see Figure 4 for the physical locations of the neurons in x and y axis). To manually create erose measurements, we divide the neurons into three subsets based on their location on the x-axis (marked in different colors in Figure 4), and neurons in each subset are randomly observed with high ($\sqrt{0.9}$), moderate ($\sqrt{0.5}$), and low ($\sqrt{0.1}$) probabilities.

However, when inference methods are applied on the full dataset, we find that the tested graph is always dense. This might be due to the huge amount of latent neurons in the mouse's brain since latent variables are known to lead to dense graph structures in graphical models (Wang and Allen 2021). Since most of these edges have small edge weights, here we consider testing if the edge weights are stronger than a threshold instead of testing if they are zero. Specifically, for any node pair (a, b), $H_{0,(a,b)} : |\frac{\Theta^*_{a,b}}{\Theta^*_{a,a}}| \leq 0.12$. Our GI-JOE approach can be directly extended to test such hypothesis and the detailed procedures are included in Section D of the supplementary material. For validation purposes, we also apply a special version of our GI-JOE (FDR) approach on the full dataset, where all pairwise sample sizes are equal. As suggested by Figure 4, our GI-JOE (FDR) approach works well, especially for the neuron set 1 (red) as they have larger sample sizes; it identifies the same hub neuron in set 1 as the tested graph (a) with the full dataset. The specific F1-scores of each method for each neuron set can be found in Section E of the supplementary material.

6. Discussion

In this article, we propose the *GI-JOE* (Graph Inference when Joint Observations are Erose) approach to address the graphical model inference problem under the *erose* measurement setting, where irregular observational patterns lead to vastly different sample sizes for node pairs. Our GI-JOE approach quantifies the



(a) FDR-selected graph with full data

(b) GI-JOE (FDR), applied to data with erose measurements

(c) Debiased graphical lasso with minimum sample size, applied to data with erose measurements

Figure 4. Tested functional connectivity graphs among neurons, using the Allen Brain Atlas dataset. The neurons marked in red, blue, and green have high, moderate and low sample sizes. Our GI-JOE (FDR) approach works reasonably well, especially for red neurons, while the debiased graphical lasso with minimum sample size is too conservative to find any edge.

different uncertainty levels over the graph induced by the erose measurements, including both an edge-wise inference method and an FDR control procedure. We characterize the Type I error and power for testing any node pair (a, b) by considering the sample sizes involving a, b's neighbors; We also guarantee the valid FDR control of GI-JOE (FDR) under appropriate conditions. Finally, our experiments with synthetic and real data demonstrate the efficacy of our approach for different graphs and measurement patterns.

There are still many open questions related to graph learning with erose measurements that may be worth future investigation. For instance, it is possible to extend our theory and methods to general sub-Gaussian data or semiparametric graphical models. Our problem setting is closely related to the latent variable graphical models when some pairwise sample sizes are extremely low, and some ideas in this line of work might be useful to further improve our method with relaxed sample size conditions. In addition, our current results are based on a variant of the neighborhood lasso, while one may also consider a variant of the graphical lasso or CLIME and investigate their potential in this setting. Another practical but challenging setting not considered here is data-dependent erose measurements, which requires novel methods and theory since the sample covariance would be biased and the plug-in type approach no longer works. Furthermore, erosely measured data sometimes exhibits temporal dependence and thus calls for new inference methodologies.

Supplementary Materials

The Supplementary material includes all the technical proofs, additional theoretical and empirical results. Software and code for reproducing the empirical results can be found at *https://github.com/Lili-Zheng-stat/GI-JOE*.

Acknowledgments

The authors also thank Gautam Dasarathy for helpful discussion during the preparation of this article.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

The authors gratefully acknowledge support by NSF NeuroNex-1707400, NIH 1R01GM140468, and NSF DMS-2210837.

References

Allen, G. I., and Liu, Z. (2012), "A Log-Linear Graphical Model for Inferring Genetic Networks from High-throughput Sequencing Data," in *2012 IEEE International Conference on Bioinformatics and Biomedicine*, IEEE, pp. 1–6. [2282]

Bae, J. A., Baptiste, M., Bodor, A. L., Brittain, D., Buchanan, J., Bumbarger, D. J., Castro, M. A., Celii, B., Cobos, E., Collman, F., et al. (2021), "Functional Connectomics Spanning Multiple Areas of Mouse Visual Cortex," BioRxiv . [2283]

Barber, R. F., and Candès, E. J. (2015), "Controlling the False Discovery Rate via Knockoffs," *The Annals of Statistics*, 43, 2055–2085. [2283]

Belloni, A., Chernozhukov, V., and Kaul, A. (2017), "Confidence Bands for Coefficients in High Dimensional Linear Models with Error-in-Variables," arXiv preprint arXiv:1703.00469 . [2284]

Cai, T., Cai, T. T., and Zhang, A. (2016), "Structured Matrix Completion with Applications to Genomic Data Integration," *Journal of the American Statistical Association*, 111, 621–633. [2283]

Cai, T., Liu, W., and Luo, X. (2011), "A Constrained ℓ_1 Minimization Approach to Sparse Precision Matrix Estimation," *Journal of the American Statistical Association*, 106, 594–607. [2282,2284]

Candès, E., Fan, Y., Janson, L., Lv, J., et al. (2018), "Panning for Gold: 'model-x' Knockoffs for High Dimensional Controlled Variable Selection," *Journal of the Royal Statistical Society*, Series B, 80, 551–577. [2284]

Chang, A., and Allen, G. I. (2021), "Extreme Graphical Models with Applications to Functional Neuronal Connectivity," arXiv preprint arXiv:2106.11554 . [2283]

Chen, S., Witten, D. M., and Shojaie, A. (2015), "Selection and Estimation for Mixed Graphical Models," *Biometrika*, 102, 47–64. [2282]

Chu, L.-F., Leng, N., Zhang, J., Hou, Z., Mamott, D., Vereide, D. T., Choi, J., Kendziorski, C., Stewart, R., and Thomson, J. A. (2016), "Single-Cell RNA-seq Reveals Novel Regulators of Human Embryonic Stem Cell Differentiation to Definitive Endoderm," *Genome Biology*, 17, 1–20. [2283,2290,2291]

Darmanis, S., Sloan, S. A., Zhang, Y., Enge, M., Caneda, C., Shuer, L. M., Hayden Gephart, M. G., Barres, B. A., and Quake, S. R. (2015), "A Survey of Human Brain Transcriptome Diversity at the Single Cell Level," *Proceedings of the National Academy of Sciences*, 112, 7285–7290. [2283,2290,2291]

Dasarathy, G. (2019), "Gaussian Graphical Model Selection from Size Constrained Measurements," in *2019 IEEE International Symposium on Information Theory (ISIT)*, IEEE, pp. 1302–1306. [2282,2283,2287]

Dasarathy, G., Singh, A., Balcan, M.-F., and Park, J. H. (2016), "Active Learning Algorithms for Graphical Model Selection," in *Artificial Intelligence and Statistics*, PMLR, pp. 1356–1364. [2282,2283]

Dobra, A., Hans, C., Jones, B., Nevins, J. R., Yao, G., and West, M. (2004), "Sparse Graphical Models for Exploring Gene Expression Data," *Journal of Multivariate Analysis*, 90, 196–212. [2282]

Dobra, A., and Lenkoski, A. (2011), "Copula Gaussian Graphical Models and their Application to Modeling Functional Disability Data," *The Annals of Applied Statistics*, 5, 969–993. [2282]

Friedman, J., Hastie, T., and Tibshirani, R. (2008), "Sparse Inverse Covariance Estimation with the Graphical Lasso," *Biostatistics*, 9, 432–441. [2282]

Gan, L., Vinci, G., and Allen, G. I. (2020), "Correlation Imputation in Single Cell RNA-seq Using Auxiliary Information and Ensemble Learning," in *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, pp. 1–6. [2282]

Gong, W., Kwak, I.-Y., Pota, P., Koyano-Nakagawa, N., and Garry, D. J. (2018), "Drimpute: Imputing Dropout Events in Single Cell RNA Sequencing Data," *BMC Bioinformatics*, 19, 1–10. [2282]

Gu, Q., Cao, Y., Ning, Y., and Liu, H. (2015), "Local and Global Inference for High Dimensional Gaussian Copula Graphical Models," arXiv preprint arXiv:1502.02347. [2284]

Huang, M., Wang, J., Torre, E., Dueck, H., Shaffer, S., Bonasio, R., Murray, J. I., Raj, A., Li, M., and Zhang, N. R. (2018), "Saver: Gene Expression Recovery for Single-Cell RNA Sequencing," *Nature Methods*, 15, 539–542. [2282]

Jankova, J., and Van De Geer, S. (2015), "Confidence Intervals for High-Dimensional Inverse Covariance Estimation," *Electronic Journal of Statistics*, 9, 1205–1229. [2283,2284,2287,2290]

Janková, J., and van de Geer, S. (2017), "Honest Confidence Regions and Optimality in High-Dimensional Precision Matrix Estimation," *Test*, 26, 143–162. [2284]

Javanmard, A., and Javadi, H. (2019), "False Discovery Rate Control via Debiased Lasso," *Electronic Journal of Statistics*, 13, 1212–1253. [2284,2288,2289]

Javanmard, A., and Montanari, A. (2014), "Confidence Intervals and Hypothesis Testing for High-Dimensional Regression," *The Journal of Machine Learning Research*, 15, 2869–2909. [2283]

Kolar, M., and Xing, E. P. (2012), "Estimating Sparse Precision Matrices from Data with Missing Values," in *Proceedings of the 29th International Coference on International Conference on Machine Learning*, pp. 635–642. [2283,2290]

Koller, D., and Friedman, N. (2009), *Probabilistic Graphical Models: Principles and Techniques*, Cambridge, MA: MIT Press. [2282]

Lee, J. D., Sun, D. L., Sun, Y., and Taylor, J. E. (2016), "Exact Post-Selection Inference, with Application to the Lasso," *The Annals of Statistics*, 44, 907–927. [2283]

Lein, E. S., Hawrylycz, M. J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A. F., Boguski, M. S., Brockway, K. S., Byrnes, E. J. et al. (2007), "Genome-Wide Atlas of Gene Expression in the Adult Mouse Brain," *Nature*, 445, 168–176. [2291]

Liu, H., Han, F., Yuan, M., Lafferty, J., and Wasserman, L. (2012), "High-Dimensional Semiparametric Gaussian Copula Graphical Models," *The Annals of Statistics*, 40, 2293–2326. [2282]

Liu, H., Lafferty, J., and Wasserman, L. (2009), "The Nonparanormal: Semiparametric Estimation of High Dimensional Undirected Graphs," *Journal of Machine Learning Research*, 10, 2295–2328. [2282]

Liu, W. (2013), "Gaussian Graphical Model Estimation with False Discovery Rate Control," *The Annals of Statistics*, 41, 2948–2978. [2284,2288,2289]

Meinshausen, N., and Bühlmann, P. (2006), "High-Dimensional Graphs and Variable Selection with the Lasso," *The Annals of Statistics*, 34, 1436–1462. [2282]

Park, S., Wang, X., and Lim, J. (2021), "Estimating High-Dimensional Covariance and Precision Matrices under General Missing Dependence," *Electronic Journal of Statistics*, 15, 4868–4915. [2283,2290]

Ravikumar, P., Wainwright, M. J., Raskutti, G., and Yu, B. (2011), "High-Dimensional Covariance Estimation by Minimizing ℓ_1-Penalized Log-Determinant Divergence," *Electronic Journal of Statistics*, 5, 935–980. [2282]

Ren, Z., Sun, T., Zhang, C.-H., and Zhou, H. H. (2015), "Asymptotic Normality and Optimalities in Estimation of Large Gaussian Graphical Models," *The Annals of Statistics*, 43, 991–1026. [2284]

Städler, N., and Bühlmann, P. (2012), "Missing values: Sparse Inverse Covariance Estimation and an Extension to Sparse Regression," *Statistics and Computing*, 22, 219–235. [2283]

Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C. B., Carandini, M., and Harris, K. D. (2019), "Spontaneous Behaviors Drive Multidimensional, Brainwide Activity," *Science*, 364, eaav7893. [2291]

Tibshirani, R. J., Taylor, J., Lockhart, R., and Tibshirani, R. (2016), 'Exact Post-Selection Inference for Sequential Regression Procedures," *Journal of the American Statistical Association*, 111, 600–620. [2283]

Van de Geer, S., Bühlmann, P., Ritov, Y., and Dezeure, R. (2014), "On Asymptotically Optimal Confidence Regions and Tests for High-Dimensional Models," *The Annals of Statistics*, 42, 1166–1202. [2283,2285,2287]

Vinci, G., Dasarathy, G., and Allen, G. I. (2019), "Graph Quilting: Graphical Model Selection from Partially Observed Covariances," arXiv preprint arXiv:1912.05573. [2291]

Vinci, G., Ventura, V., Smith, M. A., and Kass, R. E. (2018), "Adjusted Regularization in Latent Graphical Models: Application to Multiple-Neuron Spike Count Data," *The Annals of Applied Statistics*, 12, 1068. [2282,2283]

Wang, H., Fazayeli, F., Chatterjee, S., and Banerjee, A. (2014), "Gaussian Copula Precision Estimation with Missing Values," in *Artificial Intelligence and Statistics*, PMLR, pp. 978–986. [2283]

Wang, M., and Allen, G. I. (2021), "Thresholded Graphical Lasso Adjusts for Latent Variables: Application to Functional Neural Connectivity," arXiv preprint arXiv:2104.06389. [2291]

Williams, J., Bravo, H. C., Tom, J., and Paulson, J. N. (2019), "microbiomedasim: Simulating Longitudinal Differential Abundance for Microbiome Data," *F1000Research*, 8, 1769. [2282]

Yang, E., Baker, Y., Ravikumar, P., Allen, G., and Liu, Z. (2014), "Mixed Graphical Models via Exponential Families," in *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, pp. 1042–1050. [2282]

Yang, E., Ravikumar, P., Allen, G. I., and Liu, Z. (2015), "Graphical Models via Univariate Exponential Family Distributions," *The Journal of Machine Learning Research*, 16, 3813–3847. [2282]

Yu, M., Gupta, V., and Kolar, M. (2020), "Simultaneous Inference for Pairwise Graphical Models with Generalized Score Matching," *Journal of Machine Learning Research*, 21, 1–51. [2284]

Yuan, M., and Lin, Y. (2007), "Model Selection and Estimation in the Gaussian Graphical Model," *Biometrika*, 94, 19–35. [2282,2284]

Zhang, C.-H., and Zhang, S. S. (2014), "Confidence Intervals for Low Dimensional Parameters in High Dimensional Linear Models," *Journal of the Royal Statistical Society*, Series B, 76, 217–242. [2283,2287]

Zheng, L., and Allen, G. I. (2022), "Learning Gaussian Graphical Models with Differing Pairwise Sample Sizes, in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp. 5588–5592. [2283,2284]

Zheng, L., Rewolinski, Z. T., and Allen, G. I. (2022), "A Low-Rank Tensor Completion Approach for Imputing Functional Neuronal Data from Multiple Recordings," in *2022 IEEE Data Science and Learning Workshop (DSLW)*, IEEE, pp. 1–6. [2283]