StarTrek: Combinatorial Variable Selection with False Discovery Rate Control

Lu Zhang* and Junwei Lu†

Abstract

Variable selection on the large-scale networks has been extensively studied in the literature. While most of the existing methods are limited to the local functionals especially the graph edges, this paper focuses on selecting the discrete hub structures of the networks. Specifically, we propose an inferential method, called StarTrek filter, to select the hub nodes with degrees larger than a certain thresholding level in the high dimensional graphical models and control the false discovery rate (FDR). Discovering hub nodes in the networks is challenging: there is no straightforward statistic for testing the degree of a node due to the combinatorial structures; complicated dependence in the multiple testing problem is hard to characterize and control. In methodology, the StarTrek filter overcomes this by constructing p-values based on the maximum test statistics via the Gaussian multiplier bootstrap. In theory, we show that the StarTrek filter can control the FDR by providing accurate bounds on the approximation errors of the quantile estimation and addressing the dependence structures among the maximal statistics. To this end, we establish novel Cramér-type comparison bounds for the high dimensional Gaussian random vectors. Comparing to the Gaussian comparison bound via the Kolmogorov distance established by Chernozhukov et al. (2014), our Cramér-type comparison bounds establish the relative difference between the distribution functions of two high dimensional Gaussian random vectors, which is essential in the theoretical analysis of FDR control. Moreover, the StarTrek filter can be applied to general statistical models for FDR control of discovering discrete structures such as simultaneously testing the sparsity levels of multiple high dimensional linear models. We illustrate the validity of the StarTrek filter in a series of numerical experiments and apply it to the genotype-tissue expression dataset to discover central regulator genes.

Keywords. Graphical models, multiple testing, false discovery rate control, combinatorial inference, Gaussian multiplier bootstrap, comparison bounds.

1 Introduction

Graphical models are widely used for real-world problems in a broad range of fields, including social science, economics, genetics, and computational neuroscience (Newman et al., 2002; Luscombe et al., 2004; Rubinov and Sporns, 2010). Scientists and practitioners aim to understand the underlying network structure behind large-scale datasets. For a high-dimensional random vector $\boldsymbol{X} = (\boldsymbol{X}_1, \cdots, \boldsymbol{X}_d) \in \mathbb{R}^d$, we let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected graph, which encodes the conditional dependence structure among $\boldsymbol{X}$. Specifically, each component of $\boldsymbol{X}$ corresponds to some vertex

*Department of Statistics, Harvard University, Cambridge, MA 02138; Email: lu_zhang@g.harvard.edu.

†Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115; Email: junweilu@hsph.harvard.edu.

in $\mathcal{V} = \{1, 2 \cdots, d\}$, and $(j,k) \notin \mathcal{E}$ if and only if $\boldsymbol{X}_j$ and $\boldsymbol{X}_k$ are conditionally independent given the rest of variables. Many existing works in the literature seek to learn the structure of $\mathcal{G}$ via estimating the weight matrix $\boldsymbol{\Theta}$. For example, Meinshausen and Bühlmann (2006); Yuan and Lin (2007); Friedman et al. (2008); Rothman et al. (2008); Peng et al. (2009); Lam and Fan (2009); Ravikumar et al. (2011); Cai et al. (2011); Shen et al. (2012) focus on estimating the precision matrix in a Gaussian graphical model. Further, there is also a line of work developing methodology and theory to assess the uncertainty of edge estimation, i.e., constructing hypothesis tests and confidence intervals on the network edges, see Cai and Ma (2013); Gu et al. (2015); Ren et al. (2015); Cai and Zhang (2016); Janková and van de Geer (2017); Yang et al. (2018); Feng and Ning (2019); Ding and Zhou (2020). Recently, simultaneously testing multiple hypotheses on edges of the graphical models has received increasing attention (Liu, 2013; Cai et al., 2013; Xia et al., 2015, 2018; Li and Maathuis, 2019; Eisenach et al., 2020).

Most of the aforementioned works formulate the testing problems based on continuous parameters and local properties. For example, Liu (2013) proposes a method to select edges in Gaussian graphical models with asymptotic FDR control guarantees. Testing the existence of edges concerns the local structure of the graph. Under certain modeling assumptions, its null hypothesis can be translated into a single point in the continuous parameter space, for example, $\boldsymbol{\Theta}_{jk} = 0$ where $\boldsymbol{\Theta}$ is the precision matrix or the general weight matrix. However, for many scientific questions involving network structures, we need to detect and infer discrete and combinatorial signals in the networks, which does not follow from single edge testing. For example, in the study of social networks, it is interesting to discover active and impactful users, usually called "hub users," as they are connected to many other nodes in the social network (Ilyas et al., 2011; Lee et al., 2019). In gene co-expression network analysis, identifying central regulators/hub genes (Yuan et al., 2017; Liu et al., 2019b,a) is known to be extremely useful to the study of progression and prognosis of certain cancers and can support the treatment in the future. In neuroscience, researchers are interested in identifying the cerebral areas which are intensively connected to other regions (Shaw et al., 2008; van den Heuvel and Sporns, 2013; Power et al., 2013) during certain cognitive processes. The discovery of such central/hub areas can provide scientists better understanding of the mechanisms of human cognition.

Motivated by these applications in various areas, in this paper, we consider the hub node selection problem from the network models. In specific, given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the vertex set and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the edge set, we consider multiple hypotheses on whether the degree of some node $j \in \mathcal{V}$ exceeds a given threshold k_τ:

$$H_{0j} : \text{degree of node } j < k_\tau \text{ v.s. } H_{1j} : \text{degree of node } j \geq k_\tau.$$

Throughout the paper, these nodes with large degrees will be called hub nodes. For each $j \in [d]$, let $\psi_j = 1$ if H_{0j} is rejected and $\psi_j = 0$ otherwise. When selecting hub nodes, we would like to control the false discovery rate, as defined below:

$$\text{FDR} = \mathbb{E}\left[\frac{\sum_{j \in \mathcal{H}_0} \psi_j}{\max\left\{\sum_{j=1}^d \psi_j, 1\right\}}\right],$$

where $\mathcal{H}_0 = \{j \mid \text{degree of node } j < k_\tau\}$. Remark the hypotheses $H_{0j}, j \in [d]$ are not based on continuous parameters. They instead involve the degrees of the nodes, which are intrinsically discrete/combinatorial functionals. To the best of our knowledge, there is no existing literature studying such combinatorial variable selection problems. The most relevant work turns out to be Lu et al. (2017), which proposes a general framework for inference about graph invariants/combinatorial quantities on undirected graphical models. However, they study single hypothesis testing and have to decide which subgraph to be tested before running the procedure.

The combinatorial variable selection problems bring many new challenges. First, most of the existing work focus on testing continuous parameters (Liu, 2013; Javanmard and Montanari, 2013, 2014a,b; Belloni et al., 2014; Van de Geer et al., 2014; Xia et al., 2015, 2018; Javanmard and Javadi, 2019; Sur and Candès, 2019; Zhao et al., 2020). For discrete functionals, it is more difficult to construct appropriate test statistics and estimate its quantile accurately, especially in high dimensions. Second, many multiple testing procedures rely on an independence assumption (or certain dependence assumptions) on the null p-values (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001; Benjamini, 2010). However, the single hypothesis here is about the global property of the graph, which means that any reasonable test statistic has to involve the whole graph. Therefore, complicated dependence structures exist inevitably, which presents another layer of difficulty for controlling the false discoveries. Now we summarize the motivating question for this paper: how to develop a combinatorial selection procedure to discover nodes with large degrees on a graph with FDR control guarantees?

This paper introduces the StarTrek filter to select hub nodes. The filter is based on the maximum statistics, whose quantiles are approximated by the Gaussian multiplier bootstrap procedure. Briefly speaking, the Gaussian multiplier bootstrap procedure estimates the distribution of a given maximum statistic of general random vectors with unknown covariance matrices by the distribution of the maximum of a sum of the conditional Gaussian random vectors. The validity of high dimensional testing problems, such as family-wise error rate (FWER) control, relies on the non-asymptotic bounds of the Kolmogorov distance between the true distribution of the maximum statistics and the Gaussian multiplier bootstrap approximation, which is established in Chernozhukov et al. (2013). However, in order to control the FDR in the context of combinatorial variable selection, a more refined characterization of the quantile approximation errors is required. In specific, we need the so called Cramér-type comparison bounds quantifying the accuracy of the p-values in order to control the FDR in the simultaneous testing procedures (Chang et al., 2016). In our context, consider two centered Gaussian random vectors $U, V \in \mathbb{R}^d$ with different covariance matrices $\mathbf{\Sigma}^U$, $\mathbf{\Sigma}^V$ and denote the ℓ_∞ norms of U, V by $||U||_\infty, ||V||_\infty$ respectively, then the Cramér-type comparison bounds aim to control the relative error $\left|\frac{\mathbb{P}(||U||_\infty > t)}{\mathbb{P}(||V||_\infty > t)} - 1\right|$ for certain range of t. Comparing to the Kolmogorov distance $\sup_{t\in\mathbb{R}} |\mathbb{P}(||U||_\infty > t) - \mathbb{P}(||V||_\infty > t)|$ (Chernozhukov et al., 2015), the Cramér-type comparison bound leads to the relative error between two cumulative density functions, which is necessary to guarantee the FDR control. In specific, we show in this paper a novel Cramér-type Gaussian comparison bound

$$\sup_{0\le t\le C_0\sqrt{\log d}} \left|\frac{\mathbb{P}(||U||_\infty > t)}{\mathbb{P}(||V||_\infty > t)} - 1\right| = O\left(\min\left\{(\log d)^{5/2}\Delta_\infty^{1/2}, \frac{\Delta_0 \log d}{\mathfrak{p}}\right\}\right), \tag{1.1}$$

for some constant $C_0 > 0$, where $\Delta_\infty := ||\mathbf{\Sigma}^U - \mathbf{\Sigma}^V||_{\max}$ is the entrywise maximum norm difference between the two covariance matrices, $\Delta_0 := ||\mathbf{\Sigma}^U - \mathbf{\Sigma}^V||_0$ with $\|\cdot\|_0$ being the entrywise ℓ_0-norm of the matrix, and $\mathfrak{p}$ is the number of connected subgraphs in the graph whose edge set $\mathcal{E} = \{(j,k) : \mathbf{\Sigma}^U_{jk} \neq 0 \text{ or } \mathbf{\Sigma}^V_{jk} \neq 0\}$. This comparison bound in (1.1) characterizes the relative errors between Gaussian maxima via two types of rates: the ℓ_∞-norm Δ_∞ and the ℓ_0-norm Δ_0. This implies a new insight that the Cramér type bound between two Gaussian maxima is small as long as either their covariance matrices are uniformly close or only sparse entries of the two covariance matrices differ. As far as we know, the second type of rate in (1.1) has not been developed even in Kolmogorov distance results of high dimensional Gaussian maxima. In the study of FDR control, we need both types of rates: the Δ_∞ rate is used to show that the Gaussian multiplier bootstrap procedure is an accurate approximation for the maximum statistic quantiles and the Δ_0 rate is used to quantify the complicated dependence structure of the p-values for the single tests on the degree of graph nodes. In order to prove the Cramér-type comparison bound in (1.1), we develop two novel theoretic techniques to prove the two types of rates separately. For the Δ_∞ rate, we reformulate the Slepian's interpolation (Slepian, 1962) into an ordinary differential inequality such that the relative error can be controlled via the Grönwall's inequality (Grönwall, 1919). To control the Δ_0 rate, the anti-concentration inequality of Gaussian maxima developed in Chernozhukov et al. (2015) is no longer sufficient, we establish a new type of anti-concentration inequality for the derivatives of the soft-max of high dimensional Gaussian vectors. The existing works on the Cramér type comparison bounds such as Liu and Shao (2010, 2014); Chang et al. (2016) does not cover the high dimensional maximum statistics. Therefore, their techniques can not be directly extended to our case. To the best of our knowledge, it is the first time in our paper to prove the Cramér-type Gaussian comparison bounds (1.1) for high dimensional Gaussian maxima.

In summary, our paper makes the following major contributions. First, we develop a novel StarTrek filter to select combinatorial statistical signals: the hub nodes with the FDR control. This procedure involves maximum statistic and Gaussian multiplier bootstrap for quantile estimation. Second, in theory, the proposed method is shown to be valid for many different models with the network structures. In this paper, we provide two examples, the Gaussian graphical model and the bipartite network in the multiple linear models. Third, we prove a new Cramér-type Gaussian comparison bound with two types of rates: the maximum norm difference and ℓ_0 norm difference. These results are quite generic and has its own significance in the probability theory.

1.1 Related work

Canonical approaches to FDR control and multiple testing (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001; Benjamini, 2010) require that valid p-values are available, and they only allow for certain forms of dependence between these p-values. However, obtaining asymptotic p-values with sufficient accuracy is generally non-trivial for high dimensional hypothesis testing problems concerning continuous parameters (Javanmard and Montanari, 2013, 2014a,b; Belloni et al., 2014; Van de Geer et al., 2014; Sur and Candès, 2019; Zhao et al., 2020), not even to mention discrete/combinatorial functionals.

Recently, there is a line of work conducting variable selection without needing to act on a set of valid p-values, including Barber and Candès (2015, 2019); Candès et al. (2018); Xing et al. (2019); Dai et al. (2020a,b). These approaches take advantage of the symmetry of the null test statistics and establish FDR control guarantee. As their single hypothesis is often formulated as conditional independence testing, it is challenging to apply those techniques to select discrete signals for the problem studied in this paper.

Another line of work develops multiple testing procedures based on asymptotic p-values for specific high dimensional models (Liu, 2013; Liu and Luo, 2014; Javanmard and Javadi, 2019; Xia et al., 2015, 2018; Liu et al., 2020). Among them, Liu (2013) studies the edge selection problem on Gaussian graphical models, which turns out to be the most relevant work to our paper. However, their single hypothesis is about the local property of the graph. Our problem of discovering nodes with large degrees concerns the global property of the whole network, therefore requiring far more work.

There exists some recent work inferring combinatorial functionals. For example, the method proposed in Ke et al. (2020) provides a confidence interval for the number of spiked eigenvalues in a covariance matrix. Jin et al. (2020) focuses on estimating the number of communities in a network and yields confidence lower bounds. Neykov et al. (2019); Lu et al. (2017) propose a general framework for conducting inference on graph invariants/combinatorial quantities, such as the maximum degree, the negative number of connected subgraphs, and the size of the longest chain of a given graph. Shen and Lu (2020) develops methods for testing the general community combinatorial properties of the stochastic block model. Regarding the hypothesis testing problem, all these works only deal with a single hypothesis and establish asymptotic type-I error rate control. While simultaneously testing those combinatorial hypotheses is also very interesting and naturally arises from many practical problems.

1.2 Outline

In Section 2, we set up the general testing framework and introduce the StarTrek filter for selecting hub nodes. In Section 3, we present our core probabilistic tools: Cramér-type Gaussian comparison bounds in terms of maximum norm difference and ℓ_0 norm difference. To offer a relatively simpler illustration of our generic theoretical results, we first consider the hub selection problem on a bipartite network (multitask regression with linear models). Specifically, the input of the general StarTrek filter is chosen to be the estimators and quantile estimates described in Section 4. Applying the probabilistic results under this model, we establish FDR control guarantees under certain conditions. Then we move to the Gaussian graphical model in Section 5. In Section 6, we demonstrate StarTrek's performance through empirical simulations and a real data application.

1.3 Notations

Let $\phi(x), \Phi(x)$ be the probability density function (PDF) and the cumulative distribution function (CDF) respectively of the standard Gaussian distribution and denote $\bar{\Phi}(x) = 1 - \Phi(x)$. Let $\mathbf{1}_d$ be the vector of ones of dimension d. We use $\mathbb{1}(\cdot)$ to denote the indicator function of a set and $|\cdot|$ to denote the cardinality of a set. For two sets A and B, denote their symmetric difference by $A \ominus B$, i.e., $A \ominus B = (A \setminus B) \cup (B \setminus A)$; let $A \times B$ be the Cartesian product. For two positive sequences $\{x_n\}_{n=1}^{\infty}$ and $\{y_n\}_{n=1}^{\infty}$, we say $x_n = O(y_n)$ if $x_n \le C y_n$ holds for any n with some large enough $C > 0$. And we say $x_n = o(y_n)$ if $x_n / y_n \to 0$ as $n \to \infty$. For a sequence of random variables $\{X_n\}_{n=1}^{\infty}$ and a scalar a, we say $X_n \le a + o_{\mathbb{P}}(1)$ if for all $\epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}(X_n - a > \epsilon) = 0$. Let $[d]$ denote the set $\{1, \dots, d\}$. The ℓ_∞ norm and the ℓ_1 norm on $\mathbb{R}^d$ are denoted by $||\cdot||_\infty$ and $||\cdot||_1$ respectively. For a random vector X, let $||X||_\infty$ be its ℓ_∞ norm. For a matrix $\mathbf{A} \in \mathbb{R}^{d_1 \times d_2}$, we denote its minimal and maximal eigenvalues by $\lambda_{\min}(\mathbf{A}), \lambda_{\max}(\mathbf{A})$ respectively, the elementwise max norm by $\|\mathbf{A}\|_{\max} = \max_{i\in[d_1], j\in[d_2]} |\mathbf{A}_{ij}|$ and the elementwise ℓ_0 norm by $\|\mathbf{A}\|_0 = \sum_{i\in[d_1], j\in[d_2]} \mathbb{1}(\mathbf{A}_{ij} \neq 0)$. Throughout this paper, $C, C', C'', C_0, C_1, C_2, \dots$ are used as generic constants whose values may vary across different places.

2 Methodology

Before introducing our method, we set up the problem with more details. Specifically, we consider a graph $\mathcal{G} = (\mathcal{V}_1, \mathcal{V}_2, \mathcal{E})$ with the node sets $\mathcal{V}_1, \mathcal{V}_2$ and the edge set $\mathcal{E}$. Let $d_1 = |\mathcal{V}_1|$, $d_2 = |\mathcal{V}_2|$ and denote its weight matrix by $\mathbf{\Theta} \in \mathbb{R}^{d_1 \times d_2}$. In the undirected graph where $\mathcal{V}_1 = \mathcal{V}_2 := \mathcal{V}$, $\mathbf{\Theta}$ is a square matrix and its element $\mathbf{\Theta}_{jk}$ is nonzero when there is an edge between node j and node k, zero when there is no edge. In a bipartite graph where $\mathcal{V}_1 \neq \mathcal{V}_2$, elements of $\mathbf{\Theta}$ describe the existence of an edge between node j in $\mathcal{V}_1$ and node k in $\mathcal{V}_2$. Without loss of generality, we focus on one of the node sets and denote it by $\mathcal{V}$ with $|\mathcal{V}| := d$. We would like to select those nodes among $\mathcal{V}$ whose degree exceeds a certain threshold k_τ. And the selection problem is equivalent to simultaneously testing d hypotheses:

$$H_{0j} : \text{degree of node } j < k_\tau \text{ v.s. } H_{1j} : \text{degree of node } j \geq k_\tau, \tag{2.1}$$

for $j \in [d]$. Let $\psi_j = 1$ if H_{0j} is rejected and $\psi_j = 0$ otherwise, then for some multiple testing procedure with output $\{\psi_j\}_{j\in[d]}$, the false discovery proportion (FDP) and FDR can be defined as below:

$$\text{FDP} = \frac{\sum_{j\in\mathcal{H}_0}^{d} \psi_j}{\max\left\{1, \sum_{j=1}^{d} \psi_j\right\}}, \quad \text{FDR} := \mathbb{E}[\text{FDP}],$$

where $\mathcal{H}_0 = \{j \mid \text{degree of node } j < k_\tau\}$. We aim to propose a multiple testing procedure such that the FDP or FDR can be controlled at a given level $0 < q < 1$.

We illustrate the above general setup in two specific examples. In multitask regression with linear models, we are working with the bipartite graph case, then the weight matrix $\mathbf{\Theta}$ corresponds to the parameter matrix whose row represents the linear coefficients for one given response variable. Given a threshold k_τ, we want to select those rows (response variables) with ℓ_0 norm being at least k_τ. In the context of Gaussian graphical models where $\mathcal{V}_1 = \mathcal{V}_2$, $\mathbf{\Theta}$ represents the precision matrix, and we want to select those hub nodes i.e., whose degree is larger than or equal to k_τ.

2.1 StarTrek filter

Letting $\mathbf{\Theta}_j$ be the j-th row of $\mathbf{\Theta}$ and $\mathbf{\Theta}_{j,-j}$ be the vector $\mathbf{\Theta}_j$ excluding its j-th element, we formulate the testing problem for each single node as below,

$$H_{0j} : \|\mathbf{\Theta}_{j,-j}\|_0 < k_\tau \text{ v.s. } H_{1j} : \|\mathbf{\Theta}_{j,-j}\|_0 \geq k_\tau.$$

To test the above hypothesis, we need some estimator of the weight matrix $\mathbf{\Theta}$. In Gaussian graphical model, it is natural to use the estimator of a precision matrix. In the bipartite graph (multiple response model), estimated parameter matrix will suffice. Denote this generic estimator by $\widetilde{\mathbf{\Theta}}$ (without causing confusion in notation), the maximum test statistic over a given subset E of $\mathcal{V} \times \mathcal{V}$ will be

$$T_E := \max_{(j,k)\in E} \sqrt{n} \left|\widetilde{\mathbf{\Theta}}_{jk}\right|$$

and its quantile is defined as $c(\alpha, E) = \inf\{t \in \mathbb{R} \mid \mathbb{P}(T_E \leq t) \geq 1 - \alpha\}$, which is often unknown. Assume it can be estimated by $\widehat{c}(\alpha, E)$ from some procedure such as the Gaussian multiplier bootstrap, a generic method called skip-down procedure can be used, which was originally proposed in Lu et al. (2017) for testing a family of monotone graph invariants. When applied to the specific degree testing problem, it leads to the following algorithm.

Algorithm 1 Skip-down Method in Lu et al. (2017) (for testing the degree of node j)

Input: $\{\widetilde{\mathbf{\Theta}}_e\}_{e\in\mathcal{V}\times\mathcal{V}}$, significance level α.
Initialize $t = 0, E_0 = \{(j, k) : k \in [d], k \neq j\}$.
repeat
  $t \leftarrow t + 1$;
  Select the rejected edges $\mathcal{R} \leftarrow \{(j, k) \in E_{t-1} \mid \sqrt{n}|\widetilde{\mathbf{\Theta}}_{jk}| > \widehat{c}(\alpha, E_{t-1})\}$;
  $E_t \leftarrow E_{t-1}\backslash\mathcal{R}$;
until $|E_t^c| \geq k$ or $\mathcal{R} = \varnothing$
Output: $\psi_{j,\alpha} = 1$ if $|E_t^c| \geq k$ and $\psi_{j,\alpha} = 0$ otherwise.

To conduct the node selection over the whole graph, we need to determine an appropriate threshold $\widehat{\alpha}$ then reject H_{0j} if $\psi_{j,\widehat{\alpha}} = 1$. A desirable choice of $\widehat{\alpha}$ should be able to discover as many as hub nodes with the FDR remaining controlled under the nominal level q. For example, if the BHq procedure is considered, $\widehat{\alpha}$ can be defined as follows:

$$\widehat{\alpha} = \sup \left\{ \alpha \in (0,1) : \frac{\alpha d}{\max\left\{1, \sum_{j\in[d]} \psi_{j,\alpha}\right\}} \le q \right\}. \tag{2.2}$$

The above range of α is $(0,1)$, it will be very computationally expensive if we do an exhaustive search since for each α, we have to recompute the quantiles $\widehat{c}(\alpha, E)$ for a lot of sets E.

We overcome the computational difficulty and propose a efficient procedure called StarTrek filter, which is presented in Algorithm 2. Remark it only involves estimating k_τ different quantiles

Algorithm 2 StarTrek Filter

Input: $\{\widetilde{\mathbf{\Theta}}_e\}_{e\in\mathcal{V}\times\mathcal{V}}$, nominal FDR level q.
for $j \in [d]$ **do**
We order the elements in $\{|\widetilde{\mathbf{\Theta}}_{j\ell}| : \ell \ne j\}$ as $|\widetilde{\mathbf{\Theta}}_{j,(1)}| \ge |\widetilde{\mathbf{\Theta}}_{j,(2)}| \ge \ldots \ge |\widetilde{\mathbf{\Theta}}_{j,(d-1)}|$, where $|\widetilde{\mathbf{\Theta}}_{j,(\ell)}|$ is the ℓth largest entry. Compute $\alpha_j = \max_{1\le s\le k_\tau} \widehat{c}^{-1}(\sqrt{n}|\widetilde{\mathbf{\Theta}}_{j,(s)}|, E_j^{(s)})$ where $E_j^{(s)} := \{(j,\ell) : \ell \ne j, |\widetilde{\mathbf{\Theta}}_{j\ell}| \le |\widetilde{\mathbf{\Theta}}_{j,(s)}|\}$.
end for
Order α_j as $\alpha_{(1)} \le \alpha_{(2)} \le \cdots \le \alpha_{(d)}$ and set $\alpha_{(0)} = 0$, let $j_{\max} = \max\{0 \le j \le d : \alpha_{(j)} \le qj/d\}$.
Output: $S = \{j : \alpha_j \le \alpha_{(j_{\max})}\}$ if $j_{\max} > 0$; $S = \varnothing$ otherwise.

of some maximum statistics per node, which is more efficient than the Skip-down procedure (Lu et al., 2017) in terms of computation.

2.2 Accuracy of approximate quantiles

Before diving into the theoretical results, we pause to give specific forms of the estimator of $\mathbf{\Theta}$ and how to compute the estimated quantiles of the maximum statistic. Take the Gaussian graphical model as an example, suppose that $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \overset{\text{i.i.d.}}{\sim} N_d(0, \mathbf{\Sigma})$. Let $\mathbf{\Theta} = \mathbf{\Sigma}^{-1}$, which will have the same ℓ_0 elementwise norm as the adjacency matrix $\mathbf{\Theta}$. Denote $\mathbf{e}_k$ be the kth canonical basis in $\mathbb{R}^d$, we consider the following one-step estimator of $\mathbf{\Theta}_{jk}$,

$$\widehat{\mathbf{\Theta}}^{\mathrm{d}}_{jk} := \widehat{\mathbf{\Theta}}_{jk} - \frac{\widehat{\mathbf{\Theta}}_j^\top \left(\widehat{\mathbf{\Sigma}}\widehat{\mathbf{\Theta}}_k - \mathbf{e}_k\right)}{\widehat{\mathbf{\Theta}}_j^\top \widehat{\mathbf{\Sigma}}_j}, \tag{2.3}$$

where $\widehat{\mathbf{\Theta}}$ could be either the graphical Lasso (GLasso) estimator (Friedman et al., 2008) or the CLIME estimator (Cai et al., 2011). Let $\widetilde{\mathbf{\Theta}}^{\mathrm{d}}_{jk} := \widehat{\mathbf{\Theta}}^{\mathrm{d}}_{jk} / \sqrt{\widehat{\mathbf{\Theta}}^{\mathrm{d}}_{jj}\widehat{\mathbf{\Theta}}^{\mathrm{d}}_{kk}}$ and the standardized version $\{\widetilde{\mathbf{\Theta}}^{\mathrm{d}}_e\}_{\mathcal{V}\times\mathcal{V}}$ will be the input $\{\widetilde{\mathbf{\Theta}}_e\}_{\mathcal{V}\times\mathcal{V}}$ of Algorithm 2. Then the maximum test statistics (over the subset E) is defined as $T_E = \max\limits_{(j,k)\in E} \sqrt{n}|\widetilde{\mathbf{\Theta}}^{\mathrm{d}}_{jk}|$. To estimate its quantile, we construct the following Gaussian multiplier bootstrap

$$T_E^{\mathcal{B}} := \max_{(j,k)\in E} \frac{1}{\sqrt{n\,\widehat{\mathbf{\Theta}}_{jj}\widehat{\mathbf{\Theta}}_{kk}}} \left| \sum_{i=1}^n \widehat{\mathbf{\Theta}}_j^\top \left(\boldsymbol{X}_i\boldsymbol{X}_i^\top\widehat{\mathbf{\Theta}}_k - \mathbf{e}_k\right)\xi_i \right|, \tag{2.4}$$

where $\xi_i \stackrel{\text{i.i.d.}}{\sim} N(0,1)$, which produces $\widehat{c}(\alpha, E) = \inf\left\{t \in \mathbb{R} : \mathbb{P}_\xi\left(T_E^{\mathcal{B}} \le t\right) \ge 1-\alpha\right\}$ as the quantile estimate. We also denote the standardized true precision matrix $(\mathbf{\Theta}_{jk}/\sqrt{\mathbf{\Theta}_{jj}\mathbf{\Theta}_{kk}})_{j,k\in[d]}$ by $\mathbf{\Theta}^\star$. The theoretical results for Gaussian multiplier bootstrap developed in Chernozhukov et al. (2013) basically imply the above quantile estimates are accurate in the following sense:

Lemma 2.1. Suppose $\mathbf{\Theta} \in \mathcal{U}(M, s, r_0)$ and $(\log(dn))^7/n + s^2(\log dn)^4/n = o(1)$, for any edge set $E \subseteq \mathcal{V} \times \mathcal{V}$, we have

$$\lim_{(n,d)\to\infty} \sup_{\mathbf{\Theta}\in\mathcal{U}(M,s,r_0)} \sup_{\alpha\in(0,1)} \left| \mathbb{P}\left(\max_{e\in E} \sqrt{n}|\widetilde{\mathbf{\Theta}}_e^{\mathrm{d}} - \mathbf{\Theta}_e^\star| > \widehat{c}(\alpha, E)\right) - \alpha \right| = 0. \tag{2.5}$$

where $\widetilde{\mathbf{\Theta}}_e^{\mathrm{d}}$ is the standardized version of the one-step estimator (2.3).

Note that $\mathcal{U}(M, s, r_0)$ denotes the parameter space of precision matrices and is defined as below:

$$\mathcal{U}(M, s, r_0) = \left\{ \mathbf{\Theta} \in \mathbb{R}^{d\times d} \,\middle|\, \lambda_{\min}(\mathbf{\Theta}) \ge 1/r_0, \lambda_{\max}(\mathbf{\Theta}) \le r_0, \max_{j\in[d]} \|\mathbf{\Theta}_j\|_0 \le s, \|\mathbf{\Theta}\|_1 \le M \right\}.$$

The proof of Lemma 2.1 can be found in Appendix D.2. However, Lemma 2.1 is not sufficient for our multiple testing problem. Generally speaking, the probabilistic bounds in Chernozhukov et al. (2013) are in terms of Kolmogorov distance, which only provides a uniform characterization for the deviation behaviors. Their results can be used to establish FWER control for global testing problems based on the maximum test statistics. However, in order to establish FDR control, we have to show that the estimation of number of false discoveries is sufficiently accurate enough in the following sense, i.e., uniformly over certain range of α,

$$\frac{\alpha d_0}{\sum_{j\in\mathcal{H}_0} \psi_{j,\alpha}} \to 1, \qquad \text{in probability}$$

where $\mathcal{H}_0 = \{j : \|\mathbf{\Theta}_{j,-j}\|_0 < k_\tau\}$. The above result is different from the one needed for FWER control: $\mathbb{E}\left[\psi_{j,\alpha}\right] = \alpha + o(1), j \in \mathcal{H}_0$. In the context of our node selection problem, it can be reduced to the following,

$$\left| \frac{\sum_{j\in\mathcal{H}_0} \mathbb{1}(\max_{e\in E} \sqrt{n}|\widetilde{\mathbf{\Theta}}_e^{\mathrm{d}} - \mathbf{\Theta}_e^\star| \ge \widehat{c}(\alpha, E))}{d_0\alpha} - 1 \right| \to 0 \quad \text{in probability}$$

uniformly over certain range of α for some subset E. The above ratio is closedly related to the ratio in Cramér-type moderation deviation results (Liu and Shao, 2010, 2014; Liu, 2013). To this end, we establish the Cramér-type deviation bounds for the Gaussian multiplier bootstrap procedure. This type of results is built on two types of Cramér-type Gaussian comparison bounds, which are presented in Section 3.

3 Cramér-type comparison bounds for Gaussian maxima

In this section, we present the theoretic results on the Cramér-type comparison bounds for Gaussian maxima. Let $U, V \in \mathbb{R}^d$ be two centered Gaussian random vectors with different covariance matrices $\mathbf{\Sigma}^U = (\sigma_{jk}^U)_{1\le j,k\le d}$, $\mathbf{\Sigma}^V = (\sigma_{jk}^V)_{1\le j,k\le d}$. Recall that the maximal difference of the covariance matrices is $\Delta_\infty := ||\mathbf{\Sigma}^U - \mathbf{\Sigma}^V||_{\max}$ and the elementwise ℓ_0 norm difference of the covariance matrices is denoted by $\Delta_0 := \left\|\mathbf{\Sigma}^U - \mathbf{\Sigma}^V\right\|_0 = \sum_{j,k\in[d]} \mathbb{1}(\sigma_{jk}^U \ne \sigma_{jk}^U)$. The Gaussian maxima of U and V are denoted as $||U||_\infty$ and $||V||_\infty$. Now we present a Cramér-type comparison bound (CCB) between Gaussian maxima in terms of the maximum norm difference Δ_∞.

Theorem 3.1 (CCB with maximum norm difference)**.** Suppose $(\log d)^5 \Delta_\infty = O(1)$, then we have

$$\sup_{0 \le t \le C_0 \sqrt{\log d}} \left| \frac{\mathbb{P}(||U||_\infty > t)}{\mathbb{P}(||V||_\infty > t)} - 1 \right| = O\left((\log d)^{5/2} \Delta_\infty^{1/2} \right), \tag{3.1}$$

for some constant $C_0 > 0$.

Remark 3.1. *We can actually prove a more general form (see Theorem B.2 in the appendix) of the upper bound on the above term, without the assumption on Δ_∞. In fact, we bound the right hand side of (3.1) as $M_3 (\log d)^{3/2} A(\Delta_\infty) e^{M_3 (\log d)^{3/2} A(\Delta_\infty)}$, where $A(\Delta_\infty) = M_1 \log d \Delta_\infty^{1/2} \exp\left(M_2 \log^2 d \Delta_\infty^{1/2}\right)$ with the constants M_1, M_2 only depending on the variance terms $\min_{1 \le j \le d}\{\sigma_{jj}^U, \sigma_{jj}^V\}, \max_{1 \le j \le d}\{\sigma_{jj}^U, \sigma_{jj}^V\}$ and M_3 being a universal constant.*

When applying Theorem 3.1 to Gaussian multiplier bootstrap, $||\Delta||_\infty$ actually controls the maximum differences between the true covariance matrix and the empirical covariance matrix. Based on the bound of $||\Delta||_\infty$, we can show that the Cramér-type comparison bound in (3.1) will be $O((\log d)^{3/2} n^{-1/4})$ with high probability.

The proof can be found in Appendix B.1. The above result bounds the relative difference between the distribution functions of the two Gaussian maxima. Compared with the bound in terms of Kolmogorov distance, it has more refined characterization when t is large, which benefits from our iterative use of the Slepian interpolation. We denote the interpolation between U and V as $W(s) = \sqrt{s}U + \sqrt{1-s}V, s \in [0,1]$ and let $Q_t(s) = \mathbb{P}(||W(s)||_\infty > t)$. Existing results (Chernozhukov et al., 2013, 2014) quantify the difference between $Q_t(1)$ and $Q_t(0)$ uniformly over $t \in \mathbb{R}$, which leads to a bound on the Kolmogorov distance between Gaussian maxima. Our main innovation is to consider $R_t(s) = Q_t(s)/Q_t(0) - 1$ and show that for any given t, $\mathcal{R}_t : s \in [0,1] \mapsto |R_t(s)|$ is a contraction mapping with 0 being its fixed point. Specifically, we have the following upper bound on $|R_t(s)|$,

$$|R_t(s)| \le AB \int_0^s |R_t(\mu)| d\mu + AB \cdot s + A,$$

where AB and A can be controlled via the bound on the maximal difference of the covariance matrices Δ_∞ and converge to 0 under certain conditions. By Grönwall's inequality (Grönwall, 1919), we then derive the bound on $R_t(1)$ explicitly in terms of A and B, which finally lead to the desired Cramér-type comparison bound in (3.1).

The above theorem is a key ingredient for deriving Cramér-type deviation results for the Gaussian multiplier bootstrap procedure. However, in certain situations, comparison bounds in terms of maximum norm difference may not be appropriate. There exist cases where the covariance matrices of two Gaussian random vectors are not uniformly closed to each other, but have lots of identical entries. In particular, for the combinatorial variable selection problem in this paper, there exist complicated dependence structures between the maximum statistic for different nodes, since each time when the degree of one single node is tested, the statistic is computed based on the whole graph. Again, this highlights the challenge of the multiple testing problem in our paper. To establish FDR control, we need to deal with the dependence between the maximum statistic of pairs of non-hub nodes. By the definition of non-hub nodes, the covariance matrix difference between each pair of the involving Gaussian vectors actually has lots of zero entries. We would like to take advantage of this sparsity pattern when applying the comparison bound. However, the bound in (3.1) is not sharp when Δ_∞ is not negligible but Δ_0 is small. To this end, we develop a different version of the Cramér-type comparison bound as below.

Theorem 3.2 (CCB with elementwise ℓ_0-norm difference)**.** Assume the Gaussian random vectors U and V have unit variances, i.e., $\sigma_{jj}^U = \sigma_{jj}^V = 1, j \in [d]$ and there exists some $\sigma_0 < 1$ such that

$|\sigma_{jk}^V| \le \sigma_0, |\sigma_{jk}^U| \le \sigma_0$ for any $j \neq k$. Suppose there exists a disjoint $\mathfrak{p}$-partition of nodes $\cup_{\ell=1}^{\mathfrak{p}} \mathcal{C}_\ell = [d]$ such that $\sigma_{jk}^U = \sigma_{jk}^V = 0$ when $j \in \mathcal{C}_\ell$ and $k \in \mathcal{C}_{\ell'}$ for some $\ell \neq \ell'$. We have

$$\sup_{0 \le t \le C_0\sqrt{\log d}} \left| \frac{\mathbb{P}(||U||_\infty > t)}{\mathbb{P}(||V||_\infty > t)} - 1 \right| \le O\left(\frac{\Delta_0 \log d}{\mathfrak{p}}\right), \tag{3.2}$$

for some constant $C_0 > 0$.

When applying the above result to our multiple degree testing problem, specifically the covariance of maximum test statistics for pairs of non-hub nodes, Δ_0 can be controlled as k_τ^2 which is in a constant order. In Theorem 3.2, the quantity $\mathfrak{p}$ represents the number of connected subgraphs shared by the coviarance matrix networks of U and V. We refer to Theorem B.4 in the appendix for a generalized definition of $\mathfrak{p}$ to strengthen the results in (3.2). The $\mathfrak{p}$ in the denominator of the right hand side of Cramér-type comparison bound in (3.2) is necessary: it is possible that even if Δ_0 is small, when $\mathfrak{p}$ is large, the Camér-type Gaussian comparison bound is not converging to zero. For example, consider Gaussian vectors with unit variances $U = (X_1, X_2, Z, \dots, Z) \in \mathbb{R}^d$, $V = (Y_1, Y_2, Z, \dots, Z) \in \mathbb{R}^d$, where $\mathrm{corr}(X_1, X_2) = 0.9, \mathrm{corr}(Y_1, Y_2) = 0$ and $(X_1, X_2) \perp\!\!\!\perp Z$, $(Y_1, Y_2) \perp\!\!\!\perp Z$. For this case, the Camér-type Gaussian comparison bound

$$\sup_{0 \le t \le C_0\sqrt{\log d}} \left| \frac{\mathbb{P}(||U||_\infty > t)}{\mathbb{P}(||V||_\infty > t)} - 1 \right| = \sup_{0 \le t \le C_0\sqrt{\log d}} \left| \frac{\mathbb{P}(\max\{|X_1|, |X_2|, |Z|\} > t)}{\mathbb{P}(\max\{|Y_1|, |Y_2|, |Z|\} > t)} - 1 \right|$$

is not converging to zero as d goes to infinity even if the corresponding Δ_0 is 1 but $\mathfrak{p} = 2$.

Compared with Theorem 3.1, the above theorem provides a sharper comparison bound for large $\mathfrak{p}$ and small Δ_0. The two theorems together describe a interesting phase transition phenomenon, i.e., the dependence on $\mathbf{\Sigma}^U - \mathbf{\Sigma}^V$ of the Cramér-type comparison bound exhibits a difference behavior in the regime of large $\mathfrak{p}$ and small Δ_0 versus the regime of small Δ_∞.

The proof of Theorem 3.2 can be found in Appendix B.2. Our main technical innovation is to establish a new type of anti-concentration bound for "derivatives" of Gaussian maxima. Since both the indicator function and maximum function are discontinuous, we follows the idea of using smoothing approximation as in the proof of Theorem 3.1, specifically, we bound the following term

$$\mathbb{E}[|\partial_j \partial_k \varphi(W(s))| \cdot \mathbb{1}(t - \epsilon \le ||W(s)||_\infty \le t + \epsilon)], \tag{3.3}$$

where φ is the same approximation function of the indicator of ℓ_∞ norm with certain smoothing parameter β. Note that $\mathbb{E}[\mathbb{1}(t - \epsilon \le ||W(s)||_\infty \le t + \epsilon)]$ is the anti-concentration bound for Gaussian maxima (Chernozhukov et al., 2014). A non-uniform version is also established in Kuchibhotla et al. (2021). (3.3) can be viewed as the anti-concentration bound on the second order partial derivatives of the smooth approximation function. When deriving the comparison bound in terms of ℓ_0 norm difference, we have to deal with such terms as (3.3) when $\sigma_{jk}^U \neq \sigma_{jk}^V$. We show (3.3) can be controlled as

$$\mathbb{E}[|\partial_j \partial_k \varphi(W(s))| \cdot \mathbb{1}(t - \epsilon \le ||W(s)||_\infty \le t + \epsilon)] \lesssim \frac{\mathbb{P}\left(||V||_\infty > t\right)(\log d)^2}{\epsilon \beta \mathfrak{p}}.$$

The above anti-concentration bound is non-uniform and has only a logarithm dependence on the dimension d. It provides a relatively sharp characterization when t is large and the Gaussian graphical model is not highly connected (i.e., the number of connected components $\mathfrak{p}$ being large).

4 Discovering hub responses in multitask regression

The theoretical results presented in Section 3 will be the cornerstone for establishing FDR control of the multiple testing problem described in Section 2. As seen previously, the testing problem (2.1) is set up in a quite general way: $\mathbf{\Theta}$ is a weight matrix, and we would like to select rows whose ℓ_0 norm exceeds some threshold. This section considers the specific application to multitask/multiple response regression, which turns out to be less involved. We take advantage of it and demonstrate how to utilize the probabilistic tools in Section 3. After that, the theoretical results on FDR control for the Gaussian graphical models are presented and discussed in Section 5.

In multitask regression problem, multiple response variables are regressed on a common set of predictors. We can view this example as a bipartite graph $\mathcal{G} = (\mathcal{V}_1, \mathcal{V}_2, \mathcal{E}), |\mathcal{V}_1| = d_1, |\mathcal{V}_2| = d_2$, where $\mathcal{V}_1$ contains the response variables and $\mathcal{V}_2$ represents the common set of predictors. Each entry of the weight matrix $\mathbf{\Theta}$ indicates whether a given predictor is non-null or not for a given response variable. In the case of parametric model, $\mathbf{\Theta} \in \mathbb{R}^{d_1 \times d_2}$ corresponds to the parameter matrix. One might be interested in identifying shared sparsity patterns across different response variables. It can be solved by selecting a set of predictors being non-null for all response variables (Obozinski et al., 2006; Dai and Barber, 2016). This section problem is column-wise in the sense that we want to select columns of $\mathbf{\Theta}$, denoted by $\mathbf{\Theta}_{\cdot j}$, such that $||\mathbf{\Theta}_{\cdot j}||_0 = d_1$. It is also interesting to consider a row-wise selection problem formalized in (2.1). Under the multitask regression setup, we would like to select response variables with at least a certain amount of non-null predictors. We will call this type of response variables hub responses throughout the section. This has practical applications in real-world problems such as the gene-disease network.

Consider the multitask regression problem with linear models, we have n i.i.d. pairs of the response vector and the predictor vector, denoted by $(\boldsymbol{Y}_1, \boldsymbol{X}_1), (\boldsymbol{Y}_2, \boldsymbol{X}_2), \ldots, (\boldsymbol{Y}_n, \boldsymbol{X}_n)$, where $\boldsymbol{Y}_i \in \mathbb{R}^{d_1}, \boldsymbol{X}_i \in \mathbb{R}^{d_2}$ satisfy the following relationship,

$$\boldsymbol{Y}_i = \mathbf{\Theta}\boldsymbol{X}_i + \boldsymbol{E}_i, \text{ where } \boldsymbol{E}_i \sim \mathcal{N}(0, \mathbf{D}_{d_1 \times d_1}) \text{ and } \boldsymbol{X}_i \perp\!\!\!\perp \boldsymbol{E}_i, \tag{4.1}$$

where $\mathbf{\Theta} \in \mathbb{R}^{d_1 \times d_2}$ is the parameter matrix and $\mathbf{D}$ is a d_1 by d_1 diagonal matrix whose diagonal elements σ_j^2 is the noise variance for response variable $\boldsymbol{Y}^{(j)}$. Let $\boldsymbol{X}$ be the design matrix with rows $\boldsymbol{X}_1^\top, \ldots, \boldsymbol{X}_n^\top$, shared by different response variables, and assume the noise variables are independent conditional on the design matrix $\boldsymbol{X}$. Let $s = \max_{j \in [d_1]} ||\mathbf{\Theta}_j||_0$ be the sparsity level of the parameter matrix $\mathbf{\Theta}$, we want to select columns of the parameter matrix which has at least k_τ nonzero entries, i.e., select nodes with large degree among $[d_1]$ in the bipartite graph $\mathcal{G} = (\mathcal{V}_1, \mathcal{V}_2, \mathcal{E})$.

As mentioned in Section 2, some estimator of the parameter matrix is needed to conduct hypothesis testing. Debiased Lasso is widely used for parameter estimation and statistical inference in high dimensional linear models (Javanmard and Montanari, 2014a,b). For each response variable $\boldsymbol{Y}^{(j)}, j \in [d_1]$, we compute the debiased Lasso estimator, denoted by $\widetilde{\mathbf{\Theta}}_j^{\mathrm{d}}$ as

$$\widetilde{\mathbf{\Theta}}_j^{\mathrm{d}} = \widehat{\mathbf{\Theta}}_j + \frac{1}{n}\mathbf{M}\boldsymbol{X}^\top(\boldsymbol{Y}^{(j)} - \boldsymbol{X}\widehat{\mathbf{\Theta}}_j), \text{ where } \widehat{\mathbf{\Theta}}_j = \arg\min_{\beta \in \mathbb{R}^{d_2}} \left\{\frac{1}{2n}\|\boldsymbol{Y}^{(j)} - \boldsymbol{X}\beta\|_2^2 + \lambda\|\beta\|_1\right\}. \tag{4.2}$$

Note the above $\mathbf{M}$ is defined as $\mathbf{M} = (m_1, \ldots, m_{d_2})^\top$ where

$$m_i = \operatorname*{argmin}_m m^\top \widehat{\mathbf{\Sigma}} m, \quad \text{s.t. } \|\widehat{\mathbf{\Sigma}} m - e_i\|_\infty \le \mu\,, \tag{4.3}$$

and here $\widehat{\mathbf{\Sigma}} = (\boldsymbol{X}^\top\boldsymbol{X})/n$.

Then the debiased estimator of the parameter matrix, defined by $\widetilde{\mathbf{\Theta}}^{\mathrm{d}} := (\widetilde{\mathbf{\Theta}}_1^{\mathrm{d}}, \cdots, \widetilde{\mathbf{\Theta}}_{d_1}^{\mathrm{d}})^\top$, will be used the input $\{\widetilde{\mathbf{\Theta}}_e\}_{e \in \mathcal{V}_1 \times \mathcal{V}_2}$ of Algorithm 2. In addition, we also need to compute the quantile of the maximum statistics. There exist many work studying the asymptotic distribution of the debiased Lasso estimator. Among them, the results in Javanmard and Montanari (2014a) (when translated into our multitask regression setup) imply, for each response variable $\boldsymbol{Y}^{(j)}, j \in [d_1]$,

$$\sqrt{n}(\widetilde{\boldsymbol{\Theta}}_j^{\mathrm{d}} - \boldsymbol{\Theta}_j) = Z + \Xi, \quad Z|\boldsymbol{X} \sim \mathcal{N}(0, \sigma_j^2 M \widehat{\boldsymbol{\Sigma}} M^\top), \tag{4.4}$$

under proper assumptions. Additionally with a natural probabilistic model of the design matrix, the bias term can be showed to be $||\Xi||_\infty = O(\frac{s \log d_2}{\sqrt{n}})$ with high probability. As discussed in (Javanmard and Montanari, 2014a), the asymptotic normality result can be used for deriving confidence intervals and statistical hypothesis tests. As the noise variance σ_j is unknown, the scaled Lasso is used for its estimation (Javanmard and Montanari, 2014a; Sun and Zhang, 2012), given by the following joint optimization problem,

$$\{\widehat{\boldsymbol{\Theta}}_j, \widehat{\sigma}_j\} = \arg \min_{\beta \in \mathbb{R}^{d_2}, \sigma > 0} \Big\{ \frac{1}{2\sigma n} \|\boldsymbol{Y}^{(j)} - \boldsymbol{X}\beta\|_2^2 + \frac{\sigma}{2} + \lambda \|\beta\|_1 \Big\}. \tag{4.5}$$

Regarding our testing problem, intuitively we can use the quantile of the Gaussian maxima of $\mathcal{N}(0, \widehat{\sigma}_j^2 M \widehat{\boldsymbol{\Sigma}} M^\top)$ to approximate the quantile of maximum statistic $T_E = \max_{(j,k) \in E} \sqrt{n} |\widetilde{\boldsymbol{\Theta}}_{jk}^{\mathrm{d}}|$ for some given subset E. Specifically, let $Z_j \mid \boldsymbol{X}, \boldsymbol{Y}^{(j)} \sim \mathcal{N}(0, \widehat{\sigma}_j^2 M \widehat{\boldsymbol{\Sigma}} M^\top)$ where $Z_j \in \mathbb{R}^{d_2}$ and consider the subset $E \subset \{j\} \times \mathcal{V}_2$, we approximate the quantile of T_E by the following

$$T_E^{\mathcal{N}} := \max_{(j,k) \in E} |Z_{jk}|, \quad \widehat{c}(\alpha, E) = \inf \left\{ t \in \mathbb{R} : \mathbb{P}_Z \left(T_E^{\mathcal{N}} \le t \right) \ge 1 - \alpha \right\}. \tag{4.6}$$

Indeed, under proper scaling conditions, similar results as (2.5) can be established, i.e., as $n, d \to \infty$,

$$\sup_{\alpha \in (0,1)} \left| \mathbb{P} \left(\max_{(j,k) \in E} \sqrt{n} |\widetilde{\boldsymbol{\Theta}}_{jk}^{\mathrm{d}} - \boldsymbol{\Theta}_{jk}| > \widehat{c}(\alpha, E) \right) - \alpha \right| \to 0. \tag{4.7}$$

The above result is based on two ingredients: the asymptotic normality result and the control of the bias term Ξ. Below we list the required assumptions for those two ingredients, i.e., (4.4) and $||\Xi||_\infty = O(\frac{s \log d_2}{\sqrt{n}})$.

Assumption 4.1 (Debiased Lasso with random designs)**.** The following assumptions are from the ones of Theorems 7 and 8 in Javanmard and Montanari (2014a).

- Let $\boldsymbol{\Sigma} = \mathbb{E}\left[\boldsymbol{X}_1 \boldsymbol{X}_1^\top\right] \in \mathbb{R}^{d_2 \times d_2}$ be such that $\sigma_{\min}(\boldsymbol{\Sigma}) \ge C_{\min} > 0$, and $\sigma_{\max}(\boldsymbol{\Sigma}) \le C_{\max} < \infty$, and $\max_{j \in [d_2]} \boldsymbol{\Sigma}_{jj} \le 1$. Assume $\boldsymbol{X}\boldsymbol{\Sigma}^{-1/2}$ have independent subgaussian rows, with zero mean and subgaussian norm $\|\boldsymbol{\Sigma}^{-1/2} \boldsymbol{X}_i\|_{\psi_2} = \kappa$, for some constant $\kappa \in (0, \infty)$.
- $\mu = a\sqrt{(\log d_2)/n}$, and $n \ge \max(\nu_0 s \log(d_2/s), \nu_1 \log d_2)$, $\nu_1 = \max(1600\kappa^4, a/4)$, and $\lambda = \sigma\sqrt{(c^2 \log d_2)/n}$.

Remark that there may exist other ways of obtaining a consistent estimator of $\boldsymbol{\Theta}$ and sufficiently accurate quantile estimates under different assumptions. Since it is not the main focus of this paper, we will not elaborate on it. As mentioned before, the Kolmogorov type result in (4.7) can be immediately applied to the global testing problem to guarantee FWER control. However, it is not sufficient for FDR control of the multiple testing problem in this paper. And this is when the Cramér-type comparison bound for Gaussian maxima established in Section 3 play its role. In addition, signal strength condition is needed. Recall that $\mathcal{H}_0 = \{j \in [d_1] : ||\boldsymbol{\Theta}_j||_0 < k_\tau\}$ with $d_0 = |\mathcal{H}_0|$, we consider the following rows of $\boldsymbol{\Theta}$,

$$\mathcal{B} := \{j \in \mathcal{H}_0^c : \forall k \in \mathrm{supp}(\boldsymbol{\Theta}_j), |\boldsymbol{\Theta}_{jk}| > c\sqrt{\log d_2/n}\}, \tag{4.8}$$

and define the proportion of such rows as $\rho = |\mathcal{B}|/d_1$. In the context of multitask regression, ρ measures the proportion of hub response variables whose non-null parameter coefficients all exceed certain thresholds, thus characterizes the overall signal strength. Below we present our result on FDP/FDR control under appropriate assumptions.

Theorem 4.2 (FDP/FDR control)**.** Under Assumption 4.1 and the scaling condition $\frac{d_2 \log d_2 + d_0}{d_0 d_2 \rho} + \frac{s \log^2 d_2}{n^{1/2}} + \frac{\log^2 d_2}{(n\rho)^{1/5}} = o(1)$, if we implement the StarTrek procedure in Algorithm 2 with $\boldsymbol{\Theta}$ estimated by (4.2) and the quantiles approximated by (4.6), as $(n, d_1, d_2) \to \infty$, we have

$$\mathrm{FDP} \leq q\frac{d_0}{d_1} + o_{\mathbb{P}}(1) \quad \text{and} \quad \lim_{(n,d_1,d_2)\to\infty} \mathrm{FDR} \leq q\frac{d_0}{d_1}. \tag{4.9}$$

The proof of Theorem 4.2 can be found in Appendix A.3. Note that signal strength conditions which require some entries of parameter matrix $\boldsymbol{\Theta}$ have magnitudes exceeding $c\sqrt{\log d_2/n}$ are usually assumed in existing work studying FDR control problem for high dimensional models (Liu, 2013; Liu and Shao, 2014; Liu and Luo, 2014; Xia et al., 2015, 2018; Javanmard and Javadi, 2019).

5 Discovering hub nodes in Gaussian graphical models

This section focuses on the hub node selection problem on Gaussian graphical models. Recall in Section 2, we first compute the one-step estimator $\{\widehat{\boldsymbol{\Theta}}^{\mathrm{d}}_e\}_{e\in\mathcal{V}\times\mathcal{V}}$ in (2.3) then take its standardized version $\{\widetilde{\boldsymbol{\Theta}}^{\mathrm{d}}_e\}_{e\in\mathcal{V}\times\mathcal{V}}$ as the input of Algorithm 2 i.e.,

$$\widehat{\boldsymbol{\Theta}}^{\mathrm{d}}_{jk} := \widehat{\boldsymbol{\Theta}}_{jk} - \frac{\widehat{\boldsymbol{\Theta}}_j^\top \left(\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Theta}}_k - \mathbf{e}_k\right)}{\widehat{\boldsymbol{\Theta}}_j^\top \widehat{\boldsymbol{\Sigma}}_j}, \quad \widetilde{\boldsymbol{\Theta}}^{\mathrm{d}}_{jk} := \widehat{\boldsymbol{\Theta}}^{\mathrm{d}}_{jk} / \sqrt{\widehat{\boldsymbol{\Theta}}^{\mathrm{d}}_{jj}\widehat{\boldsymbol{\Theta}}^{\mathrm{d}}_{kk}}. \tag{5.1}$$

Our StarTrek filter selects nodes with large degrees based on the maximum statistics $T_E = \max_{(j,k)\in E} \sqrt{n}|\widetilde{\boldsymbol{\Theta}}^{\mathrm{d}}_{jk}|$ over certain subset E. We use the Gaussian multiplier bootstrap (2.4) to approximate the quantiles, specifically,

$$\widehat{c}(\alpha, E) = \inf\left\{t \in \mathbb{R} : \mathbb{P}_\xi\left(T_E^{\mathcal{B}} \leq t\right) \geq 1 - \alpha\right\}. \tag{5.2}$$

Chernozhukov et al. (2013) shows that this quantile approximation is accurate enough for FWER control in modern high dimensional simultaneous testing problems. Their results are based on the control of the non-asymptotic bounds in a Kolmogorov distance sense. Lu et al. (2017) also takes advantage of this result to test single hypothesis of graph properties or derive confidence bounds on graph invariants.

However, in order to conduct combinatorial variable selection with FDR control guarantees, we need more refined studies about the accuracy of the quantile approximation. This is due to the ratio nature of the definition of FDR, as explained in Section 2.2. Compared with the results in Chernozhukov et al. (2013), we provide a Cramér-type control on the approximation errors of the Gaussian multiplier bootstrap procedure. This is built on the probabilistic tools in Section 3, in particular, the Cramér-type Gaussian comparison bound with max norm difference in Theorem 3.1. Due to the dependence structure behind the hub selection problem in Graphical models, we also have to utilize Theorem 3.2. In a bit more detail, computing the maximum test statistic for testing node node actually involves the whole graph, resulting complicated dependence among the test statistics. The non-differentiability of the maximum function makes it very difficult to track this dependence. Also note that, this type of difficulty can not be easily circumvented by alternative methods, due to the discrete nature of the combinatorial inference problem. However, we figure out that the Cramér-type Gaussian comparison bound with ℓ_0 norm difference plays an important role in handling this challenge.

In general, the sparsity/density of the graph is closed related to the dependence level of multiple testing problem on graphical models. For example, Liu (2013); Xia et al. (2015, 2018) make certain assumptions on the sparsity level and control the dependence of test statistics when testing multiple hypotheses on graphical models/networks. For the hub node selection problem in this paper, a new quantity is introduced, and we will explain why it is suitable. Recall that we define the set of non-hub response variables in Section 4. Similarly, the set of non-hub nodes is denoted by $\mathcal{H}_0 = \{j \in [d] : ||\mathbf{\Theta}_j||_0 < k_\tau\}$ with $d_0 = |\mathcal{H}_0|$. Now we consider the following set,

$$S = \{(j_1, j_2, k_1, k_2) : j_1, j_2 \in \mathcal{H}_0, j_1 \neq j_2, k_1 \neq k_2, \mathbf{\Theta}_{j_1 j_2} = \mathbf{\Theta}_{j_1 k_1} = \mathbf{\Theta}_{j_2 k_2} = 0, \mathbf{\Theta}_{j_1 k_2} \neq 0, \mathbf{\Theta}_{j_2 k_1} \neq 0\}. \tag{5.3}$$

Remark that in the above definition, k_1 can be the same as j_2 and k_2 can be the same as j_1. If

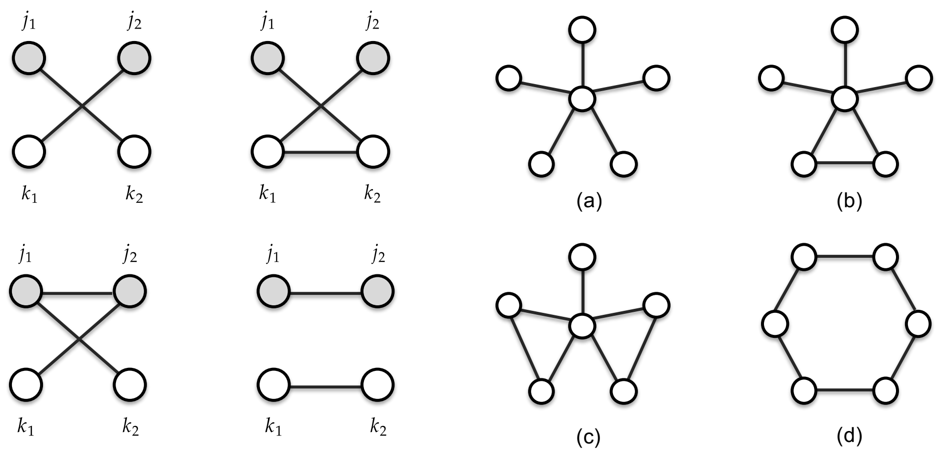


Figure 1: Left panel: a graphical demonstration of the definition of S via four examples of a 4-vertex graph; Right panel: four different graph patterns with 6 vertices. Calculating $|S|$ yields $10, 15, 24, 51$ for (a),(b),(c),(d) respectively.

there exists a large number of nodes which are neither connected to j_1 nor j_2, we then do not need to worry much about the dependence between the test statistics for non-hub nodes. Therefore, $|S|$ actually measures the dependence level via checking how a pair of non-hub nodes interact through other nodes. Liu (2013); Cai et al. (2013) also examine the connection structures in the 4-vertex graph and control the dependence level by carefully bounding the number of the 4-vertex graphs with different numbers of edges.

We provide a graphical demonstration of S and show how $|S|$ looks like in certain types of graph patterns via some simple examples. Though the definition of S does not exclude the possibility of (j_1, j_2, k_1, k_2) being a graph with 2 or 3 vertices, we only draw 4-vertex graph in Figure 1 for convenience. In the left panel of Figure 1, we consider four different cases of the 4-vertex graph. The upper two belong to the set S, while the lower three do not. In the right panel, we consider four graphs which all have 6 vertices. They have different graph patterns. For example, (a) clearly has a hub structure. All of the non-hub nodes are only connected to the hub node. While in (d), the edges are evenly distributed and each node are connected to its two nearest neighbours. For each graph, we count the value of $|S|$ and obtain $10, 15, 24, 51$ respectively, which show a increasing trend of $|S|$. This sort of matches our intuition that it is relatively easier to discover hub nodes on graph (a) compared with graph (d). See more evidence in the empirical results of Section 6.

In addition to $|S|$, we also characterize the dependence level via the connectivity of the graph, specifically let p be the number of connected components. And similarly as in Section 4, we define ρ to measure the signal strength, i.e., $\rho = |\mathcal{B}|/d$, where $\mathcal{B} := \{j \in \mathcal{H}_0^c : \forall k \in \text{supp}(\mathbf{\Theta}_j), |\mathbf{\Theta}_{jk}| > c\sqrt{\log d/n}\}$. In the following, we list our assumptions needed for FDR control.

Assumption 5.1. Suppose that $\mathbf{\Theta} \in \mathcal{U}(M, s, r_0)$ and the following conditions hold:

(i) Signal strength and scaling condition.

$$\frac{\log d}{\rho}\left(\frac{(\log d)^{19/6}}{n^{1/6}} + \frac{(\log d)^{11/6}}{\rho^{1/3}n^{1/6}} + \frac{s(\log d)^3}{n^{1/2}}\right) = o(1). \tag{5.4}$$

(ii) Dependency and connectivity condition.

$$\frac{\log d}{\rho d_0} + \frac{(\log d)^2|S|}{\rho d_0^2 p} = o(1). \tag{5.5}$$

In the above assumption, (5.4) places conditions on the signal strength and scaling. The first and the second term come from the Cramér-type large deviation bounds in the high dimensional CLT setting (Kuchibhotla et al., 2021) and the Cramér-type Gaussian comparison bound established in Theorem 3.1. And the third term comes from the fact that the relevant test statistics arise as maxima of approximate averages instead of the exact averages and thus the approximation error needs to be controlled. See similar discussions about this in (Chernozhukov et al., 2013). Remark that the signal strength condition is mild here, due to similar reasons as the discussion in Section 4. Regarding (5.5), there is a trade-off between the dependence level and connectivity level of the topological structure. $|S|/d_0^2$ characterizes how the test statistics of non-hub nodes are correlated to each other in average. p by definition describes the level of connectivity. Due to the condition (5.5), larger signal strength generally makes the hub selection problem easier. And when $|S|/d_0^2$ is small, the graph is allowed to be more connected. When there exist more sub-graphs, we allow higher correlations between the non-hub nodes. Note that the cardinality of S is directly related to the ℓ_0 norm covariance matrix difference term Δ_0, and arises from the application of Theorem 3.2. In the following, we present our core theoretical result on FDP/FDR control for hub selection using the StarTrek filter on Gaussian graphical models.

Theorem 5.2 (FDP/FDR control)**.** Under Assumption 5.1, the StarTrek procedure in Algorithm 2 with (5.1) as input and the quantiles approximated by (5.2) satisfies: as $(n, d) \to \infty$,

$$\text{FDP} \le q\frac{d_0}{d} + o_{\mathbb{P}}(1) \quad \text{and} \quad \lim_{(n,d)\to\infty} \text{FDR} \le q\frac{d_0}{d}. \tag{5.6}$$

The proof can be found in Appendix A.1. Remark that control of the FDR does not prohibit the FDP from varying. Therefore our result on FDP provides a stronger guarantee on controlling the false discoveries. See clear empirical evidence in Section 6.1. To the best of our knowledge, the proposed StarTrek filter in Section 2 and the above FDP/FDR control result are the first Algorithm and theoretical guarantee for the problem of simultaneously selecting hub nodes. Existing work like Liu (2013); Liu and Luo (2014); Xia et al. (2015, 2018); Javanmard and Javadi (2019) focus on the discovery of continuous signals and their tools are not applicable to the problem here.

6 Numerical results

6.1 Synthetic data

In this section, we apply the StarTrek filter to synthetic data and demonstrate the performance of our method. The synthetic datasets are generated from Gaussian graphical models. The corresponding precision matrices are specified based on four different types of graphs. Given the number of nodes d and the number of connected components p, we will randomly assign those nodes into p groups. Within each group (sub-graph), the way of assigning edges for different graph types will be explained below in detail. After determinning the adjacency matrix of the graph, we follow Zhao et al. (2012) to construct the precision matrix, more specifically, we set the off-diagonal elements to be of value v which control the magnitude of partial correlations and is closely related to the signal strength. In order to ensure positive-definiteness, we add some value v together with the absolute value of the minimal eigenvalues to the diagonal terms. In the following simulations, v and u are set to be 0.4 and 0.1 respectively. Now we explain how to determine the edges within each group (sub-graph) for four different graph patterns.

- **Hub graph.** We randomly pick one node as the hub node of the sub-graph, then the rest of the nodes are made to connect with this hub node. There is no edge between the non-hub nodes.
- **Random graph.** This is the Erdös-Rényi random graph. There is an edge between each pair of nodes with certain probability independently. In the following simulations, we will set this probability to be 0.15 unless stated otherwise.
- **Scale-free graph.** In this type of graphs, the degree distribution follows a power law. We construct it by the Barabási-Albert algorithm: starting with two connected nodes, then adding each new node to be connected with only one node in the existing graph; and the probability is proportional to the degree of the each node in the existing graph. The number of the edges will be the same as the number of nodes.
- **K-nearest-neighbor (knn) graph.** For a given number of k, we add edges such that each node is connected to another k nodes. In our simulations, k is sampled from $\{1, 2, 3, 4\}$ with probability mass $\{0.4, 0.3, 0.2, 0.1\}$.

See a visual demonstration of the above four different graph patterns in Appendix E.1. Throughout the simulated examples, we fix the number of nodes d to be 300 and vary other quantities such as sample size n or the number of connected components p. To estimate the precision matrix, we run the graphical Lasso algorithm with 5-fold cross-validation. Then we obtain the standardized debiased estimator as described in (2.3). To obtain the quantile estimates, we use the Gaussian multiplier bootstrap with 4000 bootstrap samples. The threshold k_τ for determining hub nodes is set to be 3. And all results (of FDR and power) are averaged over 64 independent replicates.

As we can see from Table 1, the FDRs of StarTrek filter for different types of graph are well controlled below the nominal levels. In hub graph, the FDRs are relatively small but the power is still pretty good. Similar phenomenon for multiple edge testing problem is observed (Liu, 2013). In the context of node testing, it is also unsurprising. These empirical results actually match our demonstration about $|S|$ in Figure 1: hub graphs have a relatively weaker dependence structure (smaller S values) and make it is easier to discover true hub nodes without making many errors.

The power performance of the StarTrek filter is showed in Table 2. As the sample size grows, we see the power is increasing for all four different types of graphs. When p is larger, there are more

Table 1: Empirical FDR

$d = 300$	$q = 0.1$			$q = 0.2$		
n	200	300	400	200	300	400
			$p = 20$			
hub	0.0000	0.0000	0.0007	0.0000	0.0000	0.0029
random	0.0255	0.0383	0.0467	0.0521	0.0770	0.0833
scale-free	0.0093	0.0211	0.0282	0.0352	0.0486	0.0581
knn	0.0101	0.0296	0.0370	0.0228	0.0620	0.0769
			$p = 30$			
hub	0.0013	0.0000	0.0016	0.0027	0.0054	0.0036
random	0.0347	0.0359	0.0568	0.0725	0.0753	0.0963
scale-free	0.0215	0.0335	0.0317	0.0521	0.0624	0.0584
knn	0.0297	0.0420	0.0563	0.0504	0.0857	0.1030

hub nodes in general due to the way of constructing the graphs, and we find the power is higher. Among different types of graphs, the power in hub graph and scale-free graph is higher than that in random and knn graph since the latter two are relatively denser and have more complicated topological structures.

Table 2: Power

$d = 300$	$q = 0.1$			$q = 0.2$		
n	200	300	400	200	300	400
			$p = 20$			
hub	0.7109	0.9453	0.9898	0.7805	0.9648	0.9938
random	0.3343	0.7815	0.9408	0.4520	0.8514	0.9604
scale-free	0.4524	0.8145	0.9363	0.5281	0.8614	0.9568
knn	0.0905	0.5306	0.8067	0.1634	0.6511	0.8630
			$p = 30$			
hub	0.6848	0.9244	0.9706	0.7588	0.9459	0.9784
random	0.4882	0.8863	0.9790	0.5770	0.9225	0.9870
scale-free	0.6472	0.9047	0.9810	0.7197	0.9331	0.9870
knn	0.2409	0.6841	0.8922	0.3298	0.7706	0.9241

In Figure 2 and 3, we demonstrate the performance of our method in the random graph with different parameters. Specifically, we vary the connecting probability changing from 0.1 to 0.3 in the x-axis. In those plots, we see the FDRs are all well controlled below the nominal level $q = 0.1$. As the connecting probability of the random graph grows, the graph gets denser, resulting more hub nodes. Thus we can see the height of the short blue solids lines (representing qd_0/d) is decreasing. Based on our results in Theorem 5.2, the target level of FDP/FDR control is qd_0/d. This is why we find the mean and median of each box-plot is getting smaller as the connecting probability increases (hence d_0 decreases).

The box-plots and the jittering points show that our StarTrek procedure not only controls the FDR but also prohibit it from varying too much, as implied by the theoretical results on FDP control in Section 5. Regarding the power plots, we see that the power is smaller when the graph is denser since the hub selection problem becomes more difficult with more disturbing factors. Plots with nominal FDR level $q = 0.2$ are deferred to Appendix E.3.

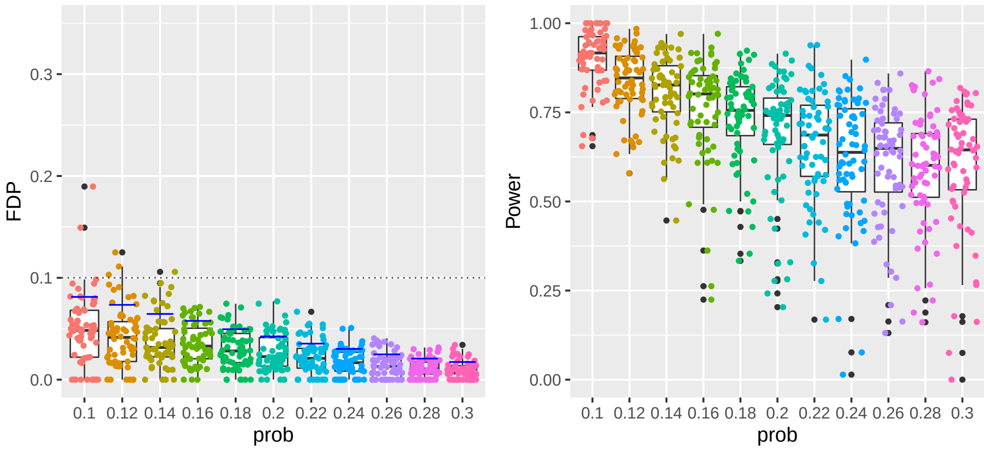


Figure 2: FDP and power plots for the StarTrek filter in the random graph. The connecting probability is varied on the x-axis. The number of samples n is chosen to be 300 and the number of connected components p equals 20. The nominal FDR level is set to be $q = 0.1$; the short blue solid lines correspond to qd_0/d, calculated by averaging over the 64 replicates. For both panels, the box plots are plotted with the black points representing the outliers. Colored points are jittered around, demonstrating how the FDP and power distribute.

6.2 Application to gene expression data

We also apply our method to the Genotype-Tissue Expression (GTEx) data studied in Lonsdale et al. (2013). Beginning with a 2.5-year pilot phase, the GTEx project establishes a great database and associated tissue bank for studying the relationship between certain genetic variations and gene expressions in human tissues. The original dataset involves 54 non-diseased tissue sites across 549 research subjects. Here we only focus on analyzing the breast mammary tissues. It is of great interest to identify hub genes over the gene expression network.

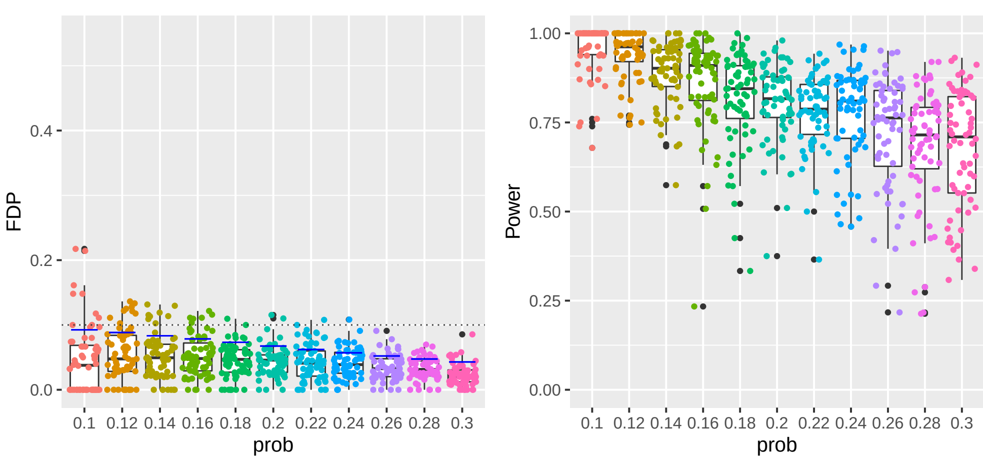


Figure 3: FDP and power plots for the StarTrek filter in the random graph. The other setups are the same as Figure 3 except for $p = 30$.

First we calculate the variances of the gene expression data and focus on the top 100 genes in the following analysis. The data involves $n = 291$ samples for male individuals and $n = 168$ samples

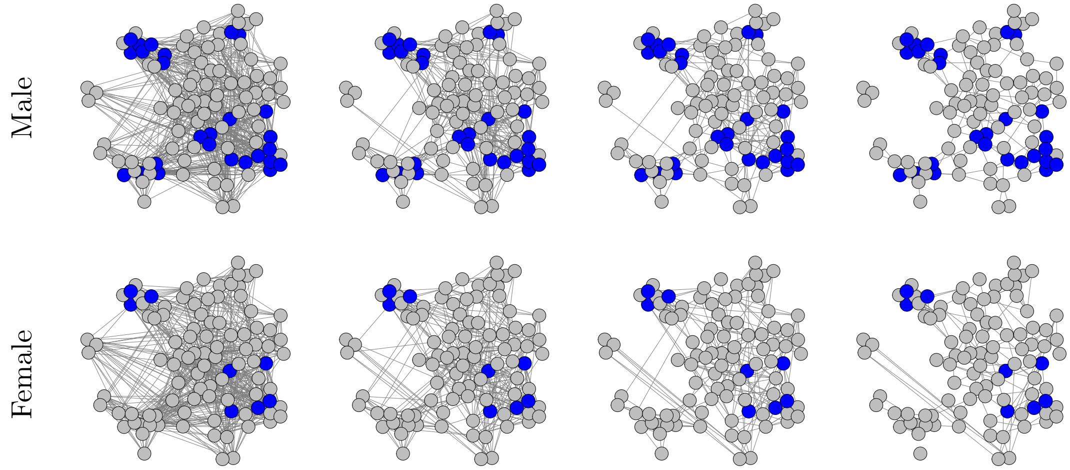


Figure 4: The above graphs are based the estimated precision matrices (the left two plots). The adjacency matrices of the other six plots are based on the standardized estimated precision matrices but thresholded at $0.025, 0.05, 0.075$ respectively. Blue vertices represent the selected hub genes.

for female individuals. The original count data is log-transformed and scaled. We then obtain the estimator of the precision matrix by the Graphical Lasso with 2-fold cross-validation. As for the hub node criterion, we set k_τ as the 50% quantile of the node degrees in the estimated precision matrix. We run StarTrek filter with 2000 bootstrap samples and nominal FDR level $q = 0.1$ to select hub genes for both the male and female datasets.

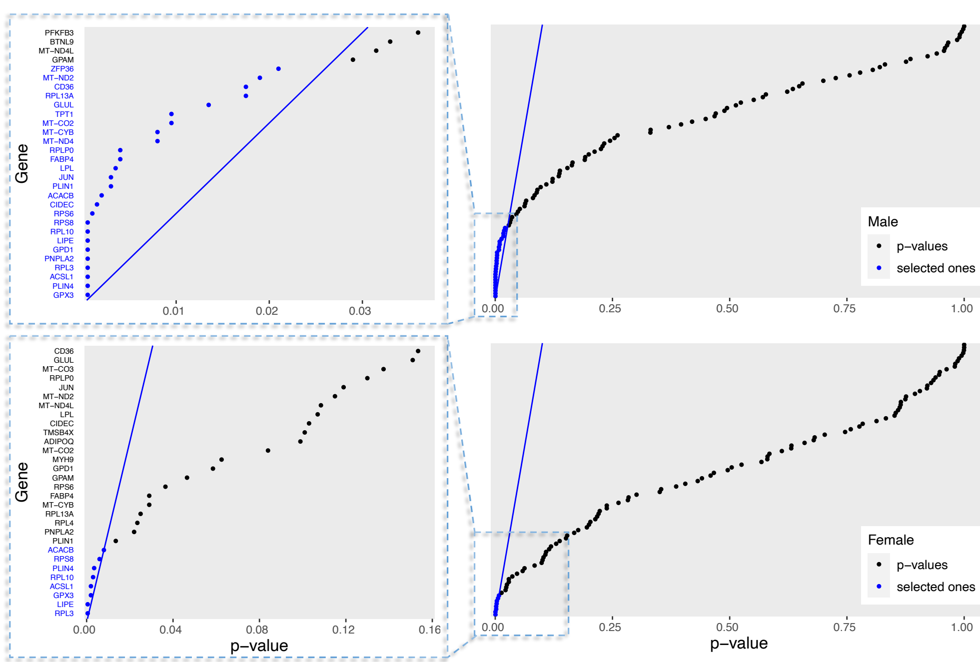


Figure 5: Plots of the sorted p-values ($\alpha_j, j \in [d]$) in Algorithm 2. Those blue points correspond to selected hub genes. The blue line is the rejection line of the BHq procedure. The coordinates of the plots are flipped. We abbreviate the names of the 100 genes and only show selected ones with blue colored text. The upper panel and the lower panel are based on male and female data respectively.

Figure 4 shows that the selected hub genes by the StarTrek filter also have large degrees on the estimated gene networks (based on the estimated precision matrices). In Figure 5, the results for male and female dataset agree with each other except that the number of selected hub genes using female dataset is smaller due to a much smaller sample size. The selected hub genes are found to play an important role in breast-related molecular processes, either as central regulators or their abnormal expressions are considered as the causes of breast cancer initiation and progression, see relevant literature in genetic research such as Hellwig et al. (2016); Blein et al. (2015); Chen et al. (2016); Li et al. (2019); Lou et al. (2020); Mohamed et al. (2014); Bai et al. (2019); Sirois et al. (2019); Marino et al. (2020); Malvia et al. (2019). Therefore, our proposed method for selecting hub nodes can be applied to the hub gene identification problem. It may improve our understanding of the mechanisms of breast cancer and provide valuable prognosis and treatment signature.

References

Bai, J., Zhang, X., Kang, X., Jin, L., Wang, P. and Wang, Z. (2019). Screening of core genes and pathways in breast cancer development via comprehensive analysis of multi gene expression datasets. *Oncology Letters* **18** 5821–5830.

Barber, R. F. and Candès, E. J. (2015). Controlling the false discovery rate via knockoffs. *The Annals of Statistics* **43** 2055 – 2085.

Barber, R. F. and Candès, E. J. (2019). A knockoff filter for high-dimensional selective inference. *The Annals of Statistics* **47** 2504 – 2537.

Belloni, A., Chernozhukov, V. and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies* **81** 608–650.

Benjamini, Y. (2010). Discovering the false discovery rate. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72** 405–416.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **57** 289–300.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics* 1165–1188.

Bentkus, V. (1990). Smooth approximations of the norm and differentiable functions with bounded support in banach space l^k_∞. *Lithuanian Mathematical Journal* **30** 223–230.

Blein, S., Barjhoux, L., investigators, G., Damiola, F., Dondon, M.-G., Eon-Marchais, S., Marcou, M., Caron, O., Lortholary, A., Buecher, B., Vennin, P., Berthet, P., Noguès, C., Lasset, C., Gauthier-Villars, M., Mazoyer, S., Stoppa-Lyonnet, D., Andrieu, N., Thomas, G., Sinilnikova, O. M. and Cox, D. G. (2015). Targeted sequencing of the mitochondrial genome of women at high risk of breast cancer without detectable mutations in brca1/2. *PLoS One* **10** e0136192.

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.

Cai, T., Liu, W. and Xia, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association* **108** 265–277.

Cai, T. T. and Ma, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19** 2359 – 2388.

Cai, T. T. and Zhang, A. (2016). Inference for high-dimensional differential correlation matrices. *Journal of Multivariate Analysis* **143** 107–126.

Candès, E. J., Fan, Y., Janson, L. and Lv, J. (2018). Panning for gold: Model-X knockoffs for high-dimensional controlled variable selection. *Journal of the Royal Statistical Society: Series B* **80** 551–577.

Chang, J., Shao, Q.-M. and Zhou, W.-X. (2016). Cramér-type moderate deviations for Studentized two-sample U-statistics with applications. *The Annals of Statistics* **44** 1931 – 1956.

Chen, W.-C., Wang, C.-Y., Hung, Y.-H., Weng, T.-Y., Yen, M.-C. and Lai, M.-D. (2016). Systematic analysis of gene expression alterations and clinical outcomes for long-chain acyl-coenzyme a synthetase family in cancer. *PLoS One* **11** e0155660.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics* **41** 2786–2819.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2014). Anti-concentration and honest, adaptive confidence bands. *The Annals of Statistics* **42** 1787–1818.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2015). Comparison and anti-concentration bounds for maxima of gaussian random vectors. *Probability Theory and Related Fields* **162** 47–70.

Dai, C., Lin, B., Xing, X. and Liu, J. S. (2020a). False discovery rate control via data splitting. *arXiv preprint arXiv:2002.08542* .

Dai, C., Lin, B., Xing, X. and Liu, J. S. (2020b). A scale-free approach for false discovery rate control in generalized linear models. *arXiv preprint arXiv:2007.01237* .

Dai, R. and Barber, R. (2016). The knockoff filter for FDR control in group-sparse and multitask regression. In *International Conference on Machine Learning*. PMLR.

Ding, X. and Zhou, Z. (2020). Estimation and inference for precision matrices of nonstationary time series. *The Annals of Statistics* **48** 2455 – 2477.

Eisenach, C., Bunea, F., Ning, Y. and Dinicu, C. (2020). High-dimensional inference for cluster-based graphical models. *Journal of Machine Learning Research* **21**.

Feng, H. and Ning, Y. (2019). High-dimensional mixed graphical model with ordinal data: Parameter estimation and statistical inference. In *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR.

Friedman, J., Hastie, T. and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441.

Grönwall, T. H. (1919). Note on the derivatives with respect to a parameter of the solutions of a system of differential equations. *Annals of Mathematics* **20** 292–296.

Gu, Q., Cao, Y., Ning, Y. and Liu, H. (2015). Local and global inference for high dimensional nonparanormal graphical models. *arXiv preprint arXiv:1502.02347* .

Hellwig, B., Madjar, K., Edlund, K., Marchan, R., Cadenas, C., Heimes, A.-S., Almstedt, K., Lebrecht, A., Sicking, I., Battista, M. J., Micke, P., Schmidt, M., Hengstler, J. G. and Rahnenführer, J. (2016). Epsin family member 3 and ribosome-related genes are associated with late metastasis in estrogen receptor-positive breast cancer and long-term survival in non-small cell lung cancer using a genome-wide identification and validation strategy. *PLoS One* **11** 1–18.

Ilyas, M. U., Shafiq, M. Z., Liu, A. X. and Radha, H. (2011). A distributed and privacy preserving algorithm for identifying information hubs in social networks. In *2011 Proceedings IEEE INFOCOM*. IEEE.

Isserlis, L. (1918). On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika* **12** 134–139.

Janková, J. and van de Geer, S. (2017). Honest confidence regions and optimality in high-dimensional precision matrix estimation. *Test* **26** 143–162.

Janková, J. and van de Geer, S. (2018). Inference in high-dimensional graphical models. *arXiv preprint arXiv:1801.08512* .

Javanmard, A. and Javadi, H. (2019). False discovery rate control via debiased lasso. *Electronic Journal of Statistics* **13** 1212 – 1253.

Javanmard, A. and Montanari, A. (2013). Nearly optimal sample size in hypothesis testing for high-dimensional regression. In *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE.

Javanmard, A. and Montanari, A. (2014a). Confidence intervals and hypothesis testing for high-dimensional regression. *Journal of Machine Learning Research* **15** 2869–2909.

Javanmard, A. and Montanari, A. (2014b). Hypothesis testing in high-dimensional regression under the gaussian random design model: Asymptotic theory. *IEEE Transactions on Information Theory* **60** 6522–6554.

Jin, J., Ke, Z. T., Luo, S. and Wang, M. (2020). Estimating the number of communities by stepwise goodness-of-fit. *arXiv preprint arXiv:2009.09177* .

Ke, Z. T., Ma, Y. and Lin, X. (2020). Estimation of the number of spiked eigenvalues in a covariance matrix by bulk eigenvalue matching analysis. *arXiv preprint arXiv:2006.00436* .

Kuchibhotla, A. K., Mukherjee, S. and Banerjee, D. (2021). High-dimensional CLT: Improvements, non-uniform extensions and large deviations. *Bernoulli* **27** 192 – 217.

Lam, C. and Fan, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *The Annals of Statistics* **37** 4254–4278.

Lee, R. K.-W., Hoang, T.-A. and Lim, E.-P. (2019). Discovering hidden topical hubs and authorities across multiple online social networks. *IEEE Transactions on Knowledge and Data Engineering* **33** 70–84.

Li, J. and Maathuis, M. H. (2019). GGM knockoff filter: False discovery rate control for gaussian graphical models. *arXiv preprint arXiv:1908.11611* .

Li, Y., Giorgi, E. E., Beckman, K. B., Caberto, C., Kazma, R., Lum-Jones, A., Haiman, C. A., Marchand, L. L., Stram, D. O., Saxena, R. and Cheng, I. (2019). Association between mitochondrial genetic variation and breast cancer risk: The multiethnic cohort. *PLoS One* **14** 1–14.

Liu, M., Xia, Y., Cai, T. and Cho, K. (2020). Integrative high dimensional multiple testing with heterogeneity under data sharing constraints. *arXiv preprint arXiv:2004.00816* .

Liu, W. (2013). Gaussian graphical model estimation with false discovery rate control. *The Annals of Statistics* **41** 2948–2978.

Liu, W. and Luo, S. (2014). Hypothesis testing for high-dimensional regression models.

Liu, W. and Shao, Q.-M. (2010). Cramér-type moderate deviation for the maximum of the periodogram with application to simultaneous tests in gene expression time series. *The Annals of Statistics* **38** 1913 – 1935.

Liu, W. and Shao, Q.-M. (2014). Phase transition and regularized bootstrap in large-scale t-tests with false discovery rate control. *The Annals of Statistics* **42** 2003 – 2025.

Liu, Y., Gu, H.-Y., Zhu, J., Niu, Y.-M., Zhang, C. and Guo, G.-L. (2019a). Identification of hub genes and key pathways associated with bipolar disorder based on weighted gene co-expression network analysis. *Frontiers in Physiology* **10** 1081.

Liu, Y., Yi, Y., Wu, W., Wu, K. and Zhang, W. (2019b). Bioinformatics prediction and analysis of hub genes and pathways of three types of gynecological cancer. *Oncology Letters* **18** 617–628.

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N. et al. (2013). The genotype-tissue expression (GTEx) project. *Nature Genetics* **45** 580–585.

Lou, W., Ding, B., Wang, S. and Fu, P. (2020). Overexpression of gpx3, a potential biomarker for diagnosis and prognosis of breast cancer, inhibits progression of breast cancer cells in vitro. *Cancer Cell International* **20** 1–15.

Lu, J., Neykov, M. and Liu, H. (2017). Adaptive inferential method for monotone graph invariants. *arXiv preprint arXiv:1707.09114* .

Luscombe, N. M., Babu, M. M., Yu, H., Snyder, M., Teichmann, S. A. and Gerstein, M. (2004). Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* **431** 308–312.

Malvia, S., Bagadi, S. A. R., Pradhan, D., Chintamani, C., Bhatnagar, A., Arora, D., Sarin, R. and Saxena, S. (2019). Study of gene expression profiles of breast cancers in indian women. *Scientific Reports* **9** 1–15.

Marino, N., German, R., Rao, X., Simpson, E., Liu, S., Wan, J., Liu, Y., Sandusky, G., Jacobsen, M., Stoval, M., Cao, S. and Storniolo, A. M. V. (2020). Upregulation of lipid metabolism genes in the breast prior to cancer diagnosis. *NPJ Breast Cancer* **6** 1–13.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics* **34** 1436–1462.

Mohamed, M. M., Sabet, S., Peng, D.-F., Nouh, M. A., El-Shinawi, M. and El-Rifai, W. (2014). Promoter hypermethylation and suppression of glutathione peroxidase 3 are associated with inflammatory breast carcinogenesis. *Oxidative Medicine and Cellular Longevity* **2014**.

Newman, M. E., Watts, D. J. and Strogatz, S. H. (2002). Random graph models of social networks. *Proceedings of the National Academy of Sciences* **99** 2566–2572.

Neykov, M., Lu, J. and Liu, H. (2019). Combinatorial inference for graphical models. *The Annals of Statistics* **47** 795–827.

Obozinski, G., Taskar, B. and Jordan, M. (2006). Multi-task feature selection. *Statistics Department, UC Berkeley, Tech. Rep* **2** 2.

Peng, J., Wang, P., Zhou, N. and Zhu, J. (2009). Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association* **104** 735–746.

Power, J. D., Schlaggar, B. L., Lessov-Schlaggar, C. N. and Petersen, S. E. (2013). Evidence for hubs in human functional brain networks. *Neuron* **79** 798–813.

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.

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.

Rothman, A. J., Bickel, P. J., Levina, E. and Zhu, J. (2008). Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics* **2** 494–515.

Rubinov, M. and Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* **52** 1059–1069.

Shaw, P., Kabani, N. J., Lerch, J. P., Eckstrand, K., Lenroot, R., Gogtay, N., Greenstein, D., Clasen, L., Evans, A., Rapoport, J. L., Giedd, J. N. and Wise, S. P. (2008). Neurodevelopmental trajectories of the human cerebral cortex. *Journal of Neuroscience* **28** 3586–3594.

Shen, S. and Lu, J. (2020). Combinatorial-probabilistic trade-off: Community properties test in the stochastic block models. *arXiv preprint arXiv:2010.15063* .

Shen, X., Pan, W. and Zhu, Y. (2012). Likelihood-based selection and sharp parameter estimation. *Journal of the American Statistical Association* **107** 223–232.

Sirois, I., Aguilar-Mahecha, A., Lafleur, J., Fowler, E., Vu, V., Scriver, M., Buchanan, M., Chabot, C., Ramanathan, A., Balachandran, B., Légaré, S., Przybytkowski, E., Lan, C., Krzemien, U., Cavallone, L., Aleynikova, O., Ferrario, C., Guilbert, M.-C., Benlimame, N., Saad, A., Alaoui-Jamali, M., Saragovi, H. U., Josephy, S., O'Flanagan, C., Hursting, S. D., Richard, V. R., Zahedi, R. P., Borchers, C. H., Bareke, E., Nabavi, S., Tonellato, P., Roy, J.-A., Robidoux, A., Marcus, E. A., Mihalcioiu, C., Majewski, J. and Basik, M. (2019). A unique morphological phenotype in chemoresistant triple-negative breast cancer reveals metabolic reprogramming and PLIN4 expression as a molecular vulnerability. *Molecular Cancer Research* **17** 2492–2507.

Slepian, D. (1962). The one-sided barrier problem for gaussian noise. *Bell System Technical Journal* **41** 463–501.

Sun, T. and Zhang, C.-H. (2012). Scaled sparse linear regression. *Biometrika* **99** 879–898.

Sur, P. and Candès, E. J. (2019). A modern maximum-likelihood theory for high-dimensional logistic regression. *Proceedings of the National Academy of Sciences* **116** 14516–14525.

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.

van den Heuvel, M. P. and Sporns, O. (2013). Network hubs in the human brain. *Trends in cognitive sciences* **17** 683–696.

Van Der Vaart, A. W. and Wellner, J. A. (1996). Weak convergence. In *Weak convergence and empirical processes*. Springer, 16–28.

van Handel, R. (2014). Probability in high dimension. Tech. rep., Princeton University.

Xia, Y., Cai, T. and Cai, T. T. (2015). Testing differential networks with applications to the detection of gene-gene interactions. *Biometrika* **102** 247–266.

Xia, Y., Cai, T. and Cai, T. T. (2018). Multiple testing of submatrices of a precision matrix with applications to identification of between pathway interactions. *Journal of the American Statistical Association* **113** 328–339.

Xing, X., Zhao, Z. and Liu, J. S. (2019). Controlling false discovery rate using gaussian mirrors. *arXiv preprint arXiv:1911.09761* .

Yang, Z., Ning, Y. and Liu, H. (2018). On semiparametric exponential family graphical models. *Journal of Machine Learning Research* **19** 2314–2372.

Yuan, L., Chen, L., Qian, K., Qian, G., Wu, C.-L., Wang, X. and Xiao, Y. (2017). Co-expression network analysis identified six hub genes in association with progression and prognosis in human clear cell renal cell carcinoma (ccrcc). *Genomics Data* **14** 132–140.

Yuan, M. and Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* **94** 19–35.

Zhao, Q., Sur, P. and Candes, E. J. (2020). The asymptotic distribution of the mle in high-dimensional logistic models: Arbitrary covariance. *arXiv preprint arXiv:2001.09351* .

Zhao, T., Liu, H., Roeder, K., Lafferty, J. and Wasserman, L. (2012). The huge package for high-dimensional undirected graph estimation in R. *Journal of Machine Learning Research* **13** 1059–1062.

Supplementary material to

StarTrek: Combinatorial Variable Selection with False Discovery Rate Control

This document contains the supplementary material to the paper "StarTrek: Combinatorial Variable Selection with False Discovery Rate Control". Appendix A presents the proofs of the FDR control results. In Appendix B, we provide the proofs of two types of Cramér-type comparison bounds for Gaussian maxima. Appendix C proves the Cramér-type deviation bounds for the Gaussian multiplier bootstrap. In Appendix D, we establish the validity and a power result of our test on the degree of a single node. Appendix E contains some plots and tables deferred from the main paper.

A Proofs for FDR control

In this section, we aim to prove Theorem 5.2. In order to prove the theorem, we need Lemma A.1 which is about the test of single node degree. Remark that this lemma proves the asymptotic validity of the test in Algorithm 1 and provides a power analysis. The signal strength condition is only required for the power analysis part. To see why Lemma A.1 is useful for establishing FDR control for our StarTrek procedure in Algorithm 2, we notice the following equivalence:

$$\{\psi_{j,\alpha} = 1\} = \{\alpha_j \leq \alpha\}, \tag{A.1}$$

where α is a given type-I error level, $\psi_{j,\alpha}$ is the test described in Algorithm 1, and α_j is defined in Algorithm 2. First, we show $\{\alpha_j \leq \alpha\} \subset \{\psi_{j,\alpha} = 1\}$. Note

$$\begin{aligned}\{\alpha_j \leq \alpha\} &= \bigcap_{1\leq s\leq k_\tau} \{\widehat{c}^{-1}(\sqrt{n}|\widetilde{\boldsymbol{\Theta}}_{j,(s)}|, E_j^{(s)}) \leq \alpha\} \\ &= \bigcap_{1\leq s\leq k_\tau} \{\sqrt{n}|\widetilde{\boldsymbol{\Theta}}_{j,(s)}| \geq \widehat{c}(\alpha, E_j^{(s)})\},\end{aligned} \tag{A.2}$$

where $E_j^{(s)} := \{(j,\ell) : \ell \neq j, |\widetilde{\boldsymbol{\Theta}}_{j\ell}| \leq |\widetilde{\boldsymbol{\Theta}}_{j,(s)}|\}$. The first equality is due to the definition of α_j and the second equality holds by the definition of $\widehat{c}^{-1}$. Examining (A.2), we immediately know $\sqrt{n}|\widetilde{\boldsymbol{\Theta}}_{j,(1)}| \geq \widehat{c}(\alpha, E_j^{(1)})$ (here $E_j^{(1)} = E_0 = \{(k,j) : k \in [d], k \neq j\}$), thus the edge corresponding to $\widetilde{\boldsymbol{\Theta}}_{j,(1)}$ will be rejected in the first iteration of Algorithm 1. Regarding the edge corresponding to $\widetilde{\boldsymbol{\Theta}}_{j,(2)}$, if $\sqrt{n}|\widetilde{\boldsymbol{\Theta}}_{j,(2)}| \geq \widehat{c}(\alpha, E_j^{(1)})$, then it will be rejected in the first iteration, too. Otherwise, Algorithm 1 enters the second iteration. Since (A.2) implies $\sqrt{n}|\widetilde{\boldsymbol{\Theta}}_{j,(2)}| \geq \widehat{c}(\alpha, E_j^{(2)})$, we know the edge corresponding to $\widetilde{\boldsymbol{\Theta}}_{j,(2)}$ must be rejected in the second iteration of Algorithm 1. Following this kind of argument, we are able to show that (A.2) implies that all those edges corresponding to $\{\widetilde{\boldsymbol{\Theta}}_{j,(s)}, 1 \leq s \leq k_\tau\}$ will be rejected according to Algorithm 1. Since the number of rejected edges is at least k_τ, we have $\psi_{j,\alpha} = 1$. Second, we show $\{\psi_{j,\alpha} = 1\} \subset \{\alpha_j \leq \alpha\}$. If $\psi_{j,\alpha} = 1\}$, we know the edges corresponding to $\{\widetilde{\boldsymbol{\Theta}}_{j,(s)}, 1 \leq s \leq k_\tau\}$ will be rejected, which immediately imply $\sqrt{n}|\widetilde{\boldsymbol{\Theta}}_{j,(1)}| \geq \widehat{c}(\alpha, E_j^{(1)})$. Regarding the edge corresponding to $\widetilde{\boldsymbol{\Theta}}_{j,(2)}$, it must get rejected in the first two iterations of Algorithm 1. In either cases, we always have $\sqrt{n}|\widetilde{\boldsymbol{\Theta}}_{j,(2)}| \geq \widehat{c}(\alpha, E_j^{(2)})$ due to $E_j^{(2)} \subset E_j^{(1)}$ and the fact that $\widehat{c}(\alpha, E) \leq \widehat{c}(\alpha, E')$ when $E \subset E'$. Finally, we establish (A.1).

Lemma A.1. Under the same conditions as Lemma 2.1, given some $1 \leq j \leq d$, we have the following results.

(i) Under the alternative hypothesis $H_{1j} : \|\boldsymbol{\Theta}_{j,-j}\|_0 \geq k_\tau$, we have for any $\alpha \in (0,1)$,

$$\lim_{(n,d)\to\infty} \mathbb{P}(\psi_{j,\alpha} = 1) = 1.$$

(ii) Additionally, suppose for any $|\boldsymbol{\Theta}_{jk}| > 0$, we also have $|\boldsymbol{\Theta}_{jk}| \geq c\sqrt{\log d/n}$ for some constant $c > 0$. Under the null hypothesis $H_{0j} : \|\boldsymbol{\Theta}_{j,-j}\|_0 < k_\tau$, we then have for any $u \in (0,1)$,

$$\lim_{(n,d)\to\infty} \mathbb{P}\left(\psi_{j,\alpha} = 1\right) \leq \alpha.$$

The proof of the above lemma is deferred to Section D.1. The maximum statistic used in our testing procedure takes the form of $T_E = \max_{(j,k)\in E} \sqrt{n}|\widetilde{\boldsymbol{\Theta}}^{\mathrm{d}}_{jk}|$. In our key proof procedure, we deal with the case where $E = \{(j,k) : \boldsymbol{\Theta}_{jk} = 0\}$. Since some of the results hold for general E, we will work with the general notations. Specifically, through out Appendices A.1 and A.2, we introduce the following notations: in order to approximate

$$T_E := \max_{(j,k)\in E} \sqrt{n}\left|(\widehat{\boldsymbol{\Theta}}^{\mathrm{d}}_{jk}/\sqrt{\widehat{\boldsymbol{\Theta}}^{\mathrm{d}}_{jj}\widehat{\boldsymbol{\Theta}}^{\mathrm{d}}_{kk}} - \boldsymbol{\Theta}_{jk}/\sqrt{\boldsymbol{\Theta}_{jj}\boldsymbol{\Theta}_{jk}})\right| \tag{A.3}$$

by the multiplier bootstrap process

$$T_E^{\mathcal{B}} := \max_{(j,k)\in E} \frac{1}{\sqrt{n\ \widehat{\boldsymbol{\Theta}}_{jj}\widehat{\boldsymbol{\Theta}}_{kk}}}\left|\sum_{i=1}^{n} \widehat{\boldsymbol{\Theta}}_j^\top(\boldsymbol{X}_i\boldsymbol{X}_i^\top\widehat{\boldsymbol{\Theta}}_k - \mathbf{e}_k)\xi_i\right|, \tag{A.4}$$

we define two intermediate processes:

$$\breve{T}_E := \max_{(j,k)\in E} \left|\frac{1}{\sqrt{n\ \boldsymbol{\Theta}_{jj}\boldsymbol{\Theta}_{kk}}}\sum_{i=1}^{n} \boldsymbol{\Theta}_j^\top(\boldsymbol{X}_i\boldsymbol{X}_i^\top\boldsymbol{\Theta}_k - \mathbf{e}_k)\right|, \tag{A.5}$$

$$\breve{T}_E^{\mathcal{B}} := \max_{(j,k)\in E} \left|\frac{1}{\sqrt{n\ \boldsymbol{\Theta}_{jj}\boldsymbol{\Theta}_{kk}}}\sum_{i=1}^{n} \boldsymbol{\Theta}_j^\top(\boldsymbol{X}_i\boldsymbol{X}_i^\top\boldsymbol{\Theta}_k - \mathbf{e}_k)\xi_i\right|. \tag{A.6}$$

A.1 Proof of Theorem 5.2

Proof of Theorem 5.2. Given some $j \in \mathcal{H}_0$, denote $N_{0j} = \{(j,k) : \boldsymbol{\Theta}_{jk} = 0\}$. By the first part of Lemma A.1, we have

$$\frac{\sum_{j\in\mathcal{B}}\psi_{j,\alpha}}{|\mathcal{B}|} \to 1 \quad \text{in probability}, \tag{A.7}$$

when $\alpha = \Omega(1/d)$, where $\mathcal{B} := \{j \in \mathcal{H}_0^c : \forall k \in \mathrm{supp}(\boldsymbol{\Theta}_j), |\boldsymbol{\Theta}_{jk}| > c\sqrt{\log d/n}\}$. Note that we have

$$\mathbb{P}\left(\frac{q|\mathcal{B}|}{d} \leq \widehat{\alpha} \leq 1\right) \geq \mathbb{P}\left(\frac{q|\mathcal{B}|/d\cdot d}{\sum_{j\in[d]}\psi_{j,q|\mathcal{B}|/d}} \leq q\right) = \mathbb{P}\left(\frac{|\mathcal{B}|}{\sum_{j\in[d]}\psi_{j,q|\mathcal{B}|/d}} \leq 1\right) \to 1 \tag{A.8}$$

in probability, where the first inequality is by (2.2) and the last convergence in probability is due to $q\frac{|\mathcal{B}|}{d} = \Omega(1/d)$ and (A.7). Rewrite the FDP (with $\widehat{\alpha}$) as

$$\mathrm{FDP}(\widehat{\alpha}) := \frac{\sum_{j\in\mathcal{H}_0}\psi_{j,\widehat{\alpha}}}{\max\left\{1, \sum_{j\in[d]}\psi_{j,\widehat{\alpha}}\right\}} = \frac{\widehat{\alpha}d}{\max\left\{1, \sum_{j\in[d]}\psi_{j,\widehat{\alpha}}\right\}}\cdot\frac{\sum_{j\in\mathcal{H}_0}\psi_{j,\widehat{\alpha}}}{d_0\widehat{\alpha}}\cdot\frac{d_0}{d},$$

and notice that

$$\frac{\widehat{\alpha} d}{\max\left\{1, \sum_{j\in[d]} \psi_{j,\widehat{\alpha}}\right\}} \cdot \frac{d_0}{d} \le \frac{q d_0}{d} \le q.$$

Then it suffices to control the FDP$(\widehat{\alpha})$ by dealing with $\left(\sum_{j\in\mathcal{H}_0} \psi_{j,\widehat{\alpha}}\right)/d_0\widehat{\alpha}$. By (A.8), the FDP control problem is now reduced to showing

$$\sup_{\alpha\in[\alpha_L,1]} \frac{\sum_{j\in\mathcal{H}_0} \psi_{j,\alpha}}{d_0\alpha} \le 1 + o_{\mathbb{P}}(1),$$

where $\alpha_L = q|\mathcal{B}|/d$, By (D.3) in the proof of the second part of Lemma A.1, $\psi_{j,\alpha} = 1$ implies that $\max_{e\in N_{0j}} \sqrt{n}|\widetilde{\mathbf{\Theta}}^{\mathrm{d}}_e - \mathbf{\Theta}^\star_e| \ge \widehat{c}(\alpha, N_{0j})$, where $N_{0j} = \{(j,k) : \mathbf{\Theta}_{jk} = 0\} = \{(j,k) : \mathbf{\Theta}^\star_{jk} = 0\}$. Therefore, we have

$$\frac{\sum_{j\in\mathcal{H}_0} \psi_{j,\alpha}}{d_0\alpha} \le \frac{\sum_{j\in\mathcal{H}_0} \mathbb{1}(\max_{e\in N_{0j}} \sqrt{n}|\widetilde{\mathbf{\Theta}}^{\mathrm{d}}_e - \mathbf{\Theta}^\star_e| \ge \widehat{c}(\alpha, N_{0j}))}{d_0\alpha}.$$

Hence it suffices to prove that

$$\sup_{\alpha\in[\alpha_L,1]} \left| \frac{\sum_{j\in\mathcal{H}_0} \mathbb{1}(\max_{e\in N_{0j}} \sqrt{n}|\widetilde{\mathbf{\Theta}}^{\mathrm{d}}_e - \mathbf{\Theta}^\star_e| \ge \widehat{c}(\alpha, N_{0j}))}{d_0\alpha} - 1 \right| \to 0 \quad \text{in probability.} \tag{A.9}$$

In order to prove (A.9), we construct a discrete grid of the interval $[\alpha_L, 1]$. The number of grid points is denoted by λ_d and will be decided later. First, we let $t_1 := \widehat{c}(1, N_{0j}) = 0$, $t_{\lambda_d} := \widehat{c}(\alpha_L, N_{0j})$. Here $\widehat{c}(\alpha_L, N_{0j}) = \inf\left\{t \in \mathbb{R} : \mathbb{P}_\xi\left(T^{\mathcal{B}}_{N_{0j}} \le t\right) \ge 1-\alpha\right\}$ is the quantile based on the Gaussian multiplier bootstrap process and depends on the data $\boldsymbol{X}$. Note that the involving random vectors in the Gaussian multiplier bootstrap process are Gaussian conditioning on the data $\boldsymbol{X}$ and have bounded variances with probability growing to 1. Since $\alpha_L = \Omega(1/d)$, then by the maximal inequalities for sub-Gaussian random variables (Lemma 5.2 in van Handel (2014)), we have $t_{\lambda_d} = O(\sqrt{\log d})$ with probability growing to 1. Second, note there exists h_d such that $h_d t_{\lambda_d} = o(1)$ and $t_{\lambda_d}/h_d = O(\log d)$. Based on such h_d, we construct equally spaced sequences $\{t_m\}_{m=1}^{\lambda_d}$ over the range $[t_1, t_{\lambda_d}] = [0, t_{\lambda_d}]$ with $t_m - t_{m-1} = h_d$. Then by setting α_m such that $t_m = \widehat{c}(\alpha_m, N_{0j})$, we obtain a discrete grid $\{\alpha_m\}_{m=1}^{\lambda_d}$ of the interval $[\alpha_L, 1]$. For such $\alpha_m, 1 \le m \le \lambda_d$, we have

$$\begin{aligned}\max_{1\le m\le\lambda_d} \left|\frac{\alpha_{m-1}}{\alpha_m} - 1\right| &= \max_{1\le m\le\lambda_d} \left| \frac{\mathbb{P}\left(T^{\mathcal{B}}_{N_{0j}} > t_{m-1}\right)}{\mathbb{P}\left(T^{\mathcal{B}}_{N_{0j}} > t_m\right)} - 1 \right| \\ &\le \max_{1\le m\le\lambda_d} C''(t_m - t_{m-1})(t_m+1)\exp(C'(t_m - t_{m-1})(t_m+1)) = o(1) \end{aligned} \tag{A.10}$$

with probability growing to 1, where the first equality holds by the definition of α_m, the first inequality holds due to part 2 and 3 of Theorem 2.1 in Kuchibhotla et al. (2021) (by first choosing $r-\epsilon, r+\epsilon$ in part 3 to be t_{m-1}, t_m respectively then letting $r-\epsilon, r$ in part 2 to be t_{m-1}, t_m respectively). And the right hand side of the inequality is $o(1)$ since $(t_m - t_{m-1})t_m \le h_d t_{\lambda_d} = o(1)$ with probability growing to 1.

Denote $I_j(\alpha) = \mathbb{1}(\max_{e\in N_{0j}} \sqrt{n}|\widetilde{\mathbf{\Theta}}^{\mathrm{d}}_e - \mathbf{\Theta}^\star_e| \ge \widehat{c}(\alpha, N_{0j}))$. Then given $\alpha_m \le \alpha \le \alpha_{m-1}$, for $m = 1, \cdots, \lambda_d$, we have

$$\frac{\sum_{j\in\mathcal{H}_0} I_j(\alpha_m)}{d_0\alpha_m} \cdot \frac{\alpha_m}{\alpha_{m-1}} \le \frac{\sum_{j\in\mathcal{H}_0} I_j(\alpha)}{d_0\alpha} \le \frac{\sum_{j\in\mathcal{H}_0} I_j(\alpha_{m-1})}{d_0\alpha_{m-1}} \cdot \frac{\alpha_{m-1}}{\alpha_m}. \tag{A.11}$$

Hence by (A.10) and (A.11), showing (A.9) is reduced to proving

$$\max_{1\le m\le\lambda_d}\left|\frac{\sum_{j\in\mathcal{H}_0} I_j(\alpha_m)}{d_0\alpha_m}-1\right|\to 0,\quad \text{in probability.} \tag{A.12}$$

Then it suffices to show that, for any $\epsilon>0$,

$$\mathbb{P}\left(\max_{1\le m\le\lambda_d}\left|\frac{\sum_{j\in\mathcal{H}_0} I_j(\alpha_m)}{d_0\alpha_m}-1\right|\ge\epsilon\right)\to 0.$$

By the union bound argument and Chebyshev's inequality, we have

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\le m\le\lambda_d}\left|\frac{\sum_{j\in\mathcal{H}_0} I_j(\alpha_m)}{d_0\alpha_m}-1\right|\ge\epsilon\right)\\
\le\ & \sum_{m=1}^{\lambda_d}\mathbb{P}\left(\left|\frac{\sum_{j\in\mathcal{H}_0} I_j(\alpha_m)}{d_0\alpha_m}-1\right|\ge\epsilon\right)\\
\le\ & \sum_{m=1}^{\lambda_d}\frac{\mathbb{E}\left[\sum_{j\in\mathcal{H}_0} I_j(\alpha_m)-d_0\alpha_m\right]^2}{\epsilon^2 d_0^2\alpha_m^2} &\text{(A.13)}\\
=\ & \underbrace{\sum_{m=1}^{\lambda_d}\frac{\sum_{j\in\mathcal{H}_0}\mathrm{Var}\left(I_j(\alpha_m)-d_0\alpha_m\right)}{\epsilon^2 d_0^2\alpha_m^2}}_{\mathrm{III}_1}+\underbrace{\sum_{m=1}^{\lambda_d}\frac{\left(\mathbb{E}\left[\sum_{j\in\mathcal{H}_0} I_j(\alpha_m)-d_0\alpha_m\right]\right)^2}{\epsilon^2 d_0^2\alpha_m^2}}_{\mathrm{III}_2} &\text{(A.14)}\\
& +\underbrace{\sum_{m=1}^{\lambda_d}\frac{\sum_{j_1,j_2\in\mathcal{H}_0,j_1\ne j_2}\mathrm{Cov}\left(I_{j_1}(\alpha_m),I_{j_2}(\alpha_m)\right)}{\epsilon^2 d_0^2\alpha_m^2}}_{\mathrm{III}_3}.
\end{aligned}$$

By Lemma A.2 and Lemma A.3, we have

$$\begin{aligned}
\mathrm{III}_1+\mathrm{III}_2+\mathrm{III}_3 &\le \frac{C' t_{\lambda_d}}{\epsilon^2 h_d}\left(\frac{d}{d_0|\mathcal{B}|}+\eta^2(d,n)\right)+\frac{C''' d}{\epsilon^2|\mathcal{B}|d_0}\cdot\frac{t_{\lambda_d}}{h_d}\cdot\left(1+\eta(d,n)d_0+\frac{|S|\log d}{d_0 p}\right)\\
&\le \frac{C_1 t_{\lambda_d}\eta^2(d,n)}{\epsilon^2 h_d}+\frac{C_2}{\epsilon^2\rho d_0}\cdot\frac{t_{\lambda_d}}{h_d}\cdot\left(1+\eta(d,n)d_0+\frac{|S|\log d}{d_0 p}\right), \qquad \text{(A.15)}
\end{aligned}$$

where we substitute $\zeta_1=s(\log d)^2/\sqrt{n}$, $\zeta_2=1/d^2$ and $\alpha_L=q|\mathcal{B}|/d=\Omega(\rho)$ in $\eta(d,n,\zeta_1,\zeta_2,\alpha_L)$ of Lemma A.2 and note $|\mathcal{B}|>0$ then obtain the concise form $\eta(d,n)$ below,

$$\eta(d,n)=\frac{(\log d)^{19/6}}{n^{1/6}}+\frac{(\log d)^{11/6}}{\rho^{1/3}n^{1/6}}+\frac{s(\log d)^3}{n^{1/2}}+\frac{1}{d}.$$

Recall that $t_{\lambda_d}=q(\alpha_L;T^B_{N_{0j}})=O\left(\sqrt{\log d}\right)$ with probability growing to 1 and $t_{\lambda_d}/h_d=O(\log d)$. Under Assumption 5.1, we have

$$\frac{\log d}{\rho}\left(\frac{(\log d)^{19/6}}{n^{1/6}}+\frac{(\log d)^{11/6}}{\rho^{1/3}n^{1/6}}+\frac{s(\log d)^3}{n^{1/2}}\right)=o(1),\quad \frac{\log d}{\rho d_0}+\frac{(\log d)^2|S|}{\rho d_0^2 p}=o(1),$$

and thus $\mathrm{III}_1+\mathrm{III}_2+\mathrm{III}_3=o(1)$ with probability growing to 1. Therefore, we have proved (A.9), and finally establish the FDP control result below,

$$\mathrm{FDP}(\widehat{\alpha})\le q\frac{d_0}{d}+o_{\mathbb{P}}(1).$$

In order to establish FDR control, it remains to check the uniformly integrability of the random variable sequence in (A.12). Note for a sequence of random variable $R_1, R_2, \cdots$, we have $\sup_n \mathbb{E}\left[|R_n|\mathbb{1}(|R_n| > x)\right] \le x^{-1}\sup_n \mathbb{E}\left[R_n^2\right]$ by Markov's inequality. Then to show the uniform integrability of the random variable sequence $\{R_n\}_{n=1}^{\infty}$, where $R_n = \max_{1\le m\le\lambda_d}\left|\frac{\sum_{j\in\mathcal{H}_0} I_j(\alpha_m)}{d_0\alpha_m} - 1\right|$, it suffices to show $\sup_n \mathbb{E}\left[R_n^2\right] < \infty$. Indeed, we have

$$\begin{aligned}
& \sup_n \mathbb{E}\left[\left(\max_{1\le m\le\lambda_d}\left|\frac{\sum_{j\in\mathcal{H}_0} I_j(\alpha_m)}{d_0\alpha_m} - 1\right|\right)^2\right] \\
\le\ & \sup_n \sum_{m=1}^{\lambda_d} \frac{\mathbb{E}\left[\sum_{j\in\mathcal{H}_0} I_j(\alpha_m) - d_0\alpha_m\right]^2}{d_0^2\alpha_m^2} \\
=\ & \sup_n \epsilon^2(\mathrm{III}_1 + \mathrm{III}_2 + \mathrm{III}_3).
\end{aligned}$$

Since $\mathrm{III}_1 + \mathrm{III}_2 + \mathrm{III}_3 = o(1)$ with probability growing to 1, we immediately have $\sup_n \mathbb{E}\left[R_n^2\right] < \infty$, thus finally establish the FDR control result:

$$\lim_{(n,d)\to\infty} \mathrm{FDR} \le q\frac{d_0}{d}.$$

□

A.2 Ancillary lemmas for Theorem 5.2

Lemma A.2. Recalling the definitions of $\mathrm{III}_1, \mathrm{III}_2$ in (A.14), we have

$$\mathrm{III}_1 + \mathrm{III}_2 \le \frac{C' t_{\lambda_d}}{\epsilon^2 h_d}\left(\frac{1}{\rho d_0} + \eta^2(d, n, \zeta_1, \zeta_2, \alpha_L)\right),$$

where $\eta(d, n, \zeta_1, \zeta_2, \alpha_L) = O\big(\frac{(\log d)^{19/6}}{n^{1/6}} + \frac{(\log d)^{11/6}}{n^{1/6}\alpha_L^{1/3}} + \zeta_1 \log d + \frac{\zeta_2}{\alpha_L}\big)$ with $\zeta_1 = s(\log d)^2/\sqrt{n}$, $\zeta_2 = 1/d^2$.

Proof of Lemma A.2. First note the definitions of $T_E, \breve{T}_E, T_E^{\mathcal{B}}$ and $\breve{T}_E^{\mathcal{B}}$ in (A.3), (A.5), (A.4) and (A.6) respectively, then we apply Proposition C.2 to $T = T_E, T_{\boldsymbol{Y}} = \breve{T}_E, T^{\mathcal{B}} = T_E^{\mathcal{B}}, T_{\mathcal{W}} = \breve{T}_E^{\mathcal{B}}$ with $E = N_{0j}$. And we can find the terms ζ_1, ζ_2 in (C.4), (C.5) to be $s(\log d)^2/\sqrt{n}, 1/d^2$ respectively, due to(D.25) and (D.26) (i.e., the bound on the differences $T_E - T_0$, $T_E^{\mathcal{B}} - T_0^{\mathcal{B}}$) in the proof of Lemma 2.1. Thus we have

$$\left|\frac{\mathbb{P}(\max_{e\in N_{0j}} \sqrt{n}|\widetilde{\boldsymbol{\Theta}}_e^{\mathrm{d}} - \boldsymbol{\Theta}_e^{\star}| \ge \widehat{c}(\alpha, N_{0j}))}{\alpha} - 1\right| = \eta(d, n, \zeta_1, \zeta_2, \alpha_L), \tag{A.16}$$

where $\boldsymbol{\Theta}_e^{\star} = 0, e \in N_{0j}$ and $\eta(d, n, \zeta_1, \zeta_2, \alpha_L) = O\left(\frac{(\log d)^{19/6}}{n^{1/6}} + \zeta_1 \log d + \frac{\zeta_2}{\alpha_L}\right)$ with $\zeta_1 = s(\log d)^2/\sqrt{n}$, $\zeta_2 = 1/d^2$. Recalling the definition of III_2 in (A.14), we have

$$\mathrm{III}_2 = \sum_{m=1}^{\lambda_d} \frac{\left(\mathbb{E}\left[\sum_{j\in\mathcal{H}_0} I_j(\alpha_m) - d_0\alpha_m\right]\right)^2}{\epsilon^2 d_0^2 \alpha_m^2},$$

where $I_j(\alpha) = \mathbb{1}(\max_{e\in N_{0j}} \sqrt{n}|\widetilde{\boldsymbol{\Theta}}^{\mathrm{d}}_e - \boldsymbol{\Theta}^{\star}_e| \geq \widehat{c}(\alpha, N_{0j}))$. Note that $\alpha_m \in [\alpha_L, 1]$, $\forall\ 1 \leq m \leq \lambda_d$, then we arrive at the following bound

$$\mathrm{III}_2 \leq \frac{\lambda_d}{\epsilon^2} \cdot \eta^2(d, n, \zeta_1, \zeta_2, \alpha_L) \leq \frac{t_{\lambda_d}}{\epsilon^2 h_d} \cdot \eta^2(d, n, \zeta_1, \zeta_2, \alpha_L) \tag{A.17}$$

up to some constant, where the first inequality holds by (A.16). As for the second inequality, we recall the construction of $\{t_m\}_{m=1}^{\lambda_d}$ (over the course of derivations from (A.9) to (A.10)) in the proof of Theorem 5.2 thus note $\alpha_1 = 1, t_1 = 0$ and $t_{\lambda_d} - t_1 = \sum_{m=2}^{\lambda_d}(t_m - t_{m-1}) = (\lambda_d - 1)h_d$. Regarding the term III_1, we have

$$\begin{aligned}
\mathrm{III}_1 &= \sum_{m=1}^{\lambda_d} \frac{\sum_{j\in\mathcal{H}_0} \mathrm{Var}\left(I_j(\alpha_m) - d_0\alpha_m\right)}{\epsilon^2 d_0^2 \alpha_m^2} \\
&= \sum_{m=1}^{\lambda_d} \frac{\sum_{j\in\mathcal{H}_0} \mathbb{E}(I_j(\alpha_m))(1 - \mathbb{E}(I_j(\alpha_m))}{\epsilon^2 d_0^2 \alpha_m^2} \leq \frac{1}{\epsilon^2 d_0} \sum_{m=1}^{\lambda_d} \frac{C}{\alpha_m} \leq \frac{C}{\epsilon^2 d_0 \alpha_L} \cdot \frac{t_{\lambda_d}}{h_d},
\end{aligned} \tag{A.18}$$

where the first inequality holds due to (A.16) and the second inequality holds since $\alpha_m \geq \alpha_L$ $\forall\ 1 \leq m \leq \lambda_d$ and $t_{\lambda_d} = (\lambda_d - 1)h_d$. Therefore, combining (A.17) with (A.18), we obtain

$$\mathrm{III}_1 + \mathrm{III}_2 \leq \frac{1}{\epsilon^2} \cdot \frac{t_{\lambda_d}}{h_d} \left(\frac{C}{d_0 \alpha_L} + \eta^2(d, n, \zeta_1, \zeta_2, \alpha_L)\right) \leq \frac{C' t_{\lambda_d}}{\epsilon^2 h_d} \left(\frac{1}{\rho d_0} + \eta^2(d, n, \zeta_1, \zeta_2, \alpha_L)\right)$$

for some constant C', where the second inequality holds by the definition $\alpha_L = q|\mathcal{B}|/d$ in the proof of Theorem 5.2 and the definition $\rho = |\mathcal{B}|/d$ in Section 5. □

Lemma A.3. Recalling the definition of III_3 in (A.14), we have

$$\mathrm{III}_3 \leq \frac{C''' t_{\lambda_d}}{\rho \epsilon^2 d_0 h_d} \left(1 + \eta(d, n, \zeta_1, \zeta_2, \alpha_L) d_0 + \frac{|S| \log d}{d_0 p}\right),$$

where $\eta(d, n, \zeta_1, \zeta_2, \alpha_L) = O\big(\frac{(\log d)^{19/6}}{n^{1/6}} + \frac{(\log d)^{11/6}}{n^{1/6}\alpha_L^{1/3}} + \zeta_1 \log d + \frac{\zeta_2}{\alpha_L}\big)$ with $\zeta_1 = s(\log d)^2/\sqrt{n}$, $\zeta_2 = 1/d^2$.

Proof of Lemma A.3. Note that III_3 in (A.14) equals

$$\begin{aligned}
\mathrm{III}_3 = \sum_{m=1}^{\lambda_d} &\frac{\sum_{j_1, j_2 \in \mathcal{H}_0, j_1 \neq j_2} \mathrm{Cov}\left(I_{j_1}(\alpha_m), I_{j_2}(\alpha_m)\right)}{\epsilon^2 d_0^2 \alpha_m^2}, \\
&\text{where } I_j(\alpha) = \mathbb{1}(\max_{e\in N_{0j}} \sqrt{n}|\widetilde{\boldsymbol{\Theta}}^{\mathrm{d}}_e - \boldsymbol{\Theta}^{\star}_e| \geq \widehat{c}(\alpha, N_{0j}))
\end{aligned} \tag{A.19}$$

for $j \in \{j_1, j_2\}$. To quantify the covariance between $I_{j_1}(\alpha_m)$ and $I_{j_2}(\alpha_m)$ for $j_1, j_2 \in \mathcal{H}_0, j_1 \neq j_2$, we define

$$W_j(\alpha) = \mathbb{1}(\max_{e\in N_{0j}} |Z_e| \geq c(\alpha, N_{0j})), \tag{A.20}$$

where $(Z_e)_{e\in E}$ (with $E = N_{0j}$) is a Gaussian random vector and shares the same mean vector and covariance matrix as the term $(\frac{1}{\sqrt{n\ \boldsymbol{\Theta}_{jj}\boldsymbol{\Theta}_{kk}}} \sum_{i=1}^n \boldsymbol{\Theta}_j^\top(\boldsymbol{X}_i\boldsymbol{X}_i^\top\boldsymbol{\Theta}_k - \mathbf{e}_k))_{(j,k)\in E}$ in $\breve{T}_E$. Here $\breve{T}_E$ (with $E = N_{0j}$) has the explicit form below

$$\breve{T}_E = \max_{(j,k)\in E} \frac{1}{\sqrt{n\ \boldsymbol{\Theta}_{jj}\boldsymbol{\Theta}_{kk}}} \left|\sum_{i=1}^n \boldsymbol{\Theta}_j^\top(\boldsymbol{X}_i\boldsymbol{X}_i^\top\boldsymbol{\Theta}_k - \mathbf{e}_k)\right|.$$

Remark here $\breve{T}_E$ corresponds to the term $T_{\boldsymbol{Y}}$ in Proposition C.2 and $\max_{e\in E}|Z_e|$ corresponds to the term $T_{\boldsymbol{Z}}$ in Proposition C.1. And $c(\alpha, N_{0j})$ is the corresponding Gaussian maxima quantile $q(\alpha; T_{\boldsymbol{Z}})$ (which does not need to be computed). Since $\mathbb{P}(T_{\boldsymbol{Z}} > q(\alpha; T_{\boldsymbol{Z}})) = \alpha$, we immediately have have $\mathbb{E}[W_j(\alpha)] = \mathbb{P}\left(\max_{e\in N_{0j}} \sqrt{n}|Z_e| \geq c(\alpha, N_{0j})\right) = \alpha$.

Now we replace $I_{j_1}(\alpha), I_{j_2}(\alpha)$ in III_3 by $W_{j_1}(\alpha), W_{j_2}(\alpha)$ and define III_3' as

$$\mathrm{III}_3' := \sum_{m=1}^{\lambda_d} \frac{\sum_{j_1,j_2\in\mathcal{H}_0, j_1\neq j_2} \mathrm{Cov}\left(W_{j_1}(\alpha_m), W_{j_2}(\alpha_m)\right)}{\epsilon^2 d_0^2 \alpha_m^2}. \tag{A.21}$$

To bound the difference between III_3 and III_3', we first note $\mathrm{Cov}(I_{j_1}(\alpha), I_{j_2}(\alpha)) = \mathbb{E}[I_{j_1}(\alpha)I_{j_2}(\alpha)] - \mathbb{E}[I_{j_1}(\alpha)]\mathbb{E}[I_{j_2}(\alpha)]$ then separately deal with the term $|\mathbb{E}\left[I_{j_1}(\alpha)I_{j_2}(\alpha)\right] - \mathbb{E}\left[W_{j_1}(\alpha)W_{j_2}(\alpha)\right]|$ and the term $|\mathbb{E}\left[I_{j_1}(\alpha)\right]\mathbb{E}\left[I_{j_2}(\alpha)\right] - \mathbb{E}\left[W_{j_1}(\alpha)\right]\mathbb{E}\left[G_{j_2}(\alpha)\right]|$.

By Lemma A.5, we have up to some constant factor,

$$\frac{|\mathbb{E}\left[I_{j_1}(\alpha)I_{j_2}(\alpha)\right] - \mathbb{E}\left[W_{j_1}(\alpha)W_{j_2}(\alpha)\right]|}{\alpha^2} \leq \frac{\eta(d,n,\zeta_1,\zeta_2,\alpha_L)}{\alpha}.$$

Applying the same strategy to the term $\mathbb{E}\left[I_{j_1}(\alpha)\right]\mathbb{E}\left[I_{j_2}(\alpha)\right]$, we obtain

$$\frac{|\mathbb{E}\left[I_{j_1}(\alpha)\right]\mathbb{E}\left[I_{j_2}(\alpha)\right] - \mathbb{E}\left[W_{j_1}(\alpha)\right]\mathbb{E}\left[G_{j_2}(\alpha)\right]|}{\alpha^2} \leq \frac{\eta(d,n,\zeta_1,\zeta_2,\alpha_L)}{\alpha}.$$

Combining the above two inequalities, and noting the definition of III_3' in (A.21), we derive the following bound on the difference between III_3 and III_3',

$$\left|\mathrm{III}_3 - \mathrm{III}_3'\right| \leq \frac{1}{\epsilon^2}\sum_{m=1}^{\lambda_d} \frac{\eta(d,n,\zeta_1,\zeta_2,\alpha_L)}{\alpha_m} \leq \frac{C' t_{\lambda_d}}{\rho\epsilon^2 h_d} \cdot \eta(d,n,\zeta_1,\zeta_2,\alpha_L).$$

where the second inequality holds due to the fact $\alpha_m \geq \alpha_L \ \forall\ 1 \leq m \leq \lambda_d$ and $t_{\lambda_d} = (\lambda_d - 1)h_d$, the definition $\alpha_L = q|\mathcal{B}|/d$ in the proof of Theorem 5.2, and the definition $\rho = |\mathcal{B}|/d$ in Section 5.

The above bound on $|\mathrm{III}_3 - \mathrm{III}_3'|$, when combined with Lemma A.4, immediately establishes

$$\begin{aligned}
\mathrm{III}_3 &\leq \frac{C' t_{\lambda_d}}{\rho\epsilon^2 h_d} \cdot \eta(d,n,\zeta_1,\zeta_2,\alpha_L) + \frac{C'' t_{\lambda_d}}{\rho\epsilon^2 d_0 h_d}\left(1 + C_{\boldsymbol{\Theta}}\frac{|S|\log d}{d_0 p}\right) \\
&\leq \frac{C''' t_{\lambda_d}}{\rho\epsilon^2 d_0 h_d}\left(1 + \eta(d,n,\zeta_1,\zeta_2,\alpha_L)d_0 + \frac{|S|\log d}{d_0 p}\right),
\end{aligned}$$

for some constant C'''. □

Lemma A.4. Recalling the term III_3' from (A.21) in the proof of Lemma A.3, we have

$$\mathrm{III}_3' = \sum_{m=1}^{\lambda_d} \frac{\sum_{j_1,j_2\in\mathcal{H}_0, j_1\neq j_2} \mathrm{Cov}(W_{j_1}(\alpha_m), W_{j_2}(\alpha_m))}{\epsilon^2 d_0^2 \alpha_m^2} \leq \frac{C'' t_{\lambda_d}}{\rho\epsilon^2 d_0 h_d}\left(1 + C_{\boldsymbol{\Theta}}\frac{|S|\log d}{d_0 p}\right).$$

Proof of Lemma A.4. Similarly as in the proof of Lemma A.3, we define $(Z_e)_{e\in N_{0j_1}\cup N_{0j_2}}$ to be jointly Gaussian such that this $(|N_{0j_1}|+|N_{0j_2}|)$-dimensional Gaussian random vector shares the same mean vector and covariance matrix as the term $\left(\frac{1}{\sqrt{n\ \boldsymbol{\Theta}_{jj}\boldsymbol{\Theta}_{kk}}}\sum_{i=1}^n \boldsymbol{\Theta}_j^\top(\boldsymbol{X}_i\boldsymbol{X}_i^\top\boldsymbol{\Theta}_k - \mathbf{e}_k)\right)_{(j,k)\in N_{0j_1}\cup N_{0j_2}}$. Note that the two sub-vectors $(Z_e)_{e\in N_{0j_1}}$ and $(Z_e)_{e\in N_{0j_1}}$ are generally dependent. Then we define $(Z_e')_{e\in N_{0j_1}}, (Z_e')_{e\in N_{0j_2}}$ to be two Gaussian random vectors such that

$$(Z_e')_{e\in N_{0j_1}} \overset{d}{=} (Z_e)_{e\in N_{0j_1}},\quad (Z_e')_{e\in N_{0j_2}} \overset{d}{=} (Z_e)_{e\in N_{0j_2}} \quad \text{and } (Z_e')_{e\in N_{0j_1}} \perp\!\!\!\perp (Z_e')_{e\in N_{0j_2}}. \tag{A.22}$$

Recalling the definition of $W_j(\alpha)$ in (A.20): $W_j(\alpha) = \mathbb{1}(\max_{e\in N_{0j}} |Z_e| \geq c(\alpha, N_{0j}))$, we thus have the following,

$$\begin{aligned}
\mathrm{IV}_{j_1 j_2}(\alpha) &:= \frac{|\mathrm{Cov}(W_{j_1}(\alpha_m), W_{j_2}(\alpha_m))|}{\alpha^2} \qquad \text{(A.23)}\\
&= \frac{|\mathbb{E}\left[W_{j_1}(\alpha)W_{j_2}(\alpha)\right] - \mathbb{E}\left[W_{j_1}(\alpha)\right]\mathbb{E}\left[W_{j_2}(\alpha)\right]|}{\alpha^2}\\
&= \frac{1}{\alpha^2}\Big|\mathbb{P}(\max_{e\in N_{0j_1}} |Z_e| \geq c(\alpha, N_{0j_1}), \max_{e\in N_{0j_2}} |Z_e| \geq c(\alpha, N_{0j_2})) -\\
&\quad \mathbb{P}(\max_{e\in N_{0j_1}} |Z'_e| \geq c(\alpha, N_{0j_1}), \max_{e\in N_{0j_2}} |Z'_e| \geq c(\alpha, N_{0j_2}))\Big|\\
&= \frac{1}{\alpha^2}\Big|\mathbb{P}(\max_{e\in N_{0j_1}} |Z_e| \geq t, \max_{e\in N_{0j_2}} |Z_e| \geq t) - \mathbb{P}(\max_{e\in N_{0j_1}} |Z'_e| \geq t, \max_{e\in N_{0j_2}} |Z'_e| \geq t)\Big|\\
&= \frac{1}{\alpha^2}\Big|\mathbb{P}(\max_{e\in N_{0j_1}\cup N_{0j_2}} |Z_e| \geq t) - \mathbb{P}(\max_{e\in N_{0j_1}\cup N_{0j_2}} |Z'_e| \geq t)\Big|, \qquad \text{(A.24)}
\end{aligned}$$

where the third equality follows due to the construction of $(Z_e)_{e\in N_{0j_1}\cup N_{0j_2}}, (Z'_e)_{e\in N_{0j_1}\cup N_{0j_2}}$. Note that in the fourth equality, we assume $c(\alpha, N_{0j_1}) = c(\alpha, N_{0j_2}) := t$ without loss of generality, since we can rescale one of the maximum statistic by rescaling the Gaussian random vectors. Remark that the scaling will not break down the application of Theorem 3.2, which will be explained in detail later in this proof. The last inequality holds by (A.22) and the fact that $\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B)$.

Notice that we can apply the Cramér-type Gaussian comparison bound with ℓ_0 norm to control (A.24). Specifically, we first figure out the difference between the covariance matrices of $(Z_e)_{e\in N_{0j_1}\cup N_{0j_2}}$ and $(Z'_e)_{e\in N_{0j_1}\cup N_{0j_2}}$. Denote the covariance matrices by $\mathbf{\Sigma}^Z$ and $\mathbf{\Sigma}^{Z'}$ respectively. As these two Gaussian random vectors have two sub-vectors, we write their covariance matrices in a block form

$$\mathbf{\Sigma}^Z = \begin{pmatrix} \mathbf{\Sigma}^Z_{11} & \mathbf{\Sigma}^Z_{12} \\ \mathbf{\Sigma}^Z_{21} & \mathbf{\Sigma}^Z_{22} \end{pmatrix}, \quad \mathbf{\Sigma}^{Z'} = \begin{pmatrix} \mathbf{\Sigma}^{Z'}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \mathbf{\Sigma}^{Z'}_{22} \end{pmatrix}.$$

where $\mathbf{\Sigma}^{Z'}$ is block diagonal due to (A.22). Note that we also have $\mathbf{\Sigma}^Z_{11} = \mathbf{\Sigma}^{Z'}_{11}$ and $\mathbf{\Sigma}^Z_{22} = \mathbf{\Sigma}^{Z'}_{22}$. Then we have

$$\mathbf{\Sigma}^Z - \mathbf{\Sigma}^{Z'} = \begin{pmatrix} \boldsymbol{O} & \mathbf{\Sigma}^Z_{12} \\ \mathbf{\Sigma}^Z_{21} & \boldsymbol{O} \end{pmatrix}. \qquad \text{(A.25)}$$

Throughout the following proof, we assume $\mathbf{\Theta}_{jj} = 1, j \in [d]$ without loss of generality, since the standardized version is considered in $\breve{T}_E$ (A.5). Recall that $(Z_e)_{e\in N_{0j_1}\cup N_{0j_2}}$ shares the same covariance structure as $(Y_e)_{e\in N_{0j_1}\cup N_{0j_2}}$ where Y_e (with $e = (j,k)$) is defined as

$$Y_e := \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{\Theta}_j^\top(\boldsymbol{X}_i\boldsymbol{X}_i^\top\mathbf{\Theta}_k - \mathbf{e}_k).$$

Then we are ready to calculate the covariance matrix $\mathbf{\Sigma}^Z$. Specifically, we compute the entries in each block. Regarding the block $\mathbf{\Sigma}^Z_{11}$, for any $k, k' \in N_{0j_1}$ where $N_{0j_1} = \{k : \mathbf{\Theta}_{j_1k} = 0\}$, we have the corresponding (k, k') entry in $\mathbf{\Sigma}^Z_{11}$ equals

$$\mathrm{Cov}(\mathbf{\Theta}_{j_1}^\top(\boldsymbol{X}_i\boldsymbol{X}_i^\top\mathbf{\Theta}_k - \mathbf{e}_k), \mathbf{\Theta}_{j_1}^\top(\boldsymbol{X}_i\boldsymbol{X}_i^\top\mathbf{\Theta}_{k'} - \mathbf{e}_{k'})) = \mathbf{\Theta}_{j_1j_1}\mathbf{\Theta}_{kk'} + \mathbf{\Theta}_{j_1k}\mathbf{\Theta}_{j_1k'} = \mathbf{\Theta}_{kk'}, \qquad \text{(A.26)}$$

by applying Isserlis' theorem (Isserlis, 1918) and noting $\mathbf{\Theta}_{j_1k} = \mathbf{\Theta}_{j_1k'} = 0$. Similar results hold for the block $\mathbf{\Sigma}^Z_{22}$. Regarding the block $\mathbf{\Sigma}^Z_{12}$, consider $k_1 \in N_{0j_1}, k_2 \in N_{0j_2}$, then we have the corresponding (k_1, k_2) entry in the block equals

$$\mathrm{Cov}(\boldsymbol{\Theta}_{j_1}^\top(\boldsymbol{X}_i\boldsymbol{X}_i^\top\boldsymbol{\Theta}_{k_1} - \mathbf{e}_{k_1}), \boldsymbol{\Theta}_{j_2}^\top(\boldsymbol{X}_i\boldsymbol{X}_i^\top\boldsymbol{\Theta}_{k_2} - \mathbf{e}_{k_2})) = \boldsymbol{\Theta}_{j_1j_2}\boldsymbol{\Theta}_{k_1k_2} + \boldsymbol{\Theta}_{j_1k_2}\boldsymbol{\Theta}_{j_2k_1}. \tag{A.27}$$

Now we have fully characterized the covariance matrix $\boldsymbol{\Sigma}^Z$ and the covariance matrix difference in (A.25) for any $j_1, j_2 \in \mathcal{H}_0, j_1 \neq j_2$. Specifically, we have $||\boldsymbol{\Sigma}^Z - \boldsymbol{\Sigma}^{Z'}||_0 = ||\boldsymbol{\Sigma}_{12}^Z||_0 = \sum_{k_1\in N_{0j_1}, k_2\in N_{0j_2}} \mathbb{1}(\boldsymbol{\Theta}_{j_1j_2}\boldsymbol{\Theta}_{k_1k_2} + \boldsymbol{\Theta}_{j_1k_2}\boldsymbol{\Theta}_{j_2k_1} \neq 0)$. Based on whether $\boldsymbol{\Theta}_{j_1j_2}$ is zero or not, we consider the following two cases then handle them separately:

- Case 1: $\boldsymbol{\Theta}_{j_1j_2} = 0$. If $k_1 = k_2$, then we have the covariance matrix entry (A.27) equal zero; If $k_1 \neq k_2$, then (A.27) is nonzero only if $\boldsymbol{\Theta}_{j_1k_2} \neq 0, \boldsymbol{\Theta}_{j_2k_1} \neq 0$ (i.e., $k_2 \notin N_{0j_1}$, $k_1 \notin N_{0j_2}$). By the fact $j_1, j_2 \in \mathcal{H}_0, j_1 \neq j_2$ and the definition of $\mathcal{H}_0 = \{j : \|\boldsymbol{\Theta}_{j,-j}\|_0 < k_\tau\}$, we have $\#\{(k_1, k_2) : k_1 \neq k_2, \boldsymbol{\Theta}_{j_1k_2} \neq 0, \boldsymbol{\Theta}_{j_2k_1} \neq 0\} \leq k_\tau^2$. Hence $||\boldsymbol{\Sigma}^Z - \boldsymbol{\Sigma}^{Z'}||_0 \leq k_\tau^2$.
- Case 2: $\boldsymbol{\Theta}_{j_1j_2} \neq 0$. The covariance matrix entry (A.27) is nonzero only if $\boldsymbol{\Theta}_{j_1k_2} \neq 0, \boldsymbol{\Theta}_{j_2k_1} \neq 0$ (i.e., $k_2 \notin N_{0j_1}$, $k_1 \notin N_{0j_2}$) or $\boldsymbol{\Theta}_{k_1k_2} \neq 0$.

We start from the simpler case, i.e., Case 2 where $\boldsymbol{\Theta}_{j_1j_2} \neq 0$. Simply, we obtain

$$\mathrm{IV}_{j_1j_2}(\alpha) = \frac{|\,\mathrm{Cov}(W_{j_1}(\alpha), W_{j_2}(\alpha))|}{\alpha^2} \leq \frac{\mathrm{Var}(W_{j_1}(\alpha))}{\alpha^2} + \frac{\mathrm{Var}(W_{j_2}(\alpha))}{\alpha^2} \leq \frac{C}{\alpha},$$

for some constant C since $\mathrm{Var}\left(W_j(\alpha)\right) = \mathbb{E}\left[W_j(\alpha)\right]\left(1 - \mathbb{E}\left[W_j(\alpha)\right]\right) = \alpha(1-\alpha)$ for $j = j_1, j_2$. For a fixed j_1, we also know that $|\{j_2 \in \mathcal{H}_0 : j_2 \neq j_1, \boldsymbol{\Theta}_{j_1j_2} \neq 0\}| < k_\tau$. Then we have

$$\sum_{m=1}^{\lambda_d} \sum_{\boldsymbol{\Theta}_{j_1j_2}\neq 0} \frac{\mathrm{IV}_{j_1j_2}(\alpha_m)}{\epsilon^2 d_0^2} \leq \sum_{m=1}^{\lambda_d} \frac{d_0 k_\tau}{\epsilon^2 d_0^2} \cdot \frac{C}{\alpha_m} \leq \frac{1}{\epsilon^2 d_0} \sum_{m=1}^{\lambda_d} \frac{C'}{\alpha_m}, \tag{A.28}$$

where the last inequality holds due to the same derivations for III_1 in the proof of Lemma A.2.

Regarding Case 1 where $\boldsymbol{\Theta}_{j_1j_2} = 0$, we will give a more careful treatment to $\mathrm{IV}_{j_1j_2}(\alpha)$ in (A.23). Due to the discussion about Case 1, we have $||\boldsymbol{\Sigma}^Z - \boldsymbol{\Sigma}^{Z'}||_0 \leq k_\tau^2$. This fact will be utilized to derive a nice bound on III_3'. Indeed, we can apply Theorem 3.2 to (A.24) (with U and V chosen to be $Z_e)_{e\in N_{0j_1}\cup N_{0j_2}}$ and $(Z_e')_{e\in N_{0j_1}\cup N_{0j_2}}$ respectively) and obtain

$$\mathrm{IV}_{j_1j_2}(\alpha) \leq \frac{\log d}{\alpha p}\left(\sum_{k_1\in N_{0j_1}, k_2\in N_{0j_2}, k_1\neq k_2} \mathbb{1}(\boldsymbol{\Theta}_{j_1k_2}\boldsymbol{\Theta}_{j_2k_1} \neq 0)\right). \tag{A.29}$$

when $\boldsymbol{\Theta}_{j_1j_2} = 0$ (i.e., under Case 1). Recall Theorem 3.2 assumes for Gaussian random vectors U and V, there exists a disjoint $\mathfrak{p}$-partition of nodes $\cup_{\ell=1}^{\mathfrak{p}}\mathcal{C}_\ell = [d]$ such that $\sigma_{jk}^U = \sigma_{jk}^V = 0$ when $j \in \mathcal{C}_\ell$ and $k \in \mathcal{C}_{\ell'}$ for some $\ell \neq \ell'$. This is the connectivity assumption. Theorem 3.2 also assumes that U and V have unit variances i.e., $\sigma_{jj}^U = \sigma_{jj}^V = 1, j \in [d]$ and there exists some $\sigma_0 < 1$ such that $|\sigma_{jk}^V| \leq \sigma_0$ for any $j \neq k$ and $|\{(j,k) : j \neq k, |\sigma_{jk}^U| > \sigma_0\}| \leq b_0$ for some constant b_0. Under its general version (which is actually proved in Appendix B.2), we only need to assume $a_0 \leq \sigma_{jj}^U = \sigma_{jj}^V \leq a_1$, $\forall j \in [d]$, and given any $j \in \mathcal{C}_\ell$ with some ℓ, there exists at least one $m \in \mathcal{C}_{\ell'}$ such that $\sigma_{jj}^U = \sigma_{jj}^V = \sigma_{mm}^U = \sigma_{mm}^V$ for any $\ell' \neq \ell$. From now, we will call it the general variance condition. Accordingly, we assume there exists some $\sigma_0 < 1$ such that $|\sigma_{jk}^V/\sqrt{\sigma_{jj}^V\sigma_{kk}^V}| \leq \sigma_0$ for any $j \neq k$ and $|\{(j,k) : j \neq k, |\sigma_{jk}^U|\sqrt{\sigma_{jj}^U\sigma_{kk}^U} > \sigma_0\}| \leq b_0$ for some constant b_0. Such condition is referred as the general covariance assumption. Below we give the details of applying Theorem 3.2 (with a general version of the variance assumption) by checking those three conditions.

We start from the connectivity assumption and the general variance condition. Notice that in Section 5, p denotes the number of connected components in the associated graph $\mathcal{G}$ of $\boldsymbol{X}$. Then we know there exist disjoint partitions of nodes $\cup_{\ell=1}^{p}\mathcal{C}_\ell^X = [d]$ such that $\boldsymbol{\Theta}_{jk} = 0$ when $j \in \mathcal{C}_\ell^X, k \in \mathcal{C}_{\ell'}^X$ for some $\ell \neq \ell'$. We will utilize this fact to examine the covariance matrices of $U := (Z_e)_{e\in N_{0j_1}\cup N_{0j_2}}$ and $V := (Z'_e)_{e\in N_{0j_1}\cup N_{0j_2}}$ and show the connectivity assumption holds. Note that for given $j_1, j_2 \in \mathcal{H}_0, j_1 \neq j_2$, there exist at least $p-2$ components $\cup_{\ell=1}^{p-2}\mathcal{C}_\ell^X$ such that j_1 and j_2 do not belong to them. Without loss of generality, we write $j_1, j_2 \notin \cup_{\ell=1}^{p-2}\mathcal{C}_\ell^X$. Thus we have $\cup_{l=1}^{p-2}\mathcal{C}_\ell^X \subset N_{0j_1} \cap N_{0j_2}$ by definition.

In the following, we will show the number of connected components on the associated graph of the Gaussian random vector $U := (Z_e)_{e\in N_{0j_1}\cup N_{0j_2}}$ is at least $2(p-2)$ by examining its covariance matrix $\boldsymbol{\Sigma}_Z$. First we focus on the covariance entries in the block $\boldsymbol{\Sigma}_{11}^Z$. When $\ell_1, \ell_2 \in [p-2]$ and $\ell_1 \neq \ell_2$, we have for any $k \in \mathcal{C}_{\ell_1}^X, k' \in \mathcal{C}_{\ell_2}^X$ (thus $k, k' \in N_{0j_1} \cap N_{0j_2}$), the (k, k') covariance entry (A.26) in the block $\boldsymbol{\Sigma}_{11}^Z$ equals

$$\boldsymbol{\Theta}_{j_1j_1}\boldsymbol{\Theta}_{kk'} + \boldsymbol{\Theta}_{j_1k}\boldsymbol{\Theta}_{j_1k'} = \boldsymbol{\Theta}_{j_1j_1}\boldsymbol{\Theta}_{kk'} = 0, \tag{A.30}$$

where the first equality holds since $k, k' \in N_{0j_1}$, and the second equality holds since $\ell_1 \neq \ell_2$. Similarly, we have the (k, k') covariance entry in the block $\boldsymbol{\Sigma}_{22}^Z$ also equals to zero. Next we compute the covariance entries in the block $\boldsymbol{\Sigma}_{12}^Z$. For the same (k, k'), we know that $k \in N_{0j_1}, k' \in N_{0j_2}$. Thus the corresponding covariance entry (A.27) equals

$$\boldsymbol{\Theta}_{j_1j_2}\boldsymbol{\Theta}_{kk'} + \boldsymbol{\Theta}_{j_1k'}\boldsymbol{\Theta}_{j_2k} = 0, \tag{A.31}$$

since we also have $k \in N_{0j_2}, k' \in N_{0j_1}$ and $k \in \mathcal{C}_{\ell_1}^X, k' \in \mathcal{C}_{\ell_2}^X$ for some $\ell_1 \neq \ell_2$. Denote the nodes in the associated graph of $\boldsymbol{\Sigma}^Z$ by $\mathcal{V}_Z := \{(j,k) : k \in N_{0j}, j = j_1, j_2\}$. Remark here we use a pair (j,k) to represent a node since there exists some $k \in N_{0j_1} \cap N_{0j_2}$ and we have to distinguish the covariance entries (j_1, k) and (j_2, k). Based on previous calculations, we immediately find $\cup_{\ell=1}^{2(p-2)}\mathcal{C}_\ell^Z \subset \mathcal{V}_Z$, where $\mathcal{C}_\ell^Z$ is chosen to be

$$\mathcal{C}_\ell^Z = \begin{cases} \{(j_1,k) : k \in \mathcal{C}_\ell^X\} & \text{when } 1 \leq \ell \leq p-2, \\ \{(j_2,k) : k \in \mathcal{C}_\ell^X\} & \text{when } p-1 \leq \ell \leq 2(p-2). \end{cases} \tag{A.32}$$

Further, we know they form different components on the associated graph of $\boldsymbol{\Sigma}^Z$. This is due to (A.30) and (A.31). The above results also apply to the Gaussian random vector $V := (Z'_e)_{e\in N_{0j_1}\cup N_{0j_2}}$ by construction of Z'_e, i.e., we have the same subset of nodes $\cup_{\ell=1}^{2(p-2)}\mathcal{C}_\ell^Z \subset \mathcal{V}_Z$ from different components on the associated graph of $\boldsymbol{\Sigma}^{Z'}$.

When $k \in \mathcal{C}_\ell$ for some $\ell \in [p-2]$, the corresponding diagonal entries of the covariance matrices $\boldsymbol{\Sigma}^Z, \boldsymbol{\Sigma}^{Z'}$ equal

$$\boldsymbol{\Theta}_{j_1j_1}\boldsymbol{\Theta}_{kk} + \boldsymbol{\Theta}_{j_1k}\boldsymbol{\Theta}_{j_1k} = \boldsymbol{\Theta}_{j_1j_1}\boldsymbol{\Theta}_{kk} = 1 = \boldsymbol{\Theta}_{j_2j_2}\boldsymbol{\Theta}_{kk},$$

where the first equality holds since $\boldsymbol{\Theta}_{j_1k} = 0$ when $k \in \mathcal{C}_\ell \subset N_{0j_1}$. As for the second equality, we use the fact that $\boldsymbol{\Theta}_{jj} = 1, j \in [d]$. This is because $\breve{T}_E$ in (A.5) considers the standardized version $\boldsymbol{\Theta}_{jk}/\sqrt{\boldsymbol{\Theta}_{jj}\boldsymbol{\Theta}_{kk}}$. Remark that the rescaling in Lemma A.4 is performed on one of the two random vectors $(Z'_e)_{e\in N_{0j_1}}, (Z'_e)_{e\in N_{0j_2}}$. Then we have the variances across the $p-2$ components $\cup_{\ell=1}^{p-2}\mathcal{C}_\ell^Z$ are the same. The variances across the other $p-2$ components $\cup_{\ell=p-1}^{2(p-2)}\mathcal{C}_\ell^Z$ are also the same. Finally, we show there exist at least $p-2$ components $\cup_{\ell=1}^{p-2}\mathcal{C}_\ell^Z$ (or $\cup_{\ell=p-1}^{2(p-2)}\mathcal{C}_\ell^Z$) satisfying the requirement in the connectivity assumption and the general variance condition.

Regarding the general covariance condition, we first note that $\mathbf{\Theta} \in \mathcal{U}(M, s, r_0)$ which says that $\lambda_{\min}(\mathbf{\Theta}) \geq 1/r_0, \lambda_{\max}(\mathbf{\Theta}) \leq r_0$. Thus we have $\max_{j,k\in[d],j\neq k}|\mathbf{\Theta}_{jk}| \leq \sigma_0$ for some $\sigma_0 < 1$. Below we will examine all the off-diagonal entries of $\mathbf{\Sigma}^Z$ and $\mathbf{\Sigma}^{Z'}$. Regarding the block $\mathbf{\Sigma}_{11}^Z$, for any $k, k' \in N_{0j_1}, k \neq k'$ where $N_{0j_1} = \{k : \mathbf{\Theta}_{j_1k} = 0\}$, (A.26) says that the corresponding (k, k') entry in $\mathbf{\Sigma}_{11}^Z$ equals $\mathbf{\Theta}_{kk'}$ (here we have $|\mathbf{\Theta}_{kk'}| \leq \sigma_0$). Similar results hold for the block $\mathbf{\Sigma}_{22}^Z$. Regarding the block $\mathbf{\Sigma}_{12}^Z$, consider $k_1 \in N_{0j_1}, k_2 \in N_{0j_2}$, then we have the corresponding (k_1, k_2) entry in the block equals $\mathbf{\Theta}_{j_1k_2}\mathbf{\Theta}_{j_2k_1}$. This is due to (A.27) and the fact that $\mathbf{\Theta}_{j_1j_2} = 0$ under Case 1. Only when $k_2 = j_1, k_1 = j_2$, we have $\mathbf{\Theta}_{j_1k_2}\mathbf{\Theta}_{j_2k_1} = 1$. Otherwise, $|\mathbf{\Theta}_{j_1k_2}\mathbf{\Theta}_{j_2k_1}| \leq \sigma_0^2 < \sigma_0$ always holds. As for the $\mathbf{\Sigma}^{Z'}$, since its block $\mathbf{\Sigma}_{12}^{Z'} = \boldsymbol{O}$, we immediately have the absolute values of all its off-diagonal entries is bounded by σ_0. In summary, we verify the covariance condition of Theorem 3.2 (here U and V are chosen to be $Z_e)_{e\in N_{0j_1}\cup N_{0j_2}}$ and $(Z'_e)_{e\in N_{0j_1}\cup N_{0j_2}}$ respectively).

Having checked all the three conditions, we now obtain

$$
\begin{aligned}
&\sum_{m=1}^{\lambda_d}\sum_{\mathbf{\Theta}_{j_1j_2}=0}\frac{\mathrm{IV}_{j_1j_2}(\alpha_m)}{\epsilon^2 d_0^2}\\
\leq\ & \sum_{m=1}^{\lambda_d}\left\{\frac{1}{\epsilon^2 d_0^2}\cdot\frac{\log d}{\alpha_m p}\left(\sum_{k_1\in N_{0j_1},k_2\in N_{0j_2},k_1\neq k_2}\mathbb{1}(\mathbf{\Theta}_{j_1k_2}\mathbf{\Theta}_{j_2k_1}\neq 0)\right)\right\}\\
\leq\ & \frac{C_{\mathbf{\Theta}}|S|\log d}{\epsilon^2 d_0 p}\left(\frac{1}{d_0}\sum_{m=1}^{\lambda_d}\frac{C'}{\alpha_m}\right),
\end{aligned}
\tag{A.33}
$$

where S represents the set

$$S = \{(j_1, j_2, k_1, k_2) : j_1, j_2 \in \mathcal{H}_0, j_1 \neq j_2, k_1 \neq k_2, \mathbf{\Theta}_{j_1j_2} = \mathbf{\Theta}_{j_1k_1} = \mathbf{\Theta}_{j_2k_2} = 0, \mathbf{\Theta}_{j_1k_2} \neq 0, \mathbf{\Theta}_{j_2k_1} \neq 0\}$$

as defined in Section 5, and $C_{\mathbf{\Theta}}$ is some universal constant over $\mathbf{\Theta} \in \mathcal{U}(M, s, r_0)$. Finally, combining (A.33) with (A.28), we obtain the following bound on III_3',

$$
\begin{aligned}
\mathrm{III}_3' &\leq \frac{C_{\mathbf{\Theta}}|S|\log d}{\epsilon^2 d_0 p}\left(\frac{1}{d_0}\sum_{m=1}^{\lambda_d}\frac{C'}{\alpha_m}\right) + \frac{1}{\epsilon^2 d_0}\sum_{m=1}^{\lambda_d}\frac{C'}{\alpha_m}\\
&= \left(1 + \frac{C_{\mathbf{\Theta}}|S|\log d}{d_0 p}\right)\cdot\frac{1}{\epsilon^2 d_0}\sum_{m=1}^{\lambda_d}\frac{C'}{\alpha_m}\\
&\leq \frac{C'' t_{\lambda_d}}{\rho\epsilon^2 d_0 h_d}\left(1 + \frac{C_{\mathbf{\Theta}}|S|\log d}{d_0 p}\right),
\end{aligned}
$$

where the last inequality holds due to the same derivations for III_1 in the proof of Lemma A.2. □

Lemma A.5. Recall the definitions of $I_j(\alpha)$ and $W_j(\alpha)$ in (A.19) and (A.20), for $j_1, j_2 \in \mathcal{H}_0, j_1 \neq j_2$, when $\alpha \in [\alpha_L, 1]$, we have

$$\left|\mathbb{E}\left[I_{j_1}(\alpha)I_{j_2}(\alpha)\right] - \mathbb{E}\left[W_{j_1}(\alpha)W_{j_2}(\alpha)\right]\right| \leq \eta(d, n, \zeta_1, \zeta_2, \alpha_L)\alpha. \tag{A.34}$$

Proof of Lemma A.5. First express $\left|\mathbb{E}\left[I_{j_1}(\alpha)I_{j_2}(\alpha)\right]-\mathbb{E}\left[W_{j_1}(\alpha)W_{j_2}(\alpha)\right]\right|$ as

$$\begin{aligned}
&\left|\mathbb{E}\left[I_{j_1}(\alpha)I_{j_2}(\alpha)\right]-\mathbb{E}\left[W_{j_1}(\alpha)W_{j_2}(\alpha)\right]\right| \\
=\ &\Big|\mathbb{P}\Big(\max_{e\in N_{0j_1}}\sqrt{n}|\widetilde{\boldsymbol{\Theta}}_e^{\mathrm{d}}|\geq\widehat{c}(\alpha,N_{0j_1}),\max_{e\in N_{0j_2}}\sqrt{n}|\widetilde{\boldsymbol{\Theta}}_e^{\mathrm{d}}|\geq\widehat{c}(\alpha,N_{0j_2})\Big) \\
&-\mathbb{P}\Big(\max_{e\in N_{0j_1}}|Z_e|\geq c(\alpha,N_{0j_1}),\max_{e\in N_{0j_1}}|Z_e|\geq c(\alpha,N_{0j_2})\Big)\Big| \\
=\ &\Big|\mathbb{P}\Big(T_{N_{0j_1}}\geq\widehat{c}(\alpha,N_{0j_1}),T_{N_{0j_2}}\geq\widehat{c}(\alpha,N_{0j_2})\Big) \\
&-\mathbb{P}\Big(\max_{e\in N_{0j_1}}|Z_e|\geq c(\alpha,N_{0j_1}),\max_{e\in N_{0j_1}}|Z_e|\geq c(\alpha,N_{0j_2})\Big)\Big|,
\end{aligned}\tag{A.35}$$

where the second equality holds by the definition of T_E in (A.3) and the definitions of N_{0j_1},N_{0j_2}. Now proving the bound in (A.34) is reduced to showing

$$\begin{aligned}
&\Big|\mathbb{P}\Big(T_{N_{0j_1}}\geq\widehat{c}(\alpha,N_{0j_1}),T_{N_{0j_2}}\geq\widehat{c}(\alpha,N_{0j_2})\Big)-\mathbb{P}\Big(\max_{e\in N_{0j_1}}|Z_e|\geq c(\alpha,N_{0j_1}),\max_{e\in N_{0j_1}}|Z_e|\geq c(\alpha,N_{0j_2})\Big)\Big| \\
&\leq\eta(d,n,\zeta_1,\zeta_2,\alpha_L)\alpha.
\end{aligned}\tag{A.36}$$

We first relate the notations in the above expression to the notations in Appendix C: $T_{N_{0j_1}},T_{N_{0j_2}}$ correspond to T; $\widehat{c}(\alpha,N_{0j_1}),\widehat{c}(\alpha,N_{0j_2})$ correspond to $q_\xi(\alpha,T^{\mathcal{B}})$; $\max_{e\in N_{0j_1}}|Z_e|,\max_{e\in N_{0j_2}}|Z_e|$ correspond to $T_{\boldsymbol{Z}}$; $c(\alpha,N_{0j_1}),c(\alpha,N_{0j_2})$ correspond to $q(\alpha;T_{\boldsymbol{Z}})$. In Appendix C, we prove Propositions C.1 and C.2. And the strategy can be used to derive the bound on (A.35). First, we note that $T_{N_{0j_1}},T_{N_{0j_2}}$ satisfy the conditions of Proposition C.2, i.e., (C.4) and (C.5). This is due to the same derivations as the first parapraph of the proof of Lemmma A.2. Since the proving strategy is quite similar, we omit the proof of (A.36) for simplicity. Instead, we prove (A.37), i.e., when $\alpha\in[\alpha_L,1]$,

$$\begin{aligned}
\mathrm{D}:=\ &\Big|\mathbb{P}\left(T_{\boldsymbol{Y}_1}\geq q_\xi(\alpha;T_{\boldsymbol{W}_1}),T_{\boldsymbol{Y}_2}\geq q_\xi(\alpha;T_{\boldsymbol{W}_2})\right)-\mathbb{P}\left(T_{\boldsymbol{Z}_1}\geq q(\alpha;T_{\boldsymbol{Z}_1}),T_{\boldsymbol{Z}_2}\geq q(\alpha;T_{\boldsymbol{Z}_2})\right)\Big| \\
\leq\ &C\alpha\left(\frac{(\log d)^{11/6}}{n^{1/6}\alpha_L^{1/3}}+\frac{(\log d)^{19/6}}{n^{1/6}}\right),
\end{aligned}\tag{A.37}$$

where $T_{\boldsymbol{Y}_1},T_{\boldsymbol{Y}_2}$ correspond to $\breve{T}_E$ with $E=N_{0j_1},N_{0j_2}$ respectively, $T_{\boldsymbol{W}_1},T_{\boldsymbol{W}_2}$ correspond to $\breve{T}_E^{\mathcal{B}}$ with $E=N_{0j_1},N_{0j_2}$ respectively, and $T_{\boldsymbol{Z}_1}=\max_{e\in N_{0j_1}}|Z_e|,T_{\boldsymbol{Z}_2}=\max_{e\in N_{0j_2}}|Z_e|$. As for the quantiles, $q_\xi(\alpha;T_{\boldsymbol{W}_1}),q_\xi(\alpha;T_{\boldsymbol{W}_2})$ are the Gaussian multiplier bootstrap quantiles based on $T_{\boldsymbol{W}_1},T_{\boldsymbol{W}_2}$. $q(\alpha;T_{\boldsymbol{Z}_1}),q(\alpha;T_{\boldsymbol{Z}_2})$ are the quantiles of the Gaussian maxima $T_{\boldsymbol{Z}_1},T_{\boldsymbol{Z}_2}$. Denote $A_1=\{T_{\boldsymbol{Y}_1}\geq q_\xi(\alpha;T_{\boldsymbol{W}_1})\}$, $A_2=\{T_{\boldsymbol{Y}_2}\geq q_\xi(\alpha;T_{\boldsymbol{W}_2})\}$, $B_1=\{T_{\boldsymbol{Y}_1}\geq q(\alpha;T_{\boldsymbol{Z}_1})\}$, $B_2=\{T_{\boldsymbol{Y}_2}\geq q(\alpha;T_{\boldsymbol{Z}_2})\}$, we have

$$\begin{aligned}
\mathrm{D}_{12}\ :=\ &\left|\mathbb{P}\left(T_{\boldsymbol{Y}_1}\geq q_\xi(\alpha;T_{\boldsymbol{W}_1}),T_{\boldsymbol{Y}_2}\geq q_\xi(\alpha;T_{\boldsymbol{W}_2})\right)-\mathbb{P}\left(T_{\boldsymbol{Y}_1}\geq q(\alpha;T_{\boldsymbol{Z}_1}),T_{\boldsymbol{Y}_2}\geq q(\alpha;T_{\boldsymbol{Z}_2})\right)\right| \\
\leq\ &\mathbb{P}\left((A_1\cap A_2)\ominus(B_1\cap B_2)\right) \\
=\ &\mathbb{P}\left((A_1\cap A_2)\cap(B_1^c\cup B_2^c)\right)+\mathbb{P}\left((B_1\cap B_2)\cap(A_1^c\cup A_2^c)\right) \\
\leq\ &\mathbb{P}\left(A_1\cap B_1^c\right)+\mathbb{P}\left(A_2\cap B_2^c\right)+\mathbb{P}\left(B_1\cap A_1^c\right)+\mathbb{P}\left(B_2\cap A_2^c\right) \\
=\ &\mathbb{P}\left((A_1\cap B_1^c)\cup(B_1\cap A_1^c)\right)+\mathbb{P}\left((A_2\cap B_2^c)\cup(B_2\cap A_2^c)\right) \\
=\ &\mathbb{P}\left(A_1\ominus B_1\right)+\mathbb{P}\left(A_2\ominus B_2\right).
\end{aligned}\tag{A.38}$$

By (C.14) and (C.15), we can bound (A.38) as

$$\mathrm{D}_{12}\leq\mathbb{P}\left(A_1\ominus B_1\right)+\mathbb{P}\left(A_2\ominus B_2\right)\leq\ 2C'\alpha\left(\frac{(\log d)^{11/6}}{n^{1/6}\alpha_L^{1/3}}+\frac{(\log d)^{19/6}}{n^{1/6}}\right).\tag{A.39}$$

By the triangle inequality, we have the following bound on D,

$$\begin{aligned}
\mathrm{D} = & \left|\mathbb{P}\left(T_{\boldsymbol{Y}_1} \geq q_\xi(\alpha; T_{\boldsymbol{W}_1}), T_{\boldsymbol{Y}_2} \geq q_\xi(\alpha; T_{\boldsymbol{W}_2})\right) - \mathbb{P}\left(T_{\boldsymbol{Z}_1} \geq q(\alpha; T_{\boldsymbol{Z}_1}), T_{\boldsymbol{Z}_2} \geq q(\alpha; T_{\boldsymbol{Z}_2})\right)\right| \\
\leq & \ \mathrm{D}_{12} + \left|\mathbb{P}\left(T_{\boldsymbol{Y}_1} \geq q(\alpha; T_{\boldsymbol{Z}_1}), T_{\boldsymbol{Y}_2} \geq q(\alpha; T_{\boldsymbol{Z}_2})\right) - \mathbb{P}\left(T_{\boldsymbol{Z}_1} \geq q(\alpha; T_{\boldsymbol{Z}_1}), T_{\boldsymbol{Z}_2} \geq q(\alpha; T_{\boldsymbol{Z}_2})\right)\right|. \\
\leq & \ \mathrm{D}_{12} + \left|\mathbb{P}\left(T_{\boldsymbol{Y}_1} \geq q(\alpha; T_{\boldsymbol{Z}_1})\right) - \mathbb{P}\left(T_{\boldsymbol{Z}_1} \geq q(\alpha; T_{\boldsymbol{Z}_1})\right)\right| + \left|\mathbb{P}\left(T_{\boldsymbol{Y}_2} \geq q(\alpha; T_{\boldsymbol{Z}_2})\right) - \mathbb{P}\left(T_{\boldsymbol{Z}_2} \geq q(\alpha; T_{\boldsymbol{Z}_2})\right)\right| \\
& + \underbrace{\left|\mathbb{P}\left(\{T_{\boldsymbol{Y}_1} \geq q(\alpha; T_{\boldsymbol{Z}_1})\} \cup \{T_{\boldsymbol{Y}_2} \geq q(\alpha; T_{\boldsymbol{Z}_2})\}\right) - \mathbb{P}\left(\{T_{\boldsymbol{Z}_1} \geq q(\alpha; T_{\boldsymbol{Z}_1})\} \cup \{T_{\boldsymbol{Z}_2} \geq q(\alpha; T_{\boldsymbol{Z}_2})\}\right)\right|}_{\mathrm{D}'_{12}},
\end{aligned} \tag{A.40}$$

where the last inequality holds since $\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B)$. For the second term and the third term in (A.40), we can directly apply the results (C.10) in Proposition C.1 and bound them as

$$\begin{aligned}
& \left|\mathbb{P}\left(T_{\boldsymbol{Y}_1} \geq q(\alpha; T_{\boldsymbol{Z}_1})\right) - \mathbb{P}\left(T_{\boldsymbol{Z}_1} \geq q(\alpha; T_{\boldsymbol{Z}_1})\right)\right| + \left|\mathbb{P}\left(T_{\boldsymbol{Y}_2} \geq q(\alpha; T_{\boldsymbol{Z}_2})\right) - \mathbb{P}\left(T_{\boldsymbol{Z}_2} \geq q(\alpha; T_{\boldsymbol{Z}_2})\right)\right| \\
& \leq C\alpha \cdot \frac{(\log d)^{19/6}}{n^{1/6}}
\end{aligned} \tag{A.41}$$

for some constant C. Regarding the term D'_{12}, we assume $q(\alpha; T_{\boldsymbol{Z}_2}) = q(\alpha; T_{\boldsymbol{Z}_2}) := t$ without loss of generality. This is because $q(\alpha; T_{\boldsymbol{Z}_1}), q(\alpha; T_{\boldsymbol{Z}_2})$ are all deterministic values and we can rescale the random vector inside one of the maximum statistics $T_{\boldsymbol{Z}_1}, T_{\boldsymbol{Z}_2}$. Now we rewrite D'_{12} based on $q(\alpha; T_{\boldsymbol{Z}_2}) = q(\alpha; T_{\boldsymbol{Z}_2}) = t$ and derive the following bound:

$$\begin{aligned}
\mathrm{D}'_{12} &= \left|\mathbb{P}\left(\max\{T_{\boldsymbol{Y}_1}, T_{\boldsymbol{Y}_2}\} \geq t\right) - \mathbb{P}\left(\max\{T_{\boldsymbol{Z}_1}, T_{\boldsymbol{Z}_2}\} \geq t\right)\right| & \text{(A.42)} \\
&\leq \frac{C''(\log d)^{19/6}}{n^{1/6}} \cdot \mathbb{P}\left(\max\{T_{\boldsymbol{Z}_1}, T_{\boldsymbol{Z}_2}\} \geq t\right) \\
&\leq \frac{C''(\log d)^{19/6}}{n^{1/6}} \cdot \left(\mathbb{P}\left(T_{\boldsymbol{Z}_1} \geq q(\alpha; T_{\boldsymbol{Z}_1})\right) + \mathbb{P}\left(T_{\boldsymbol{Z}_2} \geq q(\alpha; T_{\boldsymbol{Z}_2})\right)\right) \\
&= 2C''\alpha \cdot \frac{(\log d)^{19/6}}{n^{1/6}}, & \text{(A.43)}
\end{aligned}$$

where the first inequality holds by applying Corollary 5.1 of Kuchibhotla et al. (2021) similarly as in the derivation of (C.10). Here we briefly explain why Corollary 5.1 of Kuchibhotla et al. (2021) is applicable to (A.42). Note that $\max\{T_{\boldsymbol{Y}_1}, T_{\boldsymbol{Y}_2}\} = T_{\boldsymbol{Y}_{12}}$ is the maximum statistic with respect to the random vectors which concatenate the random vectors involved in $T_{\boldsymbol{Y}_1}, T_{\boldsymbol{Y}_2}$. Write $T_{\boldsymbol{Y}_1}, T_{\boldsymbol{Y}_2}$ explicitly as

$$T_{\boldsymbol{Y}_1} := \left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{Y}_i^{(1)}\right\|_\infty, \quad T_{\boldsymbol{Y}_2} := \left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{Y}_i^{(2)}\right\|_\infty,$$

and denote $\boldsymbol{Y}_i^{(12)} = (\boldsymbol{Y}_i^{(1)}, \boldsymbol{Y}_i^{(2)})$, then $T_{\boldsymbol{Y}_{12}}$ is defined as

$$T_{\boldsymbol{Y}_{12}} := \left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{Y}_i^{(12)}\right\|_\infty.$$

By the definition of $\boldsymbol{Z}_1, \boldsymbol{Z}_2$, we have $\mathrm{Cov}((\boldsymbol{Z}_1^\top, \boldsymbol{Z}_2^\top)^\top) = \mathrm{Cov}((\boldsymbol{Y}_1^\top, \boldsymbol{Y}_2^\top)^\top)$. Hence we can apply Corollary 5.1 of Kuchibhotla et al. (2021) to (A.42). Now we combine (A.39), (A.40), (A.41) with (A.43) and obtain the following bound

$$\mathrm{D} \le C\alpha \left(\frac{(\log d)^{11/6}}{n^{1/6}\alpha_L^{1/3}} + \frac{(\log d)^{19/6}}{n^{1/6}} \right),$$

for some constant C, thus (A.37) is established. The above strategy of obtaining (A.37) can be similarly applied to the term in (A.35), then establishes the bound in (A.34).

□

A.3 Proof of Theorem 4.2

Proof of Theorem 4.2. Throughout the proof, we condition on the design matrix $\boldsymbol{X}$, but without explicitly writing it out in order to simplify the notation. In the context of selecting hub response variables, we recall $\mathcal{H}_0 = \{j \in [d_1] : ||\boldsymbol{\Theta}_j||_0 \ge k_\tau\}$ and $d_0 = |\mathcal{H}_0|$. For a non-hub response variable $j \in \mathcal{H}_0$, let N_{0j} be the set of its null covariates, i.e., $N_{0j} = \{(j,k) : \boldsymbol{\Theta}_{jk} = 0\}$.

To establish FDR control, we follow the same derivations as in the proof of Theorem 5.2. Specifically, it suffices to bound

$$\sum_{m=1}^{\lambda_d} \frac{\mathrm{Var}[\sum_{j\in\mathcal{H}_0} I_j(\alpha_m) - d_0\alpha_m]}{\epsilon^2 d_0^2 \alpha_m^2} + \sum_{m=1}^{\lambda_d} \frac{(\mathbb{E}[\sum_{j\in\mathcal{H}_0} I_j(\alpha_m) - d_0\alpha_m])^2}{\epsilon^2 d_0^2 \alpha_m^2}$$
$$+ \sum_{m=1}^{\lambda_d} \frac{\sum_{j_1,j_2\in\mathcal{H}_0, j_1\neq j_2} \mathrm{Cov}(I_{j_1}(\alpha_m), I_{j_2}(\alpha_m))}{\epsilon^2 d_0^2 \alpha_m^2} := \mathrm{III}_1 + \mathrm{III}_2 + 0 \tag{A.44}$$

for any $\epsilon > 0$. In the above terms, the sequence $\{\alpha_m\}_{m=1}^{\lambda_d}$ is chosen similarly as in the proof of Theorem 5.2 and $I_j(\alpha)$ is defined as

$$I_j(\alpha) = \mathbb{1}(\max_{e\in N_{0j}} \sqrt{n}|\widetilde{\boldsymbol{\Theta}}_e^{\mathrm{d}}| \ge \widehat{c}(\alpha, N_{0j})),$$

where $\widetilde{\boldsymbol{\Theta}}_j^{\mathrm{d}}$ is the debiased Lasso estimator defined in (4.2). Note that the cross term in (A.44) equals zero as $\mathrm{Cov}(I_{j_1}(\alpha_m), I_{j_2}(\alpha_m)) = 0$. This is because $\boldsymbol{Y}^{(j)}, j \in [d_1]$ are conditionally independent given $\mathbf{X}$. Therefore it suffices to bound III_1 and III_2. By applying Lemma A.2 with the term $\eta(d, n, \zeta_1, \zeta_2, \alpha_L)$ replaced by $\eta_0(d_1, d_2, n, \zeta_1, \zeta_2, \alpha_L)$ in Lemma A.8, (A.44) can be controlled by

$$\mathrm{III}_2 + \mathrm{III}_2 \le \frac{C' t_{\lambda_{d_2}}}{\epsilon^2 h_{d_2}} \left(\frac{d_1}{d_0|\mathcal{B}|} + \eta_0(d_1, d_2, n, \zeta_1, \zeta_2, \alpha_L) \right),$$

where $\alpha_L = q|\mathcal{B}|/d_1$ and $t_{\lambda_{d_2}}, h_{d_2}$ are similarly defined as in the proof of Theorem 5.2. According to Lemma A.8, we have the explicit form of $\eta_0(d_1, d_2, n, \zeta_1, \zeta_2, \delta, \alpha_L)$:

$$\eta_0(d_1, d_2, n, \zeta_1, \zeta_2, \delta, \alpha_L) = \zeta_1 \log d_2 + (\log d_2)^{5/2}\delta^{1/2} + \frac{\eta + \zeta_2}{\alpha_L},$$

where $\zeta_1 = O(s \log d_2/\sqrt{n})$, $\zeta_2 = O(e^{-c_1 n} + d_2^{-\widetilde{c}_0 \wedge c_2})$, δ satisfies $\frac{1}{\delta}\sqrt{\frac{s\log d_2}{n}} = O(1)$ and $\eta = e^{-c_1 n} + \frac{1}{d_2} + \frac{1}{n\delta^2}$. By rearranging, we obtain the following bound on $\mathrm{III}_2 + \mathrm{III}_2$:

$$\frac{\log d_2}{\epsilon^2} \left(\frac{1}{d_0 \rho} + \frac{s(\log d_2)^2}{n^{1/2}} + (\log d_2)^{5/2}\delta^{1/2} + \frac{1}{n\delta^2\rho} + \frac{1}{\rho}\left(\frac{1}{d_2} + e^{-c_1 n} + d_2^{-\widetilde{c}_0\wedge c_2}\right) \right).$$

where $\rho = \mathcal{B}/d_1$. We choose δ to be $\frac{1}{(n\rho)^{2/5}\log d_2}$ and have $\delta > \frac{1}{n^{2/5}\log d_2}$ (since $\rho < 1$). Thus this choice of δ satisfies the requirement in Lemma A.8. Finally we have (A.44) is bounded as

$$\frac{\log d_2}{\epsilon^2}\left(\frac{1}{d_0\rho} + \frac{s(\log d_2)^2}{n^{1/2}} + \frac{(\log d_2)^2}{(n\rho)^{1/5}} + \frac{1}{\rho d_2}\right).$$

Under the stated assumption in Theorem 4.2, the above term is $o(1)$. Thus the FDP control result is established. Due to similar derivations as in Theorem 5.2, the FDR control result follows. □

A.4 Ancillary lemmas for Theorem 4.2

To prove FDR control, we will establish a key result, i.e., Lemma A.8 in this section. Recall that in Section 4, we utilize the following result

$$\sqrt{n}(\widetilde{\mathbf{\Theta}}_j^{\mathrm{d}} - \mathbf{\Theta}_j) = Z_j + \Xi, \quad Z_j|\boldsymbol{X} \sim \mathcal{N}(0, \sigma_j^2 M\widehat{\mathbf{\Sigma}}M^\top).$$

and approximate the quantile of the maximum statistics $T_E = \max_{(j,k)\in E}\sqrt{n}|\widetilde{\mathbf{\Theta}}_{jk}^{\mathrm{d}}|$ by $T_E^{\mathcal{N}} = \max_{(j,k)\in E}|Z_{jk}|$. Lemma A.8 basically establishes the Cramér deviation bounds for such quantile approximation. Note that this lemma can be seen as a special case of Proposition C.1 since the involving random vector $\sqrt{n}(\widetilde{\mathbf{\Theta}}_j^{\mathrm{d}} - \mathbf{\Theta}_j)$ can be decomposed into a Gaussian random vector plus some error term. Hence we do not need to use the results in Kuchibhotla et al. (2021) to handle the case of a general random vector (and quantify Gaussian approximation errors).

In this section, we will define some notations similar to the theoretical results in Appendix C. First, we will drop the j-th subscript for simplicity. Without loss of generality, we prove relevant results for $E = \{(j,k) : k \in [d_2]\}$ and drop the subscript E. Note the results hold for any $j \in [d_1]$ and any subset of $\{(j,k) : k \in [d_2]\}$. Now we rewrite (4.4) using new notations, i.e.,

$$\sqrt{n}(\widetilde{\mathbf{\Theta}}_j^{\mathrm{d}} - \mathbf{\Theta}_j) = \boldsymbol{Z} + \Xi, \quad \boldsymbol{Z}|\boldsymbol{X} \sim \mathcal{N}(0, \sigma_j^2 M\widehat{\mathbf{\Sigma}}M^\top), \tag{A.45}$$

and denote its maximum by $T_{\boldsymbol{Z}} = ||\boldsymbol{Z}||_\infty$. Intuitively, we can use the quantile of $T_{\boldsymbol{Z}}$ to approximate the quantile of $T := \sqrt{n}||\widetilde{\mathbf{\Theta}}_j^{\mathrm{d}} - \mathbf{\Theta}_j||_\infty$. Since the covariance matrix $\sigma_j^2 M\widehat{\mathbf{\Sigma}}M^\top$ of the Gaussian random vector $\boldsymbol{Z}$ is not completely known, we can not directly compute its quantile (denoted by $q(\alpha;\boldsymbol{Z})$). Instead, we first estimate the unknown parameter σ_j by $\widehat{\sigma}_j$, which is constructed according to (4.5). Then we define $\boldsymbol{W} \sim \mathcal{N}(0, \widehat{\sigma}_j^2 M\widehat{\mathbf{\Sigma}}M^\top)$ (given the data $\boldsymbol{X}, \boldsymbol{Y}^{(j)}$), and denote its maximum by $T_{\boldsymbol{W}} = ||W||_\infty$. We will approximate the unknown quantile of T by the conditional quantile $q_\xi(\alpha;T_{\boldsymbol{W}})$. Here we use the ξ subscript to emphasize that we are conditioning on the data when defining such quantiles.

Due to the existence of the term Ξ in (A.45), there also exist additional estimation errors when we approximate the quantiles of T by the conditional quantiles $q_\xi(\alpha;T_{\boldsymbol{W}})$. Lemma A.7 characterizes such approximation errors. As for the difference between the distributions of the two Gaussian random vectors $\boldsymbol{W}$ and $\boldsymbol{Z}$, Lemma A.7 provides a bound on the maximal difference of their covariance matrices, which is denoted by Δ_∞. Finally, Lemma A.8 builds on these results and establishes the Cramér-type deviation bounds for the quantile approximation of T.

Lemma A.6. In the context of multiple linear models, we have

$$\mathbb{P}(|T - T_{\boldsymbol{Z}}| > \zeta_1) < \zeta_2,$$

where $\zeta_1 = O(s\log d_2/\sqrt{n})$ and $\zeta_2 = O(e^{-c_1 n} + d_2^{-\widetilde{c_0}\wedge c_2})$.

Proof of Lemma A.6. By Theorem 2.5 in Javanmard and Montanari (2014a), we have

$$\sqrt{n}(\widetilde{\boldsymbol{\Theta}}_j^{\mathrm{d}} - \boldsymbol{\Theta}_j) = \boldsymbol{Z} + \Xi, \quad \boldsymbol{Z}|\boldsymbol{X} \sim \mathcal{N}(0, \sigma_j^2 M \widehat{\boldsymbol{\Sigma}} M^\top),$$

and

$$\mathbb{P}\left(\|\Xi\|_\infty \geq \left(\frac{16ac\,\sigma}{C_{\min}} \right) \frac{s \log d_2}{\sqrt{n}} \right) \leq 4\, e^{-c_1 n} + 4\, d_2^{-\widetilde{c_0} \wedge c_2}.$$

Thus we immediately obtain the following bound on the difference between T and $T_{\boldsymbol{Z}}$:

$$\mathbb{P}(|T - T_{\boldsymbol{Z}}| > \zeta_1) < \zeta_2$$

where $\zeta_1 = O(s \log d_2 / \sqrt{n})$ and $\zeta_2 = O(e^{-c_1 n} + d_2^{-\widetilde{c_0} \wedge c_2})$. □

Lemma A.7. For the the maximal difference term $\Delta_\infty = ||\widehat{\sigma}^2 M \widehat{\boldsymbol{\Sigma}} M^\top - \sigma^2 M \widehat{\boldsymbol{\Sigma}} M^\top||_{\max}$, we have

$$\mathbb{P}\left(\Delta_\infty \geq \delta\right) \leq \eta, \tag{A.46}$$

where δ satisfies $\frac{1}{\delta}\sqrt{\frac{s \log d_2}{n}} = O(1)$ and $\eta = O\left(e^{-c_1 n} + \frac{1}{d_2} + \frac{1}{n\delta^2}\right)$.

Proof of Lemma A.7. To bound Δ_∞, we start with the term $|\widehat{\sigma}/\sigma - 1|$. First we denote

$$\mathcal{E}_n = \mathcal{E}_n(\phi_0, s_0, K) := \left\{ \boldsymbol{X} \in \mathbb{R}^{n \times d_1} : \min_{S:|S| \leq s_0} \phi(\widehat{\boldsymbol{\Sigma}}, S) \geq \phi_0, \max_{j \in [d_1]} \boldsymbol{\Sigma}_{jj} \leq K, \boldsymbol{\Sigma} = (\boldsymbol{X}^\top \boldsymbol{X})/n \right\}$$

similarly as in Theorem 7.(a) of Javanmard and Montanari (2014a), where $\phi(\widehat{\boldsymbol{\Sigma}}, S)$ is the compatibility constant as defined in Definition 1 of Javanmard and Montanari (2014a). Following the proof of Lemma 14 in Javanmard and Montanari (2014a), we have

$$\begin{aligned} \mathbb{P}\left(\left| \frac{\widehat{\sigma}}{\sigma} - 1 \right| \geq \epsilon \right) &\leq \mathbb{P}\left(\boldsymbol{X} \notin \mathcal{E}_n\right) + \sup_{\boldsymbol{X} \in \mathcal{E}_n} \mathbb{P}\Big(\left| \frac{\widehat{\sigma}}{\sigma} - 1 \right| \geq \epsilon \,\Big|\, \boldsymbol{X} \Big) \\ &\leq 4e^{-c_1 n} + \sup_{\boldsymbol{X} \in \mathcal{E}_n} \mathbb{P}\Big(\frac{||\boldsymbol{X}^\top \boldsymbol{E}||_\infty}{n\sigma^\star} \geq \widetilde{\lambda}/4 \,\Big|\, \boldsymbol{X} \Big) + \sup_{\boldsymbol{X} \in \mathcal{E}_n} \mathbb{P}\Big(\left| \frac{\sigma^\star}{\sigma} - 1 \right| \geq \frac{\epsilon}{10} \,\Big|\, \boldsymbol{X} \Big) \end{aligned} \tag{A.47}$$

where $\widetilde{\lambda} = 10\sqrt{(2 \log d_2)/n}$, $\sigma^\star$ is the oracle estimator of σ introduced in Sun and Zhang (2012) and ϵ satisfies $\frac{2\sqrt{s}\widetilde{\lambda}}{\sigma^\star \phi_0} \leq \frac{\epsilon}{2} < a_0$.

Now we separately bound the last two terms in (A.47). The second term in (A.47) can be bounded by the derivation in the proof of Theorem 2 (ii) (Sun and Zhang, 2012), i.e.,

$$\begin{aligned} \sup_{\boldsymbol{X} \in \mathcal{E}_n} \mathbb{P}\Big(\frac{||\boldsymbol{X}^\top \boldsymbol{E}||_\infty}{n\sigma^\star} \geq \widetilde{\lambda}/4 \,\Big|\, \boldsymbol{X} \Big) &\leq d_2 \mathbb{P}\Big(|L_k| \geq \sqrt{2 \log(d_2^{25/4})/n} \,\Big|\, \boldsymbol{X} \Big) \\ &\leq d_2 \cdot \frac{C}{d_2^{25/4} \sqrt{\log d_2}} \leq \frac{C}{d_2}, \end{aligned} \tag{A.48}$$

where L_k is the k-th element of $\frac{\boldsymbol{X}^\top \boldsymbol{E}}{n\sigma^\star}$ and $\frac{\sqrt{n-1}L_k}{\sqrt{1-L_k^2}}$ follows the Student's t-distribution with $n-1$ degrees of freedom. Then (A.48) holds due to equation (A7) in Sun and Zhang (2012) together with the union bound. As for the last term in (A.47), we note $n(\sigma^\star/\sigma)^2$ follows the χ_n^2 distribution according to Sun and Zhang (2012). Thus by Markov's inequality, we have

$$\sup_{\boldsymbol{X}\in\mathcal{E}_n} \mathbb{P}\Big(\Big|\frac{\sigma^\star}{\sigma}-1\Big| \geq \frac{\epsilon}{10} \,\Big|\, \boldsymbol{X}\Big) \;\leq\; \frac{C'\mathbb{E}\left[(n(\sigma^\star/\sigma)^2-n)^2\right]}{n^2\epsilon^2} \leq \frac{2C'}{n\epsilon^2}. \tag{A.49}$$

Now we arrive at the following bound on Δ_∞:

$$\begin{aligned} & \mathbb{P}\Big(\Delta_\infty \geq (\epsilon^2+2\epsilon)\cdot\sigma^2||M\widehat{\boldsymbol{\Sigma}}M^\top||_{\max}\Big) \\ = \;& \mathbb{P}\Big(||\widehat{\sigma}^2 M\widehat{\boldsymbol{\Sigma}}M^\top - \sigma^2 M\widehat{\boldsymbol{\Sigma}}M^\top||_{\max} \geq (\epsilon^2+2\epsilon)\cdot\sigma^2||M\widehat{\boldsymbol{\Sigma}}M^\top||_{\max}\Big) \\ \leq \;& \mathbb{P}\Big(\Big|\frac{\widehat{\sigma}}{\sigma}-1\Big| \geq \epsilon\Big) \\ \leq \;& 4e^{-c_1 n} + \frac{C}{d_2} + \frac{2C'}{n\epsilon^2}, \end{aligned}$$

where the last inequality comes from combining (A.47), (A.48) with (A.49). Note that the proof of Theorem 16 in Javanmard and Montanari (2014a) shows that $||M\widehat{\boldsymbol{\Sigma}}M^\top||_{\max} = O(1)$. Hence, we finally establish (A.46) with

$$\delta = \sigma^2(\epsilon^2+2\epsilon)||M\widehat{\boldsymbol{\Sigma}}M^\top||_{\max} = C_\sigma\epsilon, \quad \eta = O\left(e^{-c_1 n} + \frac{1}{d_2} + \frac{1}{n\delta^2}\right).$$

where C_σ is some constant and $\delta = C_\sigma\epsilon$ satisfies $\frac{1}{\delta}\sqrt{\frac{s\log d_2}{n}} = O(1)$ due to the choice of ϵ. □

Lemma A.8. Based on the result about the approximation error between T and $T_{\boldsymbol{Z}}$ (Lemma A.6) and the bound on $||\Delta||_\infty$ in Lemma A.7, we have

$$\sup_{\alpha\in[\alpha_L,1]} \left|\frac{\mathbb{P}(T>q(\alpha;T_{\boldsymbol{W}}))}{\mathbb{P}(T_{\boldsymbol{Z}}>q(\alpha;T_{\boldsymbol{Z}}))} - 1\right| = O\left(\eta_0(d_1,d_2,n,\zeta_1,\zeta_2,\delta,\alpha_L)\right), \tag{A.50}$$

where $\eta_0(d_1,d_2,n,\zeta_1,\zeta_2,\delta,\alpha_L) := \zeta_1\log d_2 + (\log d_2)^{5/2}\delta^{1/2} + \frac{\eta+\zeta_2}{\alpha_L}$ with $\zeta_1 = O(s\log d_2/\sqrt{n})$, $\zeta_2 = O(e^{-c_1 n} + d_2^{-\widetilde{c_0}\wedge c_2})$. Here δ is a term to be determined and we requite $\frac{1}{\delta}\sqrt{\frac{s\log d_2}{n}} = O(1)$. η depends on δ, i.e., $\eta = e^{-c_1 n} + \frac{1}{d_2} + \frac{1}{n\delta^2}$.

Proof of Lemma A.8. First we have $\mathbb{P}(|T - T_{\boldsymbol{Z}}| > \zeta_1) < \zeta_2$ by Lemma A.6, thus we obtain

$$\left|\frac{\mathbb{P}(T>q(\alpha;T_{\boldsymbol{W}}))}{\mathbb{P}(T_{\boldsymbol{Z}}>q(\alpha;T_{\boldsymbol{Z}}))} - 1\right| \leq \max\{\mathrm{II}_1,\mathrm{II}_2\} + \frac{2\zeta_2}{\alpha}$$

for $\alpha\in[\alpha_L,1]$, where II_1 and II_2 are defined as:

$$\mathrm{II}_1 := \left|\frac{\mathbb{P}(T_{\boldsymbol{Z}}>q(\alpha;T_{\boldsymbol{W}})+\zeta_1)}{\mathbb{P}(T_{\boldsymbol{Z}}>q(\alpha;T_{\boldsymbol{Z}}))} - 1\right|, \quad \mathrm{II}_2 := \left|\frac{\mathbb{P}(T_{\boldsymbol{Z}}>q(\alpha;T_{\boldsymbol{W}})-\zeta_1)}{\mathbb{P}(T_{\boldsymbol{Z}}>q(\alpha;T_{\boldsymbol{Z}}))} - 1\right|.$$

The above two terms can be bounded similarly. Take II_1 as an example, we use similar strategy as in Proposition C.1. Consider the event $S := \{\Delta_\infty \leq \delta\}$ where δ satisfies $\frac{1}{\delta}\sqrt{\frac{s\log d_2}{n}} = O(1)$, we apply Lemma C.1 and bound II_1 by

$$\frac{1}{1-\pi(\delta)}\cdot\mathrm{II}_{11} + \mathrm{II}_{12} + \frac{\mathbb{P}\left(\Delta_\infty>\delta\right)}{\alpha},$$

where II_{11} and II_{11} are defined as

$$\begin{aligned} \mathrm{II}_{11} &:= \frac{1-\pi(\delta)}{\alpha}\Big|\mathbb{P}\Big(T_{\boldsymbol{Z}} > q(\frac{\alpha}{1-\pi(\delta)};T_{\boldsymbol{Z}}) + \zeta_1\Big) - \mathbb{P}(T_{\boldsymbol{Z}} > q(\frac{\alpha}{1-\pi(\delta)};T_{\boldsymbol{Z}}))\Big|, \\ \mathrm{II}_{12} &:= \frac{1}{\alpha}\Big|\mathbb{P}\Big(T_{\boldsymbol{Z}} > q(\frac{\alpha}{1-\pi(\delta)};T_{\boldsymbol{Z}})\Big) - \mathbb{P}(T_{\boldsymbol{Z}} > q(\alpha;T_{\boldsymbol{Z}}))\Big|, \end{aligned} \tag{A.51}$$

where $\pi(\Delta_\infty) = [A(\Delta_\infty)+1]e^{M_1(\log d)^{3/2}A(\Delta_\infty)} - 1$. By applying the part 3 of Theorem 2.1 in Kuchibhotla et al. (2021) (with $r+\epsilon = q(\frac{\alpha}{1-\pi(\delta)};T_{\boldsymbol{Z}}) + \zeta_1, r-\epsilon = q(\frac{\alpha}{1-\pi(\delta)};T_{\boldsymbol{Z}})$) to the Gaussian random vector $\boldsymbol{Z}$, we have

$$\mathrm{II}_{11} \le K_4\zeta_1\big(q(\frac{\alpha}{1-\pi(\delta)};T_{\boldsymbol{Z}}) + \zeta_1/2\big) \le C\zeta_1\log d_2. \tag{A.52}$$

where the second inequality holds due to the similar reason stated in the proof of Proposition C.2. And the term II_{11} can be simply derived as

$$\mathrm{II}_{12} = \frac{1}{\alpha}\Big|\frac{\alpha}{1-\pi(\delta)} - \alpha\Big| = \frac{\pi(\delta)}{1-\pi(\delta)}. \tag{A.53}$$

Combing the results above, we have

$$\left|\frac{\mathbb{P}(T > q(\alpha;T_{\boldsymbol{W}}))}{\mathbb{P}(T_{\boldsymbol{Z}} > q(\alpha;T_{\boldsymbol{Z}}))} - 1\right| \le C'\zeta_1\log d_2 + \frac{\pi(\delta)}{1-\pi(\delta)} + \frac{\pi(\delta)}{1+\pi(\delta)} + \frac{2\mathbb{P}\left(\Delta_\infty > \delta\right)}{\alpha} + \frac{2\zeta_2}{\alpha}.$$

Applying the bound in Lemma A.7, we finally establish (A.50) i.e., $\eta_0(d_1,d_2,n,\zeta_1,\zeta_2,\alpha_L) := \zeta_1\log d_2 + (\log d_2)^{5/2}\delta^{1/2} + \frac{\eta+\zeta_2}{\alpha_L}$ up to some constant factor, where $\eta = e^{-c_1 n} + \frac{1}{d_2} + \frac{1}{n\delta^2}$. □

B Proofs of Cramér-type comparison bounds

In this section, we will prove two types of Cramér-type comparison bounds: Theorems 3.1 and 3.2. One of the challenges to derive the comparison bounds for Gaussian maxima is that the maximum function is non-smooth. In order to show the Cramér-type comparison bound, we first consider smooth approximation of the maximum. The following lemma from Bentkus (1990) show the existence of such smooth approximation.

Lemma B.1 (Theorem 1, Bentkus (1990))**.** Consider the Euclidean space $\mathbb{R}^d$ with ℓ_∞-norm, for any $t,\epsilon \ge 0$, there exists a smooth approximating function $\varphi_{r,\epsilon}$ satisfying the following:

(a) $\varphi_{r,\epsilon}: \mathbb{R}^d \to [0,1], \varphi_{r,\epsilon} \in \mathbb{C}^\infty$, where $\mathbb{C}^\infty$ is the smooth function class with functions differentiable for all degrees of differentiation.

(b) $\varphi_{r,\epsilon}(x) = 1$ if $||x||_\infty \le r$, $\varphi_{r,\epsilon}(x) = 0$ if $||x||_\infty \ge r+\epsilon$,

(c) $\sup_{x\in\mathbb{R}^d} ||D^j\varphi_{r,\epsilon}(x)||_1 \le c(j)\epsilon^{-j}\log^{j-1}(d+1)$,

where $||D^j\varphi_{r,\epsilon}(x)||_1 = \sum_{i_1=1}^d \cdots \sum_{i_j=1}^d \left|\frac{\partial^j\varphi_{r,\epsilon}(x)}{\partial x_{i_1}\cdots\partial x_{i_j}}\right|$ and the constants $c(j)$ only depends on j.

Remark B.1. *Kuchibhotla et al. (2021) gives a concrete example of $\varphi_{r,\epsilon}(x)$ satisfying the three properties in Lemma B.1:*

$$\varphi_{r,\epsilon}(x) = g_0\left(\frac{2(F_\beta(z_x - r\mathbf{1}_{2d}) - \epsilon/2)}{\epsilon}\right), \tag{B.1}$$

where $\beta = 2\log(2d)/\epsilon$, $g_0(t) := 30\mathbb{1}(0 \le t \le 1)\int_t^1 s^2(1-s)^2 ds + \mathbb{1}(t \le 0)$, $F_\beta(\cdot)$ *is the "softmax" function*

$$F_\beta(z) := \frac{1}{\beta}\log\Big(\sum_{m=1}^{2d}\exp\left(\beta z_m\right)\Big) \qquad \textit{for } z \in \mathbb{R}^{2d},$$

$z_x = (x^\top, -x^\top)^\top$, *and* $\mathbf{1}_{2d}$ *is the vector of* 1*'s of dimension* $2d$.

In fact, in the proof of Theorem 3.1, we do not need a specific form of $\varphi_{r,\epsilon}(x)$ *and any function satisfying Lemma B.1 will work. While in the proof of Theorem 3.2, we need to utilize the specific form in* (B.1).

B.1 Proof of Theorem 3.1

As mentioned in Remark 3.1, we can prove the Cramér-type comparison bound with max norm difference as $M_3(\log d)^{3/2}A(\Delta_\infty)e^{M_3(\log d)^{3/2}A(\Delta_\infty)}$, without the assumption on Δ_∞. Therefore we state the more general form of Theorem 3.1 below and give its proof. Note that under the assumption $(\log d)^5\Delta_\infty = O(1)$ and the discussions in Remark 3.1, the bound (3.1) in Theorem 3.1 immediately follows from Theorem B.2.

Theorem B.2 (CCB with max norm difference)**.** Let U and V be two Gaussian random vectors and we have

$$\sup_{0\le t\le C_0\sqrt{\log d}}\left|\frac{\mathbb{P}(||U||_\infty > t)}{\mathbb{P}(||V||_\infty > t)} - 1\right| \le M_3(\log d)^{3/2}A(\Delta_\infty)e^{M_1(\log d)^{3/2}A(\Delta_\infty)}, \tag{B.2}$$

where $C_0 > 0$ is some constant, $A(\Delta_\infty) = M_1\log d\Delta_\infty^{1/2}\exp\left(M_2\log^2 d\Delta_\infty^{1/2}\right)$, the constants M_1, M_2 only depend on $\min_{1\le j\le d}\{\sigma_{jj}^U, \sigma_{jj}^V\}, \max_{1\le j\le d}\{\sigma_{jj}^U, \sigma_{jj}^V\}$, and M_3 is a universal constant.

Proof of Theorem B.2. Using the smooth approximation in Lemma B.1, we can bound the difference between the distribution functions of Gaussian maxima as

$$\begin{aligned} & \left|\mathbb{P}(||U||_\infty > t) - \mathbb{P}(||V||_\infty > t)\right| \\ = \; & \left|\mathbb{E}[\mathbb{1}(||U||_\infty \le t) - \mathbb{1}(||V||_\infty \le t)]\right| \\ \le \; & \mathbb{P}(t-\epsilon \le ||V||_\infty \le t+\epsilon) + \max_{j=1,2}\left|\mathbb{E}\varphi_j(U) - \mathbb{E}\varphi_j(V)\right|, \end{aligned} \tag{B.3}$$

where $\varphi_1(x) := \varphi_{t,\epsilon}(x), \varphi_2(x) := \varphi_{t-\epsilon,\epsilon}(x)$. Regarding the inequality in (B.3), we first notice that

$$\mathbb{1}(||x||_\infty \le t) = \varphi_{t,\epsilon}(x) - \mathbb{1}(t < ||x||_\infty < t+\epsilon)\cdot\varphi_{t,\epsilon}(x) = \varphi_{t-\epsilon,\epsilon}(x) - \mathbb{1}(t-\epsilon < ||x||_\infty < t)\cdot\varphi_{t-\epsilon,\epsilon}(x),$$

where the first equality is due to property (b) in Lemma B.1. Hence we have

$$\begin{aligned} \mathbb{1}(||U||_\infty \le t) &\le \varphi_j(U), \quad j = 1,2 \\ \mathbb{1}(||V||_\infty \le t) &\ge \varphi_1(V) - \mathbb{1}(t < \|V\| < t+\epsilon), \\ \mathbb{1}(||V||_\infty \le t) &\ge \varphi_2(V) - \mathbb{1}(t-\epsilon < \|V\| < t), \end{aligned}$$

then (B.3) immediately follows by combining the above three inequalities.

The first term in (B.3) is related to the anti-concentration inequalities for the Gaussian maxima. By applying Theorem 2.1 in Kuchibhotla et al. (2021), we have

$$\mathbb{P}(t-\epsilon \le ||V||_\infty \le t+\epsilon) \le K_1(t+1)\epsilon\exp(K_2(t+1)\epsilon)\mathbb{P}(||V||_\infty > t). \tag{B.4}$$

The explicit forms of K_1, K_2 can be found in Theorem 2.1 of Kuchibhotla et al. (2021). They only depend on $\min_{1\le j\le d}\{\sigma^U_{jj}, \sigma^V_{jj}\}, \max_{1\le j\le d}\{\sigma^U_{jj}, \sigma^V_{jj}\}$ and the median of Gaussian maxima. Remark that the median of $||V||_\infty$ is bounded by $O(\sqrt{\log d})$ by the maximal inequalities for sub-Gaussian random variables (Lemma 5.2 in van Handel (2014)). Plugging this into the explicit form of K_1, K_2 in Theorem 2.1 of Kuchibhotla et al. (2021), we have $K_1 = O(\log d), K_2 = O(\log^2 d)$. Then (B.4) can be written as

$$\mathbb{P}(t - \epsilon \le ||V||_\infty \le t + \epsilon) \le M_1 \log d(t+1)\epsilon \exp(M_2 \log^2 d\ (t+1)\epsilon)\mathbb{P}(||V||_\infty > t),$$

for some constants M_1, M_2 only depending on $\min_{1\le j\le d}\{\sigma^U_{jj}, \sigma^V_{jj}\}, \max_{1\le j\le d}\{\sigma^U_{jj}, \sigma^V_{jj}\}$.

Overall the above bound has only a logarithmic dependence on the dimension d, similar to the anti-concentration bounds from Chernozhukov et al. (2014). But it quantifies the deviation with respect to the tail probability of the Gaussian maxima, thus offers a more refined characterization, which is crucial to our proof.

Now we deal with the second term in (B.3). It is not hard to check that the following proof works for both φ_1 and φ_2. Therefore, without loss of generality, we use a unified notation φ to represent either functions. We consider the Slepian interpolation between U and V: $W(s) := \sqrt{s}U + \sqrt{1-s}V, \quad s \in [0,1]$. Let $\Psi_t(s) = \mathbb{E}[\varphi(W(s))]$, then we have

$$|\mathbb{E}\varphi(U) - \mathbb{E}\varphi(V)| = |\Psi_t(1) - \Psi_t(0)| = \left|\int_0^1 \Psi_t'(s)ds\right|, \tag{B.5}$$

where $\Psi_t'(s) = \frac{1}{2}\sum_{j=1}^d \mathbb{E}[\partial_j\varphi(W(s))(s^{-1/2}U_j - (1-s)^{-1/2}V_j)]$. Applying Stein's identity (Lemma 2 of Chernozhukov et al. (2015)) to $(s^{-1/2}U_j - (1-s)^{-1/2}V_j, W(s)^\top)^\top$ and $\partial_j\varphi(W(s))$, we have

$$\Psi_t'(s) = \frac{1}{2}\sum_{j,k=1}^d (\sigma^U_{jk} - \sigma^V_{jk})\mathbb{E}[\partial_j\partial_k\varphi(W(s))]. \tag{B.6}$$

Hence we obtain the following bound on (B.5),

$$\begin{aligned}
\left|\int_0^1 \Psi_t'(s)ds\right| &\le \frac{1}{2}\sum_{j,k=1}^d |\sigma^U_{jk} - \sigma^V_{jk}| \cdot \left|\int_0^1 \mathbb{E}[\partial_j\partial_k\varphi(W(s))]ds\right| \\
&\le \frac{\Delta_\infty}{2}\int_0^1 \sum_{j,k=1}^d \mathbb{E}[|\partial_j\partial_k\varphi(W(s))|]ds \\
&\le \frac{\Delta_\infty}{2}\int_0^1 \sum_{j,k=1}^d \mathbb{E}[|\partial_j\partial_k\varphi(W(s))| \cdot \mathbb{1}(t-\epsilon \le ||W(s)||_\infty \le t+\epsilon)]ds \\
&\le \frac{\Delta_\infty}{2}\int_0^1 \sup_{x\in\mathbb{R}^d} ||D^2\varphi(x)||_1 \cdot \mathbb{E}[\mathbb{1}(t-\epsilon \le ||W(s)||_\infty \le t+\epsilon))]ds \\
&\le \frac{c(2)\Delta_\infty \log(d+1)}{2\epsilon^2}\int_0^1 \mathbb{P}(t-\epsilon \le ||W(s)||_\infty \le t+\epsilon)ds
\end{aligned} \tag{B.7}$$

where the second inequality is by the definition of Δ_∞ and the third one comes from the property (b) in Lemma B.1 for $\varphi_j(x), j = 1, 2$ (recalling $\varphi_1(x) = \varphi_{t,\epsilon}(x)$ and $\varphi_2(x) = \varphi_{t-\epsilon,\epsilon}(x)$). Note that property (c) gives a upper bound for the partial derivative terms. Thus the fourth inequality holds.

By the definition of Slepian interpolation, we have, for any $s \in [0,1]$, $W(s)$ is a Gaussian random vector and the variances can be controlled between $\min_{1\le j\le d}\{\sigma^U_{jj}, \sigma^V_{jj}\}$ and $\max_{1\le j\le d}\{\sigma^U_{jj}, \sigma^V_{jj}\}$.

The median of $||W(s)||_\infty$ can also be similarly bounded by $O(\sqrt{\log d})$ as $||V||_\infty$. Applying the anti-concentration inequalities again to $W(s)$ in (B.7), we thus obtain

$$\left|\int_0^1 \Psi_t'(s)ds\right| \leq \frac{c(2)\Delta_\infty \log(d+1)}{2\epsilon^2} \cdot M_1 \log d(t+1)\epsilon \exp(M_2 \log^2 d(t+1)\epsilon) \cdot \int_0^1 \mathbb{P}(||W(s)||_\infty > t)ds. \tag{B.8}$$

Let $Q_t(u) = \mathbb{P}(||W(u)||_\infty > t)$ and $R_t(u) = Q_t(u)/Q_t(0) - 1$. Combining (B.3), (B.4), (B.5) and (B.8), we have

$$\begin{aligned}
|Q_t(1) - Q_t(0)| &= \big|\mathbb{P}(||U||_\infty > t) - \mathbb{P}(||V||_\infty > t)\big| \\
&\leq M_1 \log d(t+1)\epsilon \exp(M_2 \log^2 d(t+1)\epsilon) Q_t(0) \\
&\quad + \frac{c(2)\Delta_\infty \log(d+1)}{2\epsilon^2} M_1 \log d(t+1)\epsilon \exp(M_2 \log^2 d(t+1)\epsilon) \int_0^1 Q_t(s)ds. \qquad (B.9)
\end{aligned}$$

If starting with the interpolation between $W(s)$ and V instead of that between U and V, we can similarly obtain the bound on $|Q_t(s) - Q_t(0)|$. And the integral $\int_0^1 Q_t(s)ds$ in (B.9) can be directly replaced by $\int_0^u Q_t(s)ds$. Namely, we have

$$\frac{|Q_t(u) - Q_t(0)|}{|Q_t(0)|} = |R_t(u)| \leq A(t,\epsilon)B(\Delta_\infty,\epsilon) \int_0^u |R_t(s)|ds + A(t,\epsilon)B(\Delta_\infty,\epsilon) \cdot u + A(t,\epsilon), \tag{B.10}$$

where we denote $A(t,\epsilon) = M_1 \log d(t+1)\epsilon \exp(M_2 \log^2 d(t+1)\epsilon)$ and $B(\Delta_\infty,\epsilon) = \frac{c(2)\Delta_\infty \log(d+1)}{2\epsilon^2}$.

Notice that (B.10) is an integral inequality and we can thus bound $R_t(s)$ by Grönwall's inequality (Grönwall, 1919)

$$|R_t(u)| \leq (A(t,\epsilon)B(\Delta_\infty,\epsilon)u + A(t,\epsilon))e^{A(t,\epsilon)B(\Delta_\infty,\epsilon)u}.$$

In particular, we have $|R_t(1)| \leq (A(t,\epsilon)B(\Delta_\infty,\epsilon) + A(t,\epsilon))e^{A(t,\epsilon)B(\Delta_\infty,\epsilon)}$. Remember that ϵ is the smoothing parameter that controls the level of approximation. Choosing $\epsilon = \Delta_\infty^{1/2}/(t+1)$, we then have $A(\Delta_\infty) := A(t,\epsilon) = M_1 \log d\Delta_\infty^{1/2} \exp(M_2 \log^2 d\Delta_\infty^{1/2})$ for some constants M_1, M_2 only depending on $\min_{1\leq j\leq d}\{\sigma_{jj}^U, \sigma_{jj}^V\}, \max_{1\leq j\leq d}\{\sigma_{jj}^U, \sigma_{jj}^V\}$ and $B(t) := B(\Delta_\infty,\epsilon) = \frac{c(2)\log(d+1)(t+1)^2}{2}$. When $0 \leq t \leq C_0\sqrt{\log d}$, we have $B(t) \leq M_1(\log d)^{3/2}$ for some universal constant M_3. Therefore the bound in (B.2) is established, i.e.,

$$\sup_{0\leq t\leq C_0\sqrt{\log d}} |R_t(1)| \leq M_3(\log d)^{3/2} A(\Delta_\infty) e^{M_3(\log d)^{3/2}A(\Delta_\infty)}.$$

□

B.2 Proof of Theorem 3.2

Before proving Theorem 3.2, we note its assumption about the connectivity can be relaxed. Therefore, we first present Theorem B.4 with a weaker connectivity assumption, which is stated below.

Assumption B.3 ($\mathfrak{p}$-connectivity property)**.** We say two Gaussian random vectors U and V satisfy the $\mathfrak{p}$-connectivity property if for any j such that $\sigma_{jk}^U \neq \sigma_{jk}^V$ for some k, there exists a subset $\mathcal{E}_0 \subset [d]$ satisfying the following three requirements:

(a) $j \in \mathcal{E}_0, |\mathcal{E}_0| = \mathfrak{p} + 1$;

(b) When $m, m' \in \mathcal{E}_0$ and $m \neq m'$, $\sigma_{mm}^U = \sigma_{m'm'}^U$ and $\sigma_{mm'}^U = \sigma_{mm'}^V = 0$ hold;

(c) $\forall\, k \in [d]$, $|\{m \in \mathcal{E}_0 : |\sigma^U_{km}| + |\sigma^V_{km}| \neq 0\}| \leq c_0$ for some constant c_0.

This assumption gives a characterization of the connectivity of the associated graphs of the Gaussian random vectors U and V. Below we give a few sufficient conditions (SC) for it.

SC1 U and V have unit variances. There exists a disjoint $(\mathfrak{p}+2)$-partition of nodes $\cup_{\ell=1}^{\mathfrak{p}+2}\mathcal{C}_\ell = [d]$ such that $\sigma^U_{jk} = \sigma^V_{jk} = 0$ when $j \in \mathcal{C}_\ell$ and $k \in \mathcal{C}_{\ell'}$ for some $\ell \neq \ell'$.

SC2 U and V have unit variances. There exist disjoint partitions of nodes $\cup_{\ell=1}^{\mathfrak{p}+2}\mathcal{C}^U_\ell = \cup_{\ell=1}^{\mathfrak{p}+2}\mathcal{C}^V_\ell = [d]$, such that σ^U_{jk} (σ^V_{jk}) equals 0 when j, k belong to different elements $\mathcal{C}^U_\ell$ ($\mathcal{C}^V_\ell$), and $\forall \ell \in [\mathfrak{p}+2]$, $\mathcal{C}^U_\ell \cap \mathcal{C}^V_\ell \neq \varnothing$.

SC3 $\forall\, s \in [0,1]$, the Gaussian random vector $W(s) := \sqrt{s}U + \sqrt{1-s}V$ always has the same variances σ_s^2 across different components. The associated graph of $W(s)$ has at least $\mathfrak{p}+2$ components, i.e., there exists a disjoint partition of nodes $\cup_{\ell=1}^{\mathfrak{p}+2}\mathcal{C}^W_\ell = [d]$, such that each $\mathcal{C}^W_\ell$ comes from a different component. And the partition $\cup_{\ell=1}^{\mathfrak{p}+2}\mathcal{C}^W_\ell = [d]$ works any $s \in [0,1]$.

Remark B.2. *Note that the above first condition SC1 is the main assumption of Theorem 3.2 (except that $\mathfrak{p}+2$ is replaced by $\mathfrak{p}$). It is immediate that the condition SC1 implies SC2. We will verify SC2 is indeed a sufficient condition of Assumption B.3 in the following paragraph. Regarding SC3, its sufficiency can be verified similarly, thus we omit the details.*

Simply, we have $\sigma^U_{jj} = \sigma^V_{jj} = 1, j \in [d]$ by the unit variance assumption. For any j such that $\sigma^U_{jk} \neq \sigma^V_{jk}$ for some k, we will construct a subset $\mathcal{E}_0$ and show it satisfies the three requirements (a), (b) and (c). Note that the condition SC1 assumes the existence of disjoint partitions of nodes $\cup_{\ell=1}^{\mathfrak{p}+2}\mathcal{C}^U_\ell = \cup_{\ell=1}^{\mathfrak{p}+2}\mathcal{C}^V_\ell = [d]$. We suppose $j \in \mathcal{C}^U_{\ell_1} \cap \mathcal{C}^V_{\ell_2}$ for some ℓ_1, ℓ_2, then $\mathcal{E}_0$ is constructed by including j and picking one element m_ℓ from $\mathcal{C}^U_\ell \cap \mathcal{C}^V_\ell$ for each $\ell \in [\mathfrak{p}+2] \setminus \{\ell_1, \ell_2\}$. As $\mathcal{C}^U_\ell \cap \mathcal{C}^V_\ell \neq \varnothing, \forall \ell \in [\mathfrak{p}+2]$, we have $|\mathcal{E}_0| \geq 1 + \mathfrak{p}$, hence the requirement (a) is satisfied. Regarding the requirement (b), when $m, m' \in \mathcal{E}_0, m \neq m'$, we immediately have $\sigma^U_{mm} = \sigma^V_{m'm'} = 1$ by the unit variance assumption. Since every element in $\mathcal{E}_0$ comes from a different component $\mathcal{C}^U_\ell$ ($\mathcal{C}^V_\ell$), we also have $\sigma^U_{mm'} = \sigma^V_{mm'} = 0$ when $m, m' \in \mathcal{E}_0, m \neq m'$. Lastly, due to the same reason, we have $\forall k \in [d]$, $|\{m \in \mathcal{E}_0 : |\sigma^U_{km}| + |\sigma^V_{km}| \neq 0\}| \leq 2$. Hence the requirement (c) is also satisfied.

Now we prove Theorem B.4, which is stated below. Note that it requires weaker connectivity assumption compared with Theorem 3.2 but needs to assume minimal eigenvalue conditions.

Theorem B.4 (CCB with elementwise ℓ_0 norm difference)**.** Consider the two Gaussian random vectors U and V to have equal variances $\sigma^U_{jj} = \sigma^V_{jj} = O(1)$, for $j \in [d]$ and we assume $\lambda_{\min}(\mathbf{\Sigma}^U) \geq 1/b_0 > 0, \lambda_{\min}(\mathbf{\Sigma}^V) \geq 1/b_0 > 0$ for some constant $b_0 > 0$. Suppose U and V also satisfy Assumption B.3, we then have

$$\sup_{0 \leq t \leq C_0\sqrt{\log d}} \left| \frac{\mathbb{P}(||U||_\infty > t)}{\mathbb{P}(||V||_\infty > t)} - 1 \right| = O\left(\frac{\Delta_0 \log d}{\mathfrak{p}} \right). \tag{B.11}$$

for some constant $C_0 > 0$.

Proof of Theorem B.4. Following the same derivations as in Theorem B.2, we have

$$\begin{aligned} & |\mathbb{P}(||U||_\infty > t) - \mathbb{P}(||V||_\infty > t)| \\ \leq\ & M_1 \log d(t+1)\epsilon \exp(M_2 \log^2 d(t+1)\epsilon)\mathbb{P}(||V||_\infty > t) + \max_{j=1,2} |\mathbb{E}[\varphi_j(U)] - \mathbb{E}[\varphi_j(V)]| \\ \leq\ & M_1 \log d(t+1)\epsilon \exp(M_2 \log^2 d(t+1)\epsilon)\mathbb{P}(||V||_\infty > t) + \left| \int_0^1 \Psi'_t(s) ds \right|, \end{aligned} \tag{B.12}$$

where $\epsilon = c/\max\{(\log d)^{3/2}, \mathfrak{p}\log d\}$ for some small enough constant $c > 0$, and the constants M_1, M_2 only depend on $\min_{1\le j\le d}\{\sigma_{jj}^U\}$, $\max_{1\le j\le d}\{\sigma_{jj}^U\}$. The above two inequalities hold by (B.3), (B.4) and (B.5). We further bound $\left|\int_0^1 \Psi_t'(s)ds\right|$ as below,

$$
\begin{aligned}
&\left|\int_0^1 \Psi_t'(s)ds\right| \\
\le\ & \frac{1}{2}\sum_{j,k=1}^{d} |\sigma_{jk}^U - \sigma_{jk}^V| \left|\int_0^1 \mathbb{E}[\partial_j\partial_k\varphi(W(s))]ds\right| \\
\le\ & \frac{M}{2}\sum_{j\neq k, \sigma_{jk}^U\neq\sigma_{jk}^V} \int_0^1 \mathbb{E}[|\partial_j\partial_k\varphi(W(s))|]ds \\
\le\ & \frac{M}{2}\sum_{j\neq k, \sigma_{jk}^U\neq\sigma_{jk}^V} \int_0^1 \mathbb{E}[|\partial_j\partial_k\varphi(W(s))|\cdot \mathbb{1}(t-\epsilon\le ||W(s)||_\infty \le t+\epsilon)]ds \\
\le\ & \frac{M\Delta_0}{2}\max_{j\neq k, \sigma_{jk}^U\neq\sigma_{jk}^V} \int_0^1 \mathbb{E}[|\partial_j\partial_k\varphi(W(s))|\cdot \mathbb{1}(t-\epsilon\le ||W(s)||_\infty \le t+\epsilon)]ds, \qquad \text{(B.13)}
\end{aligned}
$$

where the first inequality holds due to (B.6), the second inequality is because $\sigma_{jk}^U = O(1), \sigma_{jk}^V = O(1)$ for all j, k and the constant M only depends on the maximal variances of the elements of U, V, the third inequality holds by the property (b) in Lemma B.1 for $\varphi_j(x), j = 1, 2$, and the last inequality holds by the definition of Δ_0. Note that $\varphi_1(x) := \varphi_{t,\epsilon}(x), \varphi_2(x) := \varphi_{t-\epsilon,\epsilon}(x)$ as defined in the proof of Theorem B.2. We use the same strategy to deal with $\varphi_1(x)$ and $\varphi_2(x)$. Below we give the derivations when $\varphi = \varphi_1(x)$ and it is not hard to check these derivations work for $\varphi_2(x)$ as well. Recall the explicit construction of $\varphi : \mathbb{R}^d \to \mathbb{R}$ introduced in Remark B.1,

$$
\varphi(x) = \varphi_{r,\epsilon}(x) = g_0\left(\frac{2(F_\beta(z_x - r\mathbf{1}_{2d}) - \epsilon/2)}{\epsilon}\right),
$$

where $\beta = 2\log(2d)/\epsilon$, $g_0(t) := 30\mathbb{1}(0\le t\le 1)\int_t^1 s^2(1-s)^2ds + \mathbb{1}(t\le 0)$, F_β is the "softmax" function

$$
F_\beta(z) := \frac{1}{\beta}\log\Big(\sum_{m=1}^{2d}\exp(\beta z_m)\Big) \quad \text{for } z\in\mathbb{R}^{2d},
$$

$z_x = (x^\top, -x^\top)^\top$ and $\mathbf{1}_{2d}$ is the vector of 1's of dimension $2d$.

To bound (B.13), we consider the case where $j \neq k$ and $\sigma_{jk}^U \neq \sigma_{jk}^V$. Note that

$$
|\partial_j\partial_k\varphi(W(s))| \le ||g''||_\infty|\widetilde{\pi}_j(Z)\widetilde{\pi}_k(Z)| + \beta||g'||_\infty|\widetilde{\pi}_j(Z)\widetilde{\pi}_k(Z)|, \qquad \text{(B.14)}
$$

where $g(t) := g_0(\frac{2(t-\epsilon/2)}{\epsilon})$, $Z := W(s)$ and

$$
\widetilde{\pi}_j(z) := \frac{e^{\beta z_j} - e^{-\beta z_j}}{\sum_{m=1}^d e^{\beta z_m} + \sum_{m=1}^d e^{-\beta z_m}}.
$$

The above result follows from a direct calculation. Due to the boundedness of $||g_0'||_\infty, ||g_0''||_\infty$ and $\beta = 2\log(2d)/\epsilon$, we obtain the following bound on (B.14),

$$
\begin{aligned}
|\partial_j\partial_k\varphi(W(s))| &\le (||g''||_\infty + \beta||g'||_\infty)|\widetilde{\pi}_j(Z)\widetilde{\pi}_k(Z)| \\
&\le \left(\frac{4}{\epsilon^2}||g_0''||_\infty + \frac{2\beta}{\epsilon}||g_0'||_\infty\right)|\widetilde{\pi}_j(Z)\widetilde{\pi}_k(Z)| \\
&\le \frac{C_1\log(2d)}{\epsilon^2}|\widetilde{\pi}_j(Z)\widetilde{\pi}_k(Z)| \le \frac{C_1\log(2d)}{\epsilon^2}|\pi_j(Z)\pi_k(Z)|,
\end{aligned}
$$

for some constant C_1, where $\pi_j(z) = e^{\beta|z_j|}/\sum_{m=1}^d e^{\beta|z_m|}$. Recalling $Z = W(s)$, we have

$$
\begin{aligned}
&\int_0^1 \mathbb{E}\left[|\partial_j\partial_k\varphi(W(s))| \cdot \mathbb{1}(t-\epsilon \le ||W(s)||_\infty \le t+\epsilon)\right] ds \\
\le\ & \frac{C_1\log(2d)}{\epsilon^2}\int_0^1 \mathbb{E}\left[\pi_j(Z)\pi_k(Z)\cdot \mathbb{1}(t-\epsilon \le ||Z||_\infty \le t+\epsilon)\right] ds \\
=\ & \frac{C_1\log(2d)}{\epsilon^2}\mathbb{P}(||V||_\infty > t)\int_0^1 \underbrace{\frac{\mathbb{E}\left[\pi_j(Z)\pi_k(Z)\cdot \mathbb{1}(t-\epsilon \le ||Z||_\infty \le t+\epsilon)\right]}{\mathbb{P}(||V||_\infty > t)}}_{\text{II(s)}} ds. \qquad \text{(B.15)}
\end{aligned}
$$

Below we focus on bounding the term II(s) for any $s \in [0,1]$. First we rewrite $\pi_j(Z)\pi_k(Z)$ and simply derive the following inequality,

$$
\begin{aligned}
\pi_j(Z)\pi_k(Z) &= \frac{e^{\beta|Z_j|}}{\sum_{m=1}^d e^{\beta|Z_m|}} \cdot \frac{e^{\beta|Z_k|}}{\sum_{m=1}^d e^{\beta|Z_m|}} \\
&= \frac{e^{-\beta(||Z||_\infty - |Z_j|)}\cdot e^{-\beta(||Z||_\infty - |Z_k|)}}{(1+\sum_{|Z_m|\ne ||Z||_\infty} e^{-\beta(||Z||_\infty - |Z_m|)})^2} \\
&\le e^{-\beta(||Z||_\infty - |Z_j|)}\cdot e^{-\beta(||Z||_\infty - |Z_k|)}, \qquad \text{(B.16)}
\end{aligned}
$$

where the second equality comes from dividing both the numerator and denominator by $e^{2\beta||Z||_\infty}$ in the first line. Note that $\mathbb{P}(|Z_j| = |Z_k|) = 0$ since the random vector Z follows a non-degenerate d-dimensional multivariate Gaussian distribution. Hence we have

$$
1 = \mathbb{1}(|Z_j| = ||Z||_\infty, |Z_k| < ||Z||_\infty) + \mathbb{1}(|Z_j| < ||Z||_\infty), \text{ almost surely.} \qquad \text{(B.17)}
$$

Plugging the equality (B.17) into (B.16), we can further bound $\pi_j(Z)\pi_k(Z)$ as

$$
\pi_j(Z)\pi_k(Z) \le e^{-\beta(||Z||_\infty - |Z_k|)}\cdot \mathbb{1}(|Z_k| < ||Z||_\infty) + e^{-\beta(||Z||_\infty - |Z_j|)}\cdot \mathbb{1}(|Z_j| < ||Z||_\infty), \text{ almost surely.}
$$

Then we can bound II(s) by

$$
\begin{aligned}
\text{II}(s) &= \frac{\mathbb{E}\left[\pi_j(Z)\pi_k(Z)\cdot \mathbb{1}(t-\epsilon \le ||Z||_\infty \le t+\epsilon)\right]}{\mathbb{P}(||V||_\infty > t)} \\
&\le \frac{\mathbb{E}[e^{-\beta(||Z||_\infty - |Z_k|)}\cdot \mathbb{1}(|Z_k| < ||Z||_\infty)\mathbb{1}(t-\epsilon \le ||Z||_\infty \le t+\epsilon)]}{\mathbb{P}(||V||_\infty > t)} \qquad \text{(B.18)} \\
&\quad + \frac{\mathbb{E}[e^{-\beta(||Z||_\infty - |Z_j|)}\cdot \mathbb{1}(|Z_j| < ||Z||_\infty)\mathbb{1}(t-\epsilon \le ||Z||_\infty \le t+\epsilon)]}{\mathbb{P}(||V||_\infty > t)}. \qquad \text{(B.19)}
\end{aligned}
$$

We use the same strategy to bound (B.18) and (B.19). Below we give the derivations for bounding (B.19) and note these also work for (B.18).

For any $j \ne k$ and $\sigma_{jk}^U \ne \sigma_{jk}^V$, Assumption B.3 says that there exists a subset $\mathcal{E}_0 \subset [d]$ satisfying $j \in \mathcal{E}_0, |\mathcal{E}_0| = \mathfrak{p}+1$, and $\sigma_{mm}^U = \sigma_{m'm'}^U$, $\sigma_{mm'}^U = \sigma_{mm'}^V = 0$ when $m, m' \in \mathcal{E}_0, m \ne m'$. This implies the following: when $s = 0$ or 1 (i.e., $Z = U$ or V), we can find a $\mathfrak{p}-$dimensional random vector G such that (Z_j, G) are all independent and $\text{Var}(G_\ell) = \text{Var}(Z_j) = \sigma_j^2$ for $\ell \in [\mathfrak{p}]$. Note that G is constructed as $(Z_m)_{m\in\mathcal{E}_0, m\ne j}$ with $\mathcal{E}_0$ being the same for $Z = U$ and V. Therefore, for any $s \in (0,1), Z = W(s) = \sqrt{s}U + \sqrt{1-s}V$, we can construct $G = (Z_m)_{m\in\mathcal{E}_0, m\ne j}$ such that (Z_j, G) are all independent and $\text{Var}(G_\ell) = \text{Var}(Z_j) = \sigma_j^2$ for $\ell \in [\mathfrak{p}]$. Throughout the following proof and the lemmas in Appendix B.3, we will use the notation Z, G without making the dependence on s explicitly. And we denote the indices of the random variables in G (among Z) by $\mathcal{E}_G$, i.e., $\mathcal{E}_G = \mathcal{E}_0 \setminus \{j\} = \{m \in [d] : Z_m = G_\ell \text{ for some } \ell \in [\mathfrak{p}]\}$.

We will consider two separate cases based on whether $||G||_\infty = ||Z||_\infty$ holds. Formally, we write $\mathbb{1}(|Z_j| < ||Z||_\infty) \le \mathbb{1}(E_1) + \mathbb{1}(E_2)$ with E_1 and E_2 defined as

$$E_1 := \{||Z||_\infty > ||G||_\infty, ||Z||_\infty > |Z_j|\}, \tag{B.20}$$
$$E_2 := \{||G||_\infty = ||Z||_\infty > |Z_j|\}. \tag{B.21}$$

Then the numerator of the fraction in (B.19) can be bounded by the summation of the following two terms:

$$\begin{aligned} \mathrm{II}_1 &:= \mathbb{E}\left[e^{-\beta(||Z||_\infty - |Z_j|)} \cdot \mathbb{1}(E_1) \cdot \mathbb{1}(t - \epsilon \le ||Z||_\infty \le t + \epsilon)\right], \\ \mathrm{II}_2 &:= \mathbb{E}\left[e^{-\beta(||Z||_\infty - |Z_j|)} \cdot \mathbb{1}(E_2) \cdot \mathbb{1}(t - \epsilon \le ||Z||_\infty \le t + \epsilon)\right]. \end{aligned} \tag{B.22}$$

Combining (B.19) with (B.22) and applying Lemmas B.5 and B.6, we have

$$\mathrm{II}(s) \le \frac{2(\mathrm{II}_1 + \mathrm{II}_2)}{\mathbb{P}(||V||_\infty > t)} \le \frac{C'\epsilon \log d}{\beta \mathfrak{p}}, \quad \forall s \in [0,1], \tag{B.23}$$

for some constant C'. By (B.12), (B.13), (B.15) and (B.23), we thus obtain the following inequality

$$\begin{aligned} & |\mathbb{P}(||U||_\infty > t) - \mathbb{P}(||V||_\infty > t)| \\ \le\ & A(t,\epsilon)\mathbb{P}(||V||_\infty > t) + \frac{C_1 M \Delta_0 \log(2d)}{2\epsilon^2}\mathbb{P}(||V||_\infty > t) \cdot \frac{C'\epsilon \log d}{\beta \mathfrak{p}} \\ =\ & \mathbb{P}(||V||_\infty > t)\left(A(t,\epsilon) + B(\Delta_0, \mathfrak{p})\right), \end{aligned} \tag{B.24}$$

where $A(t,\epsilon) := M_1 \log d(t+1)\epsilon \exp(M_2 \log^2 d(t+1)\epsilon)$, $B(\Delta_0, \mathfrak{p}) := C''(\log d/\mathfrak{p})\Delta_0$ for some constants M_1, M_2, C''. In the last line, we also subsitute $\beta = \frac{2\log(2d)}{\epsilon}$. By re-arranging (B.24), we finally have

$$\left|\frac{\mathbb{P}(||U||_\infty > t)}{\mathbb{P}(||V||_\infty > t)} - 1\right| \le A(t,\epsilon) + B(\Delta_0, \mathfrak{p}).$$

Since $0 \le t \le C_0\sqrt{\log d}$ and $\epsilon = c/\max\{(\log d)^{3/2}, \mathfrak{p}\log d\}$ for some small enough constant $c > 0$, we have $A(t,\epsilon) = O(B(\Delta_0, \mathfrak{p}))$. Then (B.11) can be established, i.e.,

$$\sup_{0 \le t \le C_0\sqrt{\log d}} \left|\frac{\mathbb{P}(||U||_\infty > t)}{\mathbb{P}(||V||_\infty > t)} - 1\right| \le C''' B(\Delta_0, \mathfrak{p}) = O\left(\frac{\Delta_0 \log d}{\mathfrak{p}}\right).$$

□

Now we prove Theorem 3.2 using similar strategies as in Theorem B.4. Recall that the connectivity assumption in Theorem 3.2 assumes that there exists a disjoint $\mathfrak{p}$-partition of nodes $\cup_{\ell=1}^{\mathfrak{p}} \mathcal{C}_\ell = [d]$ such that $\sigma^U_{jk} = \sigma^V_{jk} = 0$ when $j \in \mathcal{C}_\ell$ and $k \in \mathcal{C}_{\ell'}$ for some $\ell \ne \ell'$. Since this connectivity assumption is stronger than that in Theorem B.4, we are able to do slightly more careful analysis in Lemma B.5. As a result, the minimal eigenvalue condition is no longer needed. Also note that Theorem 3.2 assumes the unit variance condition and there exists some $\sigma_0 < 1$ such that $|\sigma^V_{jk}| \le \sigma_0, |\sigma^U_{jk}| \le \sigma_0$ for any $j \ne k$. Both the variance condition and the covariance condition can be relaxed. In the following proof, we establish the Cramér-type comparison bound under a general variance condition. This general version is actually used in the proof of Theorem 5.2. Specifically, the general variance condition says that $a_0 \le \sigma^U_{jj} = \sigma^V_{jj} \le a_1,\ \forall j \in [d]$. After relaxing the unit variance assumption, some balanced variance assumption on the above components $\mathcal{C}_\ell$ is required. It says that given any $j \in \mathcal{C}_\ell$ with some ℓ, there exists at least one $m \in \mathcal{C}_{\ell'}$ such that $\sigma_{jj}^U = \sigma_{jj}^V = \sigma_{mm}^U = \sigma_{mm}^V$ for any $\ell' \neq \ell$. Remark this condition is mainly needed for Lemma B.12. We will call all these assumptions about variances as general variance condition. Denote $\widetilde{\sigma}_{jk}^U = \sigma_{jk}^U / \sqrt{\sigma_{jj}^U \sigma_{kk}^U}$. Accordingly, the covariance condition on σ_{jk} in Theorem 3.2 can also be relaxed into the following: there exists some $\sigma_0 < 1$ such that $|\widetilde{\sigma}_{jk}^V| = |\sigma_{jk}^V| / \sqrt{\sigma_{jj}^V \sigma_{kk}^V} \leq \sigma_0$ for any $j \neq k$ and $|\{(j,k) : j \neq k, |\widetilde{\sigma}_{jk}^U| = |\sigma_{jk}^U| / \sqrt{\sigma_{jj}^U \sigma_{kk}^U} > \sigma_0\}| \leq b_0$ for some constant b_0. We will call this condition as general covariance condition.

Proof of Theorem 3.2. Following exactly the same derivations in Theorem B.4 (up to (B.22)), we arrive at the following

$$\mathrm{II}(s) \leq \frac{2(\mathrm{II}_1 + \mathrm{II}_2)}{\mathbb{P}(||V||_\infty > t)},$$

where $\mathrm{II}(s), \mathrm{II}_1, \mathrm{II}_2$ are defined in (B.15) and (B.22), except that the random vector G can be constructed to satisfy more properties. Assuming the connectivity assumption of Theorem 3.2 and the general variance condition, we construct G by choosing one random variable Z_m from each component (except the one to which Z_j belongs) satisfying $\mathrm{Var}\,(Z_m) = \sigma_{mm}^U = \sigma_{mm}^V = \sigma_{jj}^U = \sigma_{jj}^V = \mathrm{Var}\,(Z_j)$. Such construction still satisfies the mentioned properties in Theorem B.4. Specifically, (G, Z_j) consists of $(\mathfrak{p}+1)$ i.i.d. Gaussian random variables. Moreover, for any $k \neq j, k \notin \mathcal{E}_G = \{m \in [d] : Z_m = G_\ell \text{ for some } \ell \in [\mathfrak{p}]\}$, there exists at most one $m \in \{j\} \cup \mathcal{E}_G$, such that Z_k and Z_m belong to the same component. Based on this property, we prove Lemma B.12 and Lemma B.13, which do not require minimal eigenvalue conditions compared with Lemma B.5 and Lemma B.7. We still apply Lemma B.13 to bound the term II_2. Regarding the term II_1, we control it by using Lemma B.6. Therefore, we obtain the following

$$\mathrm{II}(s) \leq \frac{C' \epsilon \log d}{\beta \mathfrak{p}} \left(1 + \frac{b_0}{\sqrt{1 - (s + (1-s)\sigma_0)^2}} \right). \tag{B.25}$$

Note a simple calculus result:

$$\int_0^1 \frac{b_0}{\sqrt{1 - (s + (1-s)\sigma_0)^2}} \leq \frac{0.5\pi b_0}{1 - \sigma_0} < C''$$

for some constant C'' when $\sigma_0 < 1$. Combining the above bound with (B.25), (B.12), (B.13), (B.15) and (B.23), we establish the bound (3.2) thus prove Theorem 3.2. □

B.3 Ancillary lemmas for Theorem B.4

Throughout the lemmas in this section, we will use Z and G without making the dependence on s explicitly, as mentioned in the proof of Theorem B.4.

Lemma B.5. Suppose $\lambda_{\min}(\mathbf{\Sigma}^U) \geq 1/b_0 > 0, \lambda_{\min}(\mathbf{\Sigma}^V) \geq 1/b_0 > 0$ for some constant $b_0 > 0$. For the term $\mathrm{II}_1 = \mathbb{E}\left[e^{-\beta(||Z||_\infty - |Z_j|)} \cdot \mathbb{1}(E_1) \cdot \mathbb{1}(t - \epsilon \leq ||Z||_\infty \leq t + \epsilon)\right]$ with E_1 defined in (B.20) and $\epsilon = c/\max\{(\log d)^{3/2}, \mathfrak{p} \log d\}$ for some small enough constant $c > 0$, whenever t satisfies $0 \leq t \leq C_0 \sqrt{\log d}$ for some constant $C_0 > 0$, we have

$$\frac{\mathrm{II}_1}{\mathbb{P}(||V||_\infty > t)} \leq \frac{C' \epsilon \log d}{\beta \mathfrak{p}}. \tag{B.26}$$

Proof of Lemma B.5. We will bound II_1 by the law of total expectation. Specifically, we first calculate the conditional expectation given (G, Z_j) then take expectation with respect to (G, Z_j). Denoting the conditional density function of $||Z||_\infty \mid Z_j = z_j, G = g$ by $f_{g,z_j}(u)$, we write out the integral form of II_1 as

$$\begin{aligned}
\mathrm{II}_1 = \; & \mathbb{E}\left[e^{-\beta(||Z||_\infty - |Z_j|)} \cdot \mathbb{1}(||Z||_\infty > ||G||_\infty, ||Z||_\infty > Z_j) \cdot \mathbb{1}(t - \epsilon \le ||Z||_\infty \le t + \epsilon)\right] \\
= \; & \mathbb{E}\left[e^{\beta|Z_j|} \cdot \mathbb{1}(||G||_\infty \le t + \epsilon, |Z_j| \le t + \epsilon) \left(\int_{t-\epsilon}^{t+\epsilon} f_{G,Z_j}(u) e^{-\beta u} \mathbb{1}(u > ||G||_\infty, u > |Z_j|) du\right)\right] \\
\le \; & \mathbb{E}\left[e^{\beta|Z_j|} \cdot \mathbb{1}(||G||_\infty \le t + \epsilon, |Z_j| \le t + \epsilon) \left(\int_{t-\epsilon}^{t+\epsilon} C\sqrt{\log d} \cdot e^{-\beta u} \mathbb{1}(u > |Z_j|) du\right)\right] \\
\le \; & C\sqrt{\log d}\, \mathbb{P}(||G||_\infty \le t + \epsilon) \mathbb{E}\left[\int_{|z_j| \le t+\epsilon} \phi\left(\frac{z_j}{\sigma_j}\right) e^{\beta|z_j|} \left(\int_{t-\epsilon}^{t+\epsilon} e^{-\beta u} \mathbb{1}(u > |z_j|) du\right) dz_j\right] \\
\le \; & C\sqrt{\log d}\, \mathbb{P}(||G||_\infty \le t + \epsilon) \underbrace{\int_{|z_j| \le t+\epsilon} \phi\left(\frac{z_j}{\sigma_j}\right) e^{\beta|z_j|} \left(\int_{t-\epsilon}^{t+\epsilon} e^{-\beta u} \mathbb{1}(u > |z_j|) du\right) dz_j}_{\mathrm{III}}, \qquad \text{(B.27)}
\end{aligned}$$

where the first inequality holds since $\mathbb{1}(u > ||G||_\infty, |Z_j|) \le \mathbb{1}(u > |Z_j|)$ and the conditional density function $f_{g,z_j}(u)$ is bounded by $C\sqrt{\log d}$ when $||g||_\infty, |z_j| < u \le t + \epsilon$ and $0 \le t \le C_0\sqrt{\log d}$, as a result of Lemma B.7. Recall that $\phi(\cdot)$ denotes the standard Gaussian PDF. We use the fact that $Z_j \perp\!\!\!\perp G$, $Z_j \sim \mathcal{N}(0, \sigma_j^2)$ and write out the integral form of the expectation with respect to Z_j, thus the second inequality follows. Then the integral III can be further rewritten as

$$\begin{aligned}
\mathrm{III} \;&=\; 2\int_{t-\epsilon}^{t+\epsilon} e^{-\beta u} \left(\int_0^u \phi\left(\frac{x}{\sigma_j}\right) e^{\beta x} dx\right) du \\
&=\; 2\int_{t-\epsilon}^{t+\epsilon} e^{-\beta u} \left(e^{\frac{\beta^2\sigma_j^2}{2}} \int_0^u \phi\left(\frac{x}{\sigma_j} - \beta\sigma_j\right) dx\right) du \\
&=\; 2\int_{t-\epsilon}^{t+\epsilon} e^{-\beta u} \left(e^{\frac{\beta^2\sigma_j^2}{2}} \int_{-\beta\sigma_j}^{u/\sigma_j - \beta\sigma_j} \phi(x)\, dx\right) du \\
&\le\; 2\int_{t-\epsilon}^{t+\epsilon} e^{-\beta u} \left(e^{\frac{\beta^2\sigma_j^2}{2}} \bar{\Phi}(\beta\sigma_j - u/\sigma_j)\right) du \\
&\le\; 2\int_{t-\epsilon}^{t+\epsilon} e^{-\beta u} \left(e^{\frac{\beta^2\sigma_j^2}{2}} \frac{e^{-\frac{(\beta\sigma_j - u/\sigma_j)^2}{2}}}{\beta\sigma_j - u/\sigma_j}\right) du \\
&\le\; \frac{4}{\beta\sigma_j} \int_{t-\epsilon}^{t+\epsilon} e^{-\beta u} \left(e^{\beta u} e^{-\frac{u}{2\sigma_j}}\right) du \le \frac{8\epsilon}{\beta\sigma_j} \exp\left(-\frac{(t-\epsilon)^2}{2\sigma_j^2}\right), \qquad \text{(B.28)}
\end{aligned}$$

where the first equality holds by Fubini's theorem, and the second equality holds by the definition of $\phi(\cdot)$. Regarding the first inequality, we use the fact that $u/\sigma_j - \beta\sigma_j < 2u/\sigma_j - \beta\sigma_j < 0$ for $u \le t + \epsilon$ and $t \le C_0\sqrt{\log d}$. This is because $\beta = \frac{2\log(2d)}{\epsilon}$ and $\epsilon = c/\max\{(\log d)^{3/2}, \mathfrak{p}\log d\}$ for some small enough constant $c > 0$. Then $\int_{-\beta\sigma_j}^{u/\sigma_j - \beta\sigma_j} \phi(x)\, dx \le \bar{\Phi}(\beta\sigma_j - u/\sigma_j)$, recalling $\bar{\Phi} = 1 - \Phi$, where Φ is the standard Gaussian CDF. The second inequality holds as a result of Lemma B.8. The third inequality holds due to $\beta\sigma_j > 2u/\sigma_j$ for $u \le t + \epsilon$.

By (B.27) and (B.28), we arrive at the following bound

$$\begin{aligned}\frac{\mathrm{II}_1}{\mathbb{P}(||V||_\infty > t)} &\le C\sqrt{\log d}\cdot \frac{\mathbb{P}(||G||_\infty \le t+\epsilon)}{\mathbb{P}(||V||_\infty > t)}\cdot \frac{8\epsilon}{\beta\sigma_j}\exp\left(-\frac{(t-\epsilon)^2}{2\sigma_j^2}\right)\\ &\le C\sqrt{\log d}\cdot\frac{C_1\epsilon}{\beta}\cdot\frac{\mathbb{P}(||G||_\infty \le t+\epsilon)}{\mathbb{P}(||G||_\infty > t)}\cdot\phi(\frac{t-\epsilon}{\sigma_j})/\sigma_j\\ &= C\sqrt{\log d}\cdot\frac{C_1\epsilon}{\beta}\cdot\underbrace{\frac{(1-2\bar{\Phi}(\frac{t+\epsilon}{\sigma_j}))^{\mathfrak{p}}}{1-(1-2\bar{\Phi}(\frac{t}{\sigma_j}))^{\mathfrak{p}}}\cdot\phi(\frac{t-\epsilon}{\sigma_j})/\sigma_j}_{\Lambda(t,\epsilon,\mathfrak{p})},\end{aligned}$$

for some constants C, C_1, where the second inequality holds due to the definition of $\phi(z)$ and $\mathbb{P}(||V||_\infty > t) \ge \mathbb{P}(||G||_\infty > t)$. This is because

$$\mathbb{P}(||V||_\infty > t) \ge \mathbb{P}(\max_{k\in\mathcal{E}_G}|V_k| > t) = \mathbb{P}(||G_V||_\infty > t) = \mathbb{P}(||G||_\infty > t), \tag{B.29}$$

where $G_V = (Z_m)_{m\in\mathcal{E}_0, m\neq j}$ with $Z = V$ has the same distribution as G. Regarding the last line, by the construction of $G = (G_\ell)_{\ell\in[\mathfrak{p}]} = (Z_m)_{m\in\mathcal{E}_0,m\neq j}$ in the proof of Theorem B.4, we have $\{G_\ell\}_{\ell\in[\mathfrak{p}]}$ are $\mathfrak{p}$ i.i.d. Gaussian random variables with $\mathrm{Var}(G_\ell) = \mathrm{Var}(Z_j) = \sigma_j^2$. By applying Lemma B.9 to the term $\Lambda(t,\epsilon,\mathfrak{p})$ in the last line, we further obtain,

$$\frac{\mathrm{II}_1}{\mathbb{P}(||V||_\infty > t)} \le C'\sqrt{\log d}\cdot\frac{\epsilon}{\beta}\frac{\sqrt{\log d}}{\mathfrak{p}} = \frac{C'\epsilon\log d}{\beta\mathfrak{p}},$$

for some constant C', therefore (B.26) is established. □

Lemma B.6. For the term $\mathrm{II}_2 = \mathbb{E}\left[e^{-\beta(||Z||_\infty - |Z_j|)}\cdot\mathbb{1}(E_2)\cdot\mathbb{1}(t-\epsilon\le ||Z||_\infty \le t+\epsilon)\right]$ with E_2 defined in (B.21) and $\epsilon = c/\max\{(\log d)^{3/2}, \mathfrak{p}\log d\}$ for some small enough constant $c > 0$, whenever t satisfies $0 \le t \le C_0\sqrt{\log d}$ for some constant $C_0 > 0$, we have

$$\frac{\mathrm{II}_2}{\mathbb{P}(||V||_\infty > t)} \le \frac{C''\epsilon\sqrt{\log d}}{\beta\mathfrak{p}}. \tag{B.30}$$

Proof of Lemma B.6. By the definition of E_2 in (B.21) and the tower property, we have

$$\begin{aligned}\mathrm{II}_2 &= \mathbb{E}\left[e^{-\beta(||Z||_\infty-|Z_j|)}\cdot\mathbb{1}(||G||_\infty = ||Z||_\infty > |Z_j|)\cdot\mathbb{1}(t-\epsilon\le||Z||_\infty\le t+\epsilon)\right]\\ &= \mathbb{E}\left[e^{-\beta(||G||_\infty-|Z_j|)}\cdot\mathbb{1}(||G||_\infty = ||Z||_\infty, ||G||_\infty > |Z_j|)\cdot\mathbb{1}(t-\epsilon\le||G||_\infty\le t+\epsilon)\right]\\ &\le \mathbb{E}\left[e^{-\beta(||G||_\infty-|Z_j|)}\cdot\mathbb{1}(||G||_\infty > |Z_j|)\cdot\mathbb{1}(t-\epsilon\le||G||_\infty\le t+\epsilon)\right]\\ &= \mathbb{E}\left[\mathbb{E}\left[e^{\beta|Z_j|}\mathbb{1}(|Z_j| < ||G||_\infty)\,|\,G\right]e^{-\beta||G||_\infty}\cdot\mathbb{1}(t-\epsilon\le||G||_\infty\le t+\epsilon)\right].\end{aligned} \tag{B.31}$$

First we bound $\mathrm{III}(g) := \mathbb{E}\left[e^{\beta|Z_j|}\mathbb{1}(|Z_j| < ||G||_\infty) \,|\, G = g\right]$ when $||g||_\infty \in [t-\epsilon, t+\epsilon]$. Specifically,

$$
\begin{aligned}
\mathrm{III}(g) = \frac{2}{\sigma_j}\int_0^{||g||_\infty} e^{\beta x}\phi\Big(\frac{x}{\sigma_j}\Big)dx \quad &= \quad \frac{2e^{\beta^2\sigma_j^2/2}}{\sigma_j}\int_0^{||g||_\infty}\phi\Big(\frac{x-\beta\sigma_j^2}{\sigma_j}\Big)dx \\
&\le \quad 2e^{\beta^2\sigma_j^2/2}\int_{-\infty}^{||g||_\infty/\sigma_j - \beta\sigma_j}\phi(y)dy \\
&= \quad 2e^{\beta^2\sigma_j^2/2}\,\bar{\Phi}(\beta\sigma_j - ||g||_\infty/\sigma_j) \\
&\le \quad 2e^{\beta^2\sigma_j^2/2}\frac{\phi\left(\beta\sigma_j - ||g||_\infty/\sigma_j\right)}{\beta\sigma_j - ||g||_\infty/\sigma_j} \\
&\le \quad \frac{4}{\beta\sigma_j}\phi\left(\frac{||g||_\infty}{\sigma_j}\right)e^{\beta||g||_\infty}, \qquad \text{(B.32)}
\end{aligned}
$$

where the first equality holds due to $Z_j \perp\!\!\!\perp G$, and the second equality comes from rearranging. The first inequality holds by the change of variable $y = (x - \beta\sigma_j^2)/\sigma_j$ and setting the lower limit of the integral as $-\infty$. Because $\beta = \frac{2\log(2d)}{\epsilon}$ and $\epsilon = c/\max\{(\log d)^{3/2}, \mathfrak{p}\log d\}$ for some small enough constant $c > 0$, we have $||g||_\infty/\sigma_j < \beta\sigma_j$ for $||g||_\infty \le t+\epsilon$ and $t \le C_0\sqrt{\log d}$. Then the second inequality holds as a result of Lemma B.8 and the fact that $\beta\sigma_j - ||g||_\infty/\sigma_j > 0$. The last inequality comes from rearranging and the fact that $\beta\sigma_j > 2||g||_\infty/\sigma_j$ for $||g||_\infty \le t+\epsilon$ and $t \le C_0\sqrt{\log d}$. Combining (B.32) with (B.31), we have

$$
\begin{aligned}
\mathrm{II}_2 \quad &\le \quad \mathbb{E}\Big[\mathrm{III}(G)\cdot e^{-\beta||G||_\infty}\cdot\mathbb{1}(t-\epsilon \le ||G||_\infty \le t+\epsilon)\Big] \\
&\le \quad \frac{4}{\beta\sigma_j}\mathbb{E}\left[\phi\Big(\frac{||G||_\infty}{\sigma_j}\Big)e^{\beta(||G||_\infty)}\cdot e^{-\beta||G||_\infty}\cdot\mathbb{1}(t-\epsilon \le ||G||_\infty \le t+\epsilon)\right] \\
&\le \quad \frac{4}{\beta\sigma_j}\int_{t-\epsilon}^{t+\epsilon}\phi\Big(\frac{y}{\sigma_j}\Big)f(y)dy, \qquad \text{(B.33)}
\end{aligned}
$$

where $f(y)$ denotes the PDF of $||G||_\infty$. As $\{G_\ell\}_{\ell\in[\mathfrak{p}]}$ are i.i.d. Gaussian random variables satisfying $\forall\, \ell \in [\mathfrak{p}]$, $\mathbb{E}[G_\ell] = 0$ and $\mathrm{Var}(G_\ell) = \sigma_j^2$, we have for $y > 0$,

$$
\mathbb{P}(||G||_\infty \le y) = \mathbb{P}\Big(\bigcup_{\ell\in[\mathfrak{p}]}|G_\ell| \le y\Big) = (1 - 2\mathbb{P}(G_\ell/\sigma_j > y/\sigma_j))^{\mathfrak{p}} = (1 - 2\bar{\Phi}(y/\sigma_j))^{\mathfrak{p}}. \qquad \text{(B.34)}
$$

Thus we have the PDF of $||G||_\infty$ equals $f(y) = \frac{2\mathfrak{p}}{\sigma_j}\left(1 - 2\bar{\Phi}(\frac{y}{\sigma_j}))^{\mathfrak{p}}\right)^{\frac{\mathfrak{p}-1}{\mathfrak{p}}}\phi(\frac{y}{\sigma_j})$. Plugging the expression of $f(y)$ into (B.33), we further derive the following bound

$$
\begin{aligned}
\frac{\mathrm{II}_2}{\mathbb{P}(||V||_\infty > t)} \quad &\le \quad \frac{8\mathfrak{p}}{\beta\sigma_j^2}\int_{t-\epsilon}^{t+\epsilon}\frac{\left(1 - 2\bar{\Phi}(\frac{y}{\sigma_j}))^{\mathfrak{p}}\right)^{\frac{\mathfrak{p}-1}{\mathfrak{p}}}\phi^2(\frac{y}{\sigma_j})}{\mathbb{P}(||V||_\infty > t)}dy \\
&\le \quad \frac{8\mathfrak{p}}{\beta\sigma_j^2}\int_{t-\epsilon}^{t+\epsilon}\frac{\left(1 - 2\bar{\Phi}(\frac{y}{\sigma_j}))^{\mathfrak{p}}\right)^{\frac{\mathfrak{p}-1}{\mathfrak{p}}}\phi^2(\frac{y}{\sigma_j})}{1 - \mathbb{P}(||G||_\infty \le t)}dy \\
&= \quad \frac{8\mathfrak{p}}{\beta\sigma_j^2}\int_{t-\epsilon}^{t+\epsilon}\frac{\left((1 - 2\bar{\Phi}(\frac{y}{\sigma_j}))^{\mathfrak{p}}\right)^{\frac{\mathfrak{p}-1}{\mathfrak{p}}}\phi^2(\frac{y}{\sigma_j})}{1 - (1 - 2\bar{\Phi}(\frac{t}{\sigma_j}))^{\mathfrak{p}}}dy \\
&\le \quad \frac{16\epsilon}{\beta\sigma_j^2\mathfrak{p}}\frac{\left((1 - 2\bar{\Phi}(\frac{t+\epsilon}{\sigma_j}))^{\mathfrak{p}}\right)^{\frac{\mathfrak{p}-1}{\mathfrak{p}}}(\mathfrak{p}\phi(\frac{t-\epsilon}{\sigma_j}))^2}{1 - (1 - 2\bar{\Phi}(\frac{t+\epsilon}{\sigma_j}))^{\mathfrak{p}}} \le \frac{C''\epsilon\sqrt{\log d}}{\beta\mathfrak{p}},
\end{aligned}
$$

for some constant C', where the second inequality holds due to (B.29), as mentioned in the proof of Lemma B.5. The equality holds as a result of substituting the expression of $\mathbb{P}(||G||_\infty \le t)$ by (B.34). The third inequality holds since $1 - 2\bar{\Phi}(z)$ is monotonically increasing and $\phi(z)$ is monotonically decreasing when $z \ge 0$. As for the last line, we apply Lemma B.10. Finally, (B.30) is established. □

Lemma B.7. Suppose $\lambda_{\min}(\mathbf{\Sigma}^U) \ge 1/b_0 > 0, \lambda_{\min}(\mathbf{\Sigma}^V) \ge 1/b_0 > 0$ for some constant $b_0 > 0$. Recall that the density function of the conditional distribution of $||Z||_\infty \mid \{Z_j = z_j, G = g\}$ is denoted by $f_{g,z_j}(z)$. Suppose $\epsilon > 0$, when $0 \le t \le C_0\sqrt{\log d}$ for some constant $C_0 > 0$ and $|z_j|, ||g||_\infty \le t + \epsilon$, we have

$$f_{g,z_j}(z) \le C\sqrt{\log d}, \quad \forall\ z \in (\max\{|z_j|, ||g||_\infty\}, t + \epsilon].$$

Proof of Lemma B.7. First we introduce some new notations. Let $(\sigma_{jk})_{1\le j,k\le d} \in \mathbb{R}^{d\times d}$ be the covariance matrix of Z. For given j, we denote

$$\sigma_{kk\cdot j} := \sigma_{kk} - \sigma_{kj}^2\sigma_{jj}^{-1} - \sum_{m\in\mathcal{E}_G} \sigma_{km}^2\sigma_{mm}^{-1}. \tag{B.35}$$

As $z \in (\max\{|z_j|, ||g||_\infty\}, t+\epsilon]$, we can choose δ such that $0 < \delta < z - \max\{|z_j|, ||g||_\infty\}$. Throughout the following proof, we will work with such δ. Since $\max\{|z_j|, ||g||_\infty\} - z < -\delta$, we have

$$\mathbb{P}\left(\big|||Z||_\infty - z\big| \le \delta \,|\, Z_j = z_j, G = g\right) = \mathbb{P}\left(\big|||X||_\infty - z\big| \le \delta \,|\, Z_j = z_j, G = g\right), \tag{B.36}$$

where X denotes the $(d - \mathfrak{p} - 1)$-dimensional random vector by excluding Z_j, G from Z and therefore $||Z||_\infty = \max\{||X||_\infty, |Z_j|, ||G||_\infty\}$.

Recalling $G = (G_\ell)_{\ell\in[\mathfrak{p}]} = (Z_m)_{m\in\mathcal{E}_G}$, where $\mathcal{E}_G$ denotes the indices of the random variables in G (among Z), i.e., $\mathcal{E}_G = \{m \in [d] : Z_m = G_\ell$ for some $\ell \in [\mathfrak{p}]\}$, we have

$$||X||_\infty = \max_{k\in[d],\ k\notin\{j,\mathcal{E}_G\}} \{\max\{Z_k, -Z_k\}\}.$$

Given j and the choice of G, we also denote

$$\underline{\sigma_{\cdot j}} := \min_{k\in[d],k\notin\{j,\mathcal{E}_G\}} \sqrt{\sigma_{kk\cdot j}}, \quad \bar{\rho}_j := \max_{k\in[d],k\notin\{j,\mathcal{E}_G\}} \frac{|\sigma_{jk}|}{\sigma_{jj}}. \tag{B.37}$$

For each $k \in [d], k \notin \{j, \mathcal{E}_G\}$, the conditional expectation $\mathbb{E}[Z_k \,|\, Z_j, G]$ has the following expression,

$$\mathbb{E}[Z_k \,|\, Z_j, G] = \sigma_{kj}\sigma_{jj}^{-1}Z_j + \sum_{m\in\mathcal{E}_G} (\sigma_{km}\sigma_{mm}^{-1}Z_m), \tag{B.38}$$

since (Z_j, G) are all independent. Note that the requirement (c) in Assumption B.3 says $\forall\ k \in [d]$, $|\{m \in \mathcal{E}_0 : |\sigma_{km}^U| + |\sigma_{km}^V| \ne 0\}| \le c_0$, we thus have

$$\begin{aligned} \sum_{m\in\mathcal{E}_G} \mathbb{1}(\sigma_{km} \ne 0) &= \sum_{m\in\mathcal{E}_G} \mathbb{1}((s\sigma_{km}^U + (1-s)\sigma_{km}^V) \ne 0) \\ &\le \sum_{m\in\mathcal{E}_G} \mathbb{1}(\sigma_{km}^U \ne 0 \text{ or } \sigma_{km}^V \ne 0) \le c_0, \end{aligned} \tag{B.39}$$

where the first equality holds by the definition of σ_{km} and $Z = W(s) = \sqrt{s}U + \sqrt{1-s}V$. Combining (B.39) with (B.38), it yields the following bound on $|\mathbb{E}[Z_k \mid Z_j = z_j, G = g]|$,

$$\begin{aligned} |\mathbb{E}[Z_k \mid Z_j = z_j, G = g]| &= \left|\sigma_{kj}\sigma_{jj}^{-1}z_j + \sum_{m\in\mathcal{E}_G}(\sigma_{km}\sigma_{mm}^{-1}z_m)\right| \\ &\le \bar{\rho}_j(|z_j| + c_0||g||_\infty), \end{aligned} \tag{B.40}$$

where $\bar{\rho}_j = \max_{k\in\mathcal{E}_X}\frac{|\sigma_{jk}|}{\sigma_{jj}}$ as defined. Denoting $\mathcal{E}_X := \{k : k \in [d],\ k \notin \{j, \mathcal{E}_G\}\}$, we define the following random variables,

$$\widetilde{W}_{2k-1} = \frac{Z_k - z}{\sqrt{\sigma_{kk\cdot j}}} + \frac{\widetilde{z}}{\underline{\sigma_{\cdot j}}}, \quad \widetilde{W}_{2k} = \frac{-Z_k - z}{\sqrt{\sigma_{kk\cdot j}}} + \frac{\widetilde{z}}{\underline{\sigma_{\cdot j}}}, \quad k \in \mathcal{E}_X, \tag{B.41}$$

where $\widetilde{z} = z + \bar{\rho}_j(|z_j| + c_0||g||_\infty)$. Then by the definitions of $\sigma_{kk\cdot j}, \underline{\sigma_{\cdot j}}$ and $\bar{\rho}_j$ in (B.35) and (B.37), we have the above random variables satisfy the following properties,

$$\mathbb{E}\left[\widetilde{W}_m \mid Z_j = z_j, G = g\right] \ge 0, \quad \mathrm{Var}\left(\widetilde{W}_m \mid Z_j = z_j, G = g\right) = 1,$$

where $m = 2k-1$ or $2k$ and $k \in \mathcal{E}_X$. Denote those random variables defined in (B.41) by $\{\widetilde{W}_m\}$ for notation simplicity. We let $q_{z_j,g}(w)$ be the PDF of the conditional distribution of $\max_m\{\widetilde{W}_m\} \mid Z_j = z_j, G = g$. Then we will apply the derivation of Step 2 in Theorem 3 of Chernozhukov et al. (2015) to bound $q_{z_j,g}(w)$. Note that for the following derivations, we always conditional on the event $Z_j = z_j, G = g$. First, we verify the condition on $\{\widetilde{W}_m\}$. Since $|\mathrm{Corr}(U_j, U_k)| \ne 1$, $|\mathrm{Corr}(V_j, V_k)| \ne 1$ for distinct $j, k \in [d]$, we then have the correlation between $\widetilde{W}_{m_1}$ and $\widetilde{W}_{m_2}$ for $m_1 \ne m_2$ is less than 1. Therefore, by applying the derivation of Step 2 in Theorem 3 of Chernozhukov et al. (2015) to $\{\widetilde{W}_m\}$, we have

$$q_{z_j,g}(w) \le h(w) := 2(w \vee 1)\exp\left\{-\frac{(w - \bar{w} - a_d)_+^2}{2}\right\},$$

where $\bar{w} = \max_m \mathbb{E}\left[\widetilde{W}_m \mid Z_j = z_j, G = g\right]$ and

$$a_d = \max_m \mathbb{E}\left[\left(\widetilde{W}_m - \mathbb{E}\left[\widetilde{W}_m \mid Z_j = z_j, G = g\right]\right) \mid Z_j = z_j, G = g\right].$$

When $w \le \bar{w} + a_d$, we have $h(w) \le 2(\bar{w} + a_d)$. To deal with the case where $w > \bar{w} + a_d$, we consider

$$\begin{aligned} \log(h(w)) &= \log(2w) - \frac{(w - \bar{w} - a_d)^2}{2}, \\ \frac{d\log(h(w))}{dw} &= \frac{1}{w} - (w - \bar{w} - a_d), \\ \frac{d^2\log(h(w))}{dw^2} &= -\frac{1}{w^2} - 1 < 0. \end{aligned}$$

Solving $\frac{d}{dw}\log(h(w)) = 0$ yields $w^\star = \frac{\bar{w} + a_d + \sqrt{(\bar{w} + a_d)^2 + 4}}{2}$. Therefore, the PDF of the conditional distribution of $\max_m\{\widetilde{W}_m\} \mid Z_j = z_j, G = g$ can be bounded by

$$h(w) \le h(w^\star) \le 3(\bar{w} + a_d). \tag{B.42}$$

Now we have

$$\begin{aligned}
& \mathbb{P}\left(\big|||Z||_\infty - z\big| \le \delta \,|\, Z_j = z_j, G = g\right) \\
= \;& \mathbb{P}\left(\big|||X||_\infty - z\big| \le \delta \,|\, Z_j = z_j, G = g\right) \\
= \;& \mathbb{P}\left(\left|\max_{k\in\mathcal{E}_X}\{Z_k, -Z_k\} - z\right| \le \delta \,|\, Z_j = z_j, G = g\right) \\
\le \;& \mathbb{P}\left(\left|\max_{k\in\mathcal{E}_X}\left\{\frac{Z_k - z}{\sqrt{\sigma_{kk\cdot j}}}, \frac{-Z_k - z}{\sqrt{\sigma_{kk\cdot j}}}\right\}\right| \le \frac{\delta}{\underline{\sigma_{\cdot j}}} \,|\, Z_j = z_j, G = g\right) \\
\le \;& \sup_{y\in\mathbb{R}} \mathbb{P}\left(\left|\max_{k\in\mathcal{E}_X}\left\{\frac{Z_k - z}{\sqrt{\sigma_{kk\cdot j}}} + \frac{\widetilde{z}}{\underline{\sigma_{\cdot j}}}, \frac{-Z_k - z}{\sqrt{\sigma_{kk\cdot j}}} + \frac{\widetilde{z}}{\underline{\sigma_{\cdot j}}}\right\} - y\right| \le \frac{\delta}{\underline{\sigma_{\cdot j}}} \,|\, Z_j = z_j, G = g\right) \\
= \;& \sup_{y\in\mathbb{R}} \mathbb{P}\left(\left|\max_m\{\widetilde{W}_m\} - y\right| \le \frac{\delta}{\underline{\sigma_{\cdot j}}} \,|\, Z_j = z_j, G = g\right) \le \frac{6\delta}{\underline{\sigma_{\cdot j}}}(\bar{w} + a_d), \qquad \text{(B.43)}
\end{aligned}$$

where the first equality holds by (B.36), the second equality holds by the definition of X and $\mathcal{E}_X$, the first inequality holds since $\underline{\sigma_{\cdot j}} = \min_{k\neq j}\sqrt{\sigma_{kk\cdot j}}$, the third equality holds by the definition of $\{\widetilde{W}_m\}$ in (B.41), and the last inequality holds by the bound on $h(w)$ in (B.42). Regarding the quantity $\bar{w} = \max_m \mathbb{E}\left[\widetilde{W}_m \,|\, Z_j = z_j, G = g\right]$, we have

$$\begin{aligned}
\bar{w} \;=\;& \max_{k\in\mathcal{E}_X}\left\{\frac{\pm\mathbb{E}\left[Z_k \,|\, Z_j = z_j, G = g\right] - z}{\sqrt{\sigma_{kk\cdot j}}} + \frac{\widetilde{z}}{\underline{\sigma_{\cdot j}}}\right\} \\
\le\;& \max_{k\in\mathcal{E}_X}\left\{\frac{\pm\mathbb{E}\left[Z_k \,|\, Z_j = z_j, G = g\right]}{\sqrt{\sigma_{kk\cdot j}}}\right\} + \max_{k\in\mathcal{E}_X}\left\{\frac{1}{\underline{\sigma_{\cdot j}}} - \frac{1}{\sqrt{\sigma_{kk\cdot j}}}\right\} z + \frac{\bar{\rho}_j(|z_j| + c_0||g||_\infty)}{\underline{\sigma_{\cdot j}}} \\
\le\;& \max_{k\in\mathcal{E}_X}\left\{\frac{\pm\left(\sigma_{kj}\sigma_{jj}^{-1}z_j + \sum_{m\in\mathcal{E}_G}(\sigma_{km}\sigma_{mm}^{-1}z_m)\right)}{\sqrt{\sigma_{kk\cdot j}}}\right\} + \frac{z}{\underline{\sigma_{\cdot j}}} + \frac{\bar{\rho}_j(|z_j| + c_0||g||_\infty)}{\underline{\sigma_{\cdot j}}} \\
\le\;& \frac{2\bar{\rho}_j(|z_j| + c_0||g||_\infty)}{\underline{\sigma_{\cdot j}}} + \frac{z}{\underline{\sigma_{\cdot j}}} \le \frac{2\bar{\rho}_j(1 + c_0) + 1}{\underline{\sigma_{\cdot j}}}(t + \epsilon), \qquad \text{(B.44)}
\end{aligned}$$

where $\max\{\pm A\} := \max\{A, -A\}$, the first inequality holds by the definition of $\widetilde{z}$, the second inequality holds by (B.38), and the last inequality holds by the definitions of $\bar{\rho}_j$ and $\underline{\sigma_{\cdot j}}$ and the fact $\sum_{m\in\mathcal{E}_G}\mathbb{1}(\sigma_{km} \neq 0) \le c_0$.

Let δ in (B.43) go to 0, we get the following bound on the density function of the conditional distribution of $||Z||_\infty \mid \{Z_j = z_j, G = g\}$, i.e., when $0 \le t \le C_0\sqrt{\log d}$ and $|z_j|, ||g||_\infty \le t + \epsilon$,

$$f_{g,z_j}(z) \le \frac{6}{\underline{\sigma_{\cdot j}}}(\bar{w} + a_d) \le \frac{6}{\underline{\sigma_{\cdot j}}}\left(\frac{2\bar{\rho}_j(1 + c_0) + 1}{\underline{\sigma_{\cdot j}}}C_1\sqrt{\log d} + C_2\sqrt{\log d}\right), \qquad \text{(B.45)}$$

for any $z \in (\max\{|z_j|, ||g||_\infty\}, t + \epsilon]$. The first inequality holds by (B.43). Regarding the second inequality, we apply the result in (B.44) and bound $(t + \epsilon)$ and a_d by $C_1\sqrt{\log d}$ for some constant C_1. Note $a_d \le C_1\sqrt{\log d}$ is because of the maximal inequalities for sub-Gaussian random variables (Lemma 5.2 in van Handel (2014)). As for $\bar{\rho}_j = \max_{k\in\mathcal{E}_X}\frac{|\sigma_{jk}|}{\sigma_{jj}}$, we have

$$\bar{\rho}_j^2 \le \max_{k\neq j}\frac{\sigma_{kk}}{\sigma_{jj}} \le \frac{\max_j \sigma_{jj}^U}{\min_j \sigma_{jj}^U} \le \frac{\max_j \sigma_{jj}^U}{\lambda_{\min}(\mathbf{\Sigma}^U)} = O(1),$$

where the first inequality holds by the Cauchy-Schwarz inequality, the second inequality holds by the definition of Z and $\sigma_{jj}^U = \sigma_{jj}^V$, the third inequality holds by the fact that $\min_j \sigma_{jj}^U \geq \lambda_{\min}(\mathbf{\Sigma}^U)$, and the last step holds under the stated assumption of Theorem B.4. As for $\underline{\sigma_{\cdot j}} = \min_{k\in\mathcal{E}_X} \sqrt{\sigma_{kk\cdot j}}$ where $\sigma_{kk\cdot j} = \sigma_{kk} - \sigma_{kj}^2\sigma_{jj}^{-1} - \sum_{m\in\mathcal{E}_G} \sigma_{km}^2\sigma_{mm}^{-1} = \mathrm{Var}\left(Z_k \mid Z_j, G\right)$, we have

$$
\begin{aligned}
\frac{1}{\underline{\sigma_{\cdot j}^2}} &= \frac{1}{\min_{k\in\mathcal{E}_X} \mathrm{Var}\left(Z_k \mid Z_j, G\right)} \\
&\leq \frac{1}{\min_k \mathrm{Var}\left(Z_k \mid Z_{\text{-}k}\right)} \\
&= \max_k ((\mathbf{\Sigma}^Z)^{-1})_{kk} \\
&\leq \lambda_{\max}((\mathbf{\Sigma}^Z)^{-1}) \\
&= 1/\lambda_{\min}(\mathbf{\Sigma}^Z) \\
&\leq (\min\{\lambda_{\min}(\mathbf{\Sigma}^U), \lambda_{\min}(\mathbf{\Sigma}^V\})^{-1} \leq b_0,
\end{aligned}
$$

under the stated assumption that $\lambda_{\min}(\mathbf{\Sigma}^U) \geq 1/b_0, \lambda_{\min}(\mathbf{\Sigma}^V) \geq 1/b_0$, where the first inequality holds since (Z_j, G) is a sub-vector of $Z_{\text{-}k} := Z_{(1:d)\setminus k}$, the second equality holds by the relationship between the partial variances and the inverse covariance matrix, and the last three hold by the definitions of $\lambda_{\min}(\cdot), \lambda_{\max}(\cdot)$. Thus we have $f_{g,z_j}(z) \leq C\sqrt{\log d}$ for some constant C, i.e., Lemma B.7 is proved.

□

Lemma B.8. For $z > 0$, we have

$$\frac{\phi(z)}{2(z\vee 1)} \leq \bar{\Phi}(z) = 1 - \Phi(z) \leq \frac{\phi(z)}{z},$$

where $\phi(z), \Phi(z)$ is the PDF and CDF of the standard Gaussian distribution respectively.

Proof of Lemma B.8. This is a simple fact derived from Mill's inequality; see the derivations in the proof of Theorem 3 in Chernozhukov et al. (2015). □

Lemma B.9. Whenever $0 \leq t \leq C_0\sqrt{\log d}$ for some constant $C_0 > 0$, and $\epsilon = c/\max\{(\log d)^{3/2}, \mathfrak{p}\log d\}$ for some small enough constant $c > 0$, we have

$$\Lambda(t,\epsilon,\mathfrak{p}) := \frac{(1-2\bar{\Phi}(\frac{t+\epsilon}{\sigma_j}))^{\mathfrak{p}}}{1-(1-2\bar{\Phi}(\frac{t}{\sigma_j}))^{\mathfrak{p}}} \cdot \phi(\frac{t-\epsilon}{\sigma_j}) = O\left(\frac{\sqrt{\log d}}{\mathfrak{p}}\right). \tag{B.46}$$

Proof of Lemma B.9. By Lemma B.8, we can simplify $\Lambda(t,\epsilon,\mathfrak{p})$ into the following

$$\Lambda(t,\epsilon,\mathfrak{p}) \leq \frac{\left(1 - \frac{\phi(\frac{t+\epsilon}{\sigma_j})}{\frac{t+\epsilon}{\sigma_j}\vee 1}\right)^{\mathfrak{p}}}{1-\left(1-\frac{\phi(\frac{t}{\sigma_j})}{\frac{t}{\sigma}\vee 1}\right)^{\mathfrak{p}}} \cdot \phi(\frac{t-\epsilon}{\sigma_j}).$$

When $\frac{t}{\sigma_j} \leq 1$, we have $\frac{t+\epsilon}{\sigma_j} \leq 2$ due to the choice of ϵ. Because $\frac{t}{\sigma_j} > 0, \frac{t+\epsilon}{\sigma_j} > 0$ and $\phi(z)$ is monotonically decreasing when $z > 0$, we then have the the bound below,

$$\Lambda(t,\epsilon,p) \leq \frac{(1-\phi(2)/2)^{\mathfrak{p}}}{1-(1-\phi(1))^{\mathfrak{p}}} = O\left(\frac{\sqrt{\log d}}{\mathfrak{p}}\right),$$

where the second inequality holds due to $0 < \phi(2) < \phi(1) < 0.5$ and $\mathfrak{p} > 1$. Now it suffices to consider the case where $\frac{t+\epsilon}{\sigma_j} > \frac{t}{\sigma_j} > 1$ and deal with the following

$$\Lambda(t,\epsilon,\mathfrak{p}) \leq \frac{\left(1 - \frac{\phi(\frac{t+\epsilon}{\sigma_j})}{\frac{t+\epsilon}{\sigma_j}}\right)^{\mathfrak{p}}}{1 - \left(1 - \frac{\phi(\frac{t+\epsilon}{\sigma_j})}{\frac{t+\epsilon}{\sigma_j}}\right)^{\mathfrak{p}}} \cdot \phi\left(\frac{t-\epsilon}{\sigma_j}\right).$$

We further bound $\Lambda(t,\epsilon,\mathfrak{p})$ as

$$\begin{aligned}
\Lambda(t,\epsilon,\mathfrak{p}) &\leq \frac{\left(1 - \frac{\phi(\frac{t+\epsilon}{\sigma_j})}{\frac{t+\epsilon}{\sigma_j}}\right)^{\mathfrak{p}}}{1 - \left(1 - \frac{\phi(\frac{t+\epsilon}{\sigma_j})}{\frac{t+\epsilon}{\sigma_j}}\right)^{\mathfrak{p}}} \cdot \phi\Big(\frac{t+\epsilon}{\sigma_j}\Big) \cdot e^{\frac{t\epsilon}{2\sigma_j^2}} \\
&\leq 2\, \frac{\left(1 - \frac{\phi(\frac{t+\epsilon}{\sigma_j})}{\frac{t+\epsilon}{\sigma_j}}\right)^{\mathfrak{p}}}{1 - \left(1 - \frac{\phi(\frac{t+\epsilon}{\sigma_j})}{\frac{t+\epsilon}{\sigma_j}}\right)^{\mathfrak{p}}} \cdot \phi\Big(\frac{t+\epsilon}{\sigma_j}\Big) \\
&:= 2e^{H(\lambda)} \cdot \frac{t+\epsilon}{\mathfrak{p}\sigma_j}
\end{aligned}$$

where the first inequality comes from rearranging, the second inequality holds since $\exp\big(\frac{t\epsilon}{2\sigma_j^2}\big) < 2$ for $t \leq C_0\sqrt{\log d}$. This is because $\epsilon = c/\max\{(\log d)^{3/2}, \mathfrak{p}\log d\}$ for some small enough constant $c > 0$. The last line holds by rewriting using some new notations: $\lambda := \mathfrak{p}\ \frac{\phi(\frac{t+\epsilon}{\sigma_j})}{\frac{t+\epsilon}{\sigma_j}}$ and

$$H(\lambda) := \log\left(\frac{(1-\frac{\lambda}{\mathfrak{p}})^{\mathfrak{p}}}{1-(1-\frac{\lambda}{\mathfrak{p}})^{\mathfrak{p}}} \cdot \lambda\right) = \mathfrak{p}\log\Big(1 - \frac{\lambda}{\mathfrak{p}}\Big) - \log\Big(1 - (1-\frac{\lambda}{\mathfrak{p}})^{\mathfrak{p}}\Big) + \log\lambda. \tag{B.47}$$

Since $\frac{t+\epsilon}{\sigma_j} > 1$, we have $0 < \lambda < \mathfrak{p}$. Below we will first deal with $H(\lambda)$ then obtain the bound on $\Lambda(t,\epsilon,\mathfrak{p})$. To bound $H(\lambda)$, consider taking the derivative of $H(\lambda)$ with respect to λ, then we have

$$\begin{aligned}
H'(\lambda) &= \frac{\mathfrak{p}}{\lambda - \mathfrak{p}} - \frac{(1-\frac{\lambda}{\mathfrak{p}})^{(\mathfrak{p}-1)}}{1-(1-\frac{\lambda}{\mathfrak{p}})^{\mathfrak{p}}} + \frac{1}{\lambda} \\
&= \frac{\mathfrak{p}}{\lambda - \mathfrak{p}} - \frac{1}{1-\frac{\lambda}{\mathfrak{p}}} \cdot \frac{(1-\frac{\lambda}{\mathfrak{p}})^{\mathfrak{p}}}{1-(1-\frac{\lambda}{\mathfrak{p}})^{\mathfrak{p}}} + \frac{1}{\lambda} \\
&\leq \frac{\mathfrak{p}}{\lambda - \mathfrak{p}} - \frac{1}{1-\frac{\lambda}{\mathfrak{p}}} \cdot \frac{1-\lambda}{1-(1-\lambda)} + \frac{1}{\lambda} \\
&= \frac{\mathfrak{p}}{\lambda - \mathfrak{p}} + \frac{1}{1-\frac{\lambda}{\mathfrak{p}}} - \frac{1}{\lambda} \cdot \frac{1}{1-\frac{\lambda}{\mathfrak{p}}} + \frac{1}{\lambda} \\
&\leq \frac{1}{\lambda}\Big(1 - \frac{\mathfrak{p}}{\mathfrak{p}-\lambda}\Big) < 0,
\end{aligned}$$

where the first inequality holds by the Bernoulli's inequality: $(1+x)^r \geq 1 + rx$ when $r \in \mathbb{N}, 1+x \geq 0$, and the last inequality holds since $0 < \lambda < \mathfrak{p}$. Now we have $H(\lambda)$ is monotone decreasing. When

$0 \le t \le C_0\sqrt{\log d}$, we will first find the lower bound on $\lambda = \mathfrak{p}\,\frac{\phi(\frac{t+\epsilon}{\sigma_j})}{\frac{t+\epsilon}{\sigma_j}}$, denoted by $\underline{\lambda}$. Then we have $H(\lambda)$ is bounded by $H(\underline{\lambda})$ due to its monotonicity. Regarding $\underline{\lambda}$, we denote $\bar{x} := 2C_0\sqrt{\log d}/\sigma_j$ and note $\frac{\phi(x)}{x}$ is monotone decreasing when $x \ge 0$. Then we have, when $0 \le t \le C_0\sqrt{\log d}$,

$$\mathfrak{p}\,\frac{\phi(\frac{t+\epsilon}{\sigma_j})}{\frac{t+\epsilon}{\sigma_j}} \ge \mathfrak{p}\,\frac{\phi(\bar{x})}{\bar{x}} \ge \frac{\mathfrak{p}}{d^{a_1}} := \underline{\lambda},$$

where $a_1 > 2$. Therefore we obtain

$$H(\lambda) \le H(\underline{\lambda}) = \log\left(\frac{(1-\frac{\lambda}{\mathfrak{p}})^{\mathfrak{p}}}{1-(1-\frac{\lambda}{\mathfrak{p}})^{\mathfrak{p}}}\cdot\lambda\right)\Bigg|_{\lambda=\underline{\lambda}} \le \log\left(\frac{\lambda}{1-(1-\mathfrak{p}\frac{\lambda}{\mathfrak{p}}+\frac{(\mathfrak{p}-1)\mathfrak{p}}{2}\frac{\lambda^2}{\mathfrak{p}^2})}\right)\Bigg|_{\lambda=\underline{\lambda}} \le C', \tag{B.48}$$

where the second inequality holds due to the fact that $(1-\frac{\lambda}{\mathfrak{p}})^{\mathfrak{p}} \le 1$, $\frac{\lambda}{\mathfrak{p}} \in [0,1]$ and Lemma B.11. The third inequality holds since $\underline{\lambda} = \frac{\mathfrak{p}}{d^{a_1}} \le \frac{1}{d^{a_1-1}} < \frac{1}{d}$, then we have

$$\left(\frac{\lambda}{1-(1-\mathfrak{p}\frac{\lambda}{\mathfrak{p}}+\frac{(\mathfrak{p}-1)\mathfrak{p}}{2}\frac{\lambda^2}{\mathfrak{p}^2})}\right)\Bigg|_{\lambda=\underline{\lambda}} = \frac{\underline{\lambda}}{\underline{\lambda}-\frac{2(\mathfrak{p}-1)}{\mathfrak{p}}\underline{\lambda}^2} \le \frac{\underline{\lambda}}{\underline{\lambda}-2\underline{\lambda}^2} = \frac{1}{1-2\underline{\lambda}} \le C_1',$$

for some constant C_1'. Now we figure out the bound on $\Lambda(t,\epsilon,\mathfrak{p})$,

$$\begin{aligned}
\Lambda(t,\epsilon,\mathfrak{p}) &\le \frac{\left(1-\frac{\phi(\frac{t+\epsilon}{\sigma_j})}{\frac{t+\epsilon}{\sigma_j}}\right)^{\mathfrak{p}}}{1-\left(1-\frac{\phi(\frac{t+\epsilon}{\sigma_j})}{\frac{t+\epsilon}{\sigma_j}}\right)^{\mathfrak{p}}}\cdot\phi\Big(\frac{t-\epsilon}{\sigma_j}\Big) \\
&\le \frac{\left(1-\frac{\phi(\frac{t+\epsilon}{\sigma_j})}{\frac{t+\epsilon}{\sigma_j}}\right)^{\mathfrak{p}}}{1-\left(1-\frac{\phi(\frac{t+\epsilon}{\sigma_j})}{\frac{t+\epsilon}{\sigma_j}}\right)^{\mathfrak{p}}}\cdot\phi\Big(\frac{t+\epsilon}{\sigma_j}\Big)\cdot e^{\frac{t\epsilon}{2\sigma_j^2}} \\
&\le 2\,\frac{\left(1-\frac{\phi(\frac{t+\epsilon}{\sigma_j})}{\frac{t+\epsilon}{\sigma_j}}\right)^{\mathfrak{p}}}{1-\left(1-\frac{\phi(\frac{t+\epsilon}{\sigma_j})}{\frac{t+\epsilon}{\sigma_j}}\right)^{\mathfrak{p}}}\cdot\phi\Big(\frac{t+\epsilon}{\sigma_j}\Big) \\
&\le 2e^{H(\lambda)}\cdot\frac{t+\epsilon}{\mathfrak{p}\sigma_j} \le \frac{C\sqrt{\log d}}{\mathfrak{p}},
\end{aligned}$$

where the second inequality comes from rearranging, the third inequality holds since $\exp\left(\frac{t\epsilon}{2\sigma_j^2}\right) < 2$ for $t \le C_0\sqrt{\log d}$. This is because $\epsilon = c/\max\{(\log d)^{3/2}, \mathfrak{p}\log d\}$ for some small enough constant $c > 0$. And the last line holds by (B.47) and (B.48). Therefore Lemma B.9 is established. □

Lemma B.10. Under the same conditions as Lemma B.9, we have

$$\frac{(1-2\bar{\Phi}(\frac{t+\epsilon}{\sigma_j}))^{\mathfrak{p}-1}}{1-(1-2\bar{\Phi}(\frac{t}{\sigma_j}))^{\mathfrak{p}}}\cdot\left(\mathfrak{p}\phi(\frac{t-\epsilon}{\sigma_j})\right)^2 = O\left(\sqrt{\log d}\right). \tag{B.49}$$

Proof. Note that the result (B.46) in Lemma B.9 can be rewritten as

$$\mathfrak{p}\ \Lambda(t,\epsilon,\mathfrak{p}) = \frac{(1-2\bar{\Phi}(\frac{t+\epsilon}{\sigma_j}))^{\mathfrak{p}}}{1-(1-2\bar{\Phi}(\frac{t}{\sigma_j}))^{\mathfrak{p}}} \cdot \left(\mathfrak{p}\phi(\frac{t-\epsilon}{\sigma_j})\right) = O\left(\sqrt{\log d}\right).$$

By similar derivations as in the proof of Lemma B.9, we can establish

$$\frac{(1-2\bar{\Phi}(\frac{t+\epsilon}{\sigma_j}))^{\mathfrak{p}-1}}{1-(1-2\bar{\Phi}(\frac{t}{\sigma_j}))^{\mathfrak{p}}} \cdot \left(\mathfrak{p}\phi(\frac{t-\epsilon}{\sigma_j})\right)^2 = O\left(\sqrt{\log d}\right).$$

□

Lemma B.11. For $x \in [0,1]$, we have $(1-x)^{\mathfrak{p}} \leq 1 - \mathfrak{p}x + 0.5\mathfrak{p}(\mathfrak{p}-1)x^2$.

Proof. When $\mathfrak{p} = 1$, the above simply holds. Now we consider the case where $\mathfrak{p} > 1$. Let $Q(x) = (1-x)^{\mathfrak{p}} - (1 - \mathfrak{p}x + 0.5\mathfrak{p}(\mathfrak{p}-1)x^2)$, we have $Q(0) = 0$ and

$$Q'(x) = -\mathfrak{p}(1-x)^{(\mathfrak{p}-1)} + \mathfrak{p} - \mathfrak{p}(\mathfrak{p}-1)x \leq -\mathfrak{p}(1-(\mathfrak{p}-1)x) + \mathfrak{p} - \mathfrak{p}(\mathfrak{p}-1)x = 0, \tag{B.50}$$

where the inequality holds by applying Bernoulli's inequality to $(1-x)^{(\mathfrak{p}-1)}$ for $\mathfrak{p} > 1, x \in [0,1]$. Therefore, $Q(x)$ is monotonically decreasing, and the statement is proved. □

B.4 Ancillary lemmas for Theorem 3.2

Remark B.3. *Recall that the connectivity assumption of Theorem 3.2 assumes that there exists a disjoint $\mathfrak{p}$-partition of nodes $\cup_{\ell=1}^{\mathfrak{p}} \mathcal{C}_\ell = [d]$ such that $\sigma_{jk}^U = \sigma_{jk}^V = 0$ when $j \in \mathcal{C}_\ell$ and $k \in \mathcal{C}_{\ell'}$ for some $\ell \neq \ell'$. The more general version of the variance condition assumes: $a_0 \leq \sigma_{jj}^U = \sigma_{jj}^V \leq a_1$, $\forall j \in [d]$; given any $j \in \mathcal{C}_\ell^U$ with some ℓ, there exists at least one $m \in \mathcal{C}_{\ell'}^U$ such that $\sigma_{jj}^U = \sigma_{mm}^U$ for any $\ell' \neq \ell$. Denote $\widetilde{\sigma}_{jk}^U = \sigma_{jk}^U / \sqrt{\sigma_{jj}^U \sigma_{kk}^U}$. And the general covariance condition says that there exists some $\sigma_0 < 1$ such that $|\widetilde{\sigma}_{jk}^V| = |\sigma_{jk}^V| / \sqrt{\sigma_{jj}^V \sigma_{kk}^V} \leq \sigma_0$ for any $j \neq k$ and $|\{(j,k) : j \neq k, |\widetilde{\sigma}_{jk}^U| = |\sigma_{jk}^U| / \sqrt{\sigma_{jj}^U \sigma_{kk}^U} > \sigma_0\}| \leq b_0$ for some constant b_0.*

Lemma B.12. For the term $\mathrm{II}_1 = \mathbb{E}\left[e^{-\beta(||Z||_\infty - |Z_j|)} \cdot \mathbb{1}(E_1) \cdot \mathbb{1}(t - \epsilon \leq ||Z||_\infty \leq t + \epsilon)\right]$ with E_1 defined in (B.20) and $\epsilon = c/\max\{(\log d)^{3/2}, \mathfrak{p}\log d\}$ for some small enough constant $c > 0$, whenever t satisfies $0 \leq t \leq C_0\sqrt{\log d}$ for some constant $C_0 > 0$, we have

$$\frac{\mathrm{II}_1}{\mathbb{P}(||V||_\infty > t)} \leq \frac{C'\epsilon\log d}{\beta\mathfrak{p}} \left(1 + \frac{b_0}{\sqrt{1-(s+(1-s)\sigma_0)^2}}\right). \tag{B.51}$$

for any $s \in (0,1)$, where $\sigma_0 < 1$ and b_0 are the constants in the assumption of Theorem 3.2.

Remark B.4. *Recall the definition of $Z = W(s)$. Hence the term II_1 depends on s. In Lemma B.5, we are able to derive a uniform upper bound when assuming the minimal eigenvalue condition as in Theorem B.4. Since Theorem 3.2 does not make assumptions about the minimal eigenvalue condition, we will bound the term II_1 differently and the upper bound depend on s, as showed in the following proof.*

Proof of Lemma B.12. We basically use the same proof strategy as Lemma but will separately deal with two cases. First recall that

$$\mathrm{II}_1 = \mathbb{E}\left[e^{-\beta(||Z||_\infty-|Z_j|)}\cdot \mathbb{1}(||Z||_\infty > ||G||_\infty, ||Z||_\infty > Z_j)\cdot \mathbb{1}(t-\epsilon \le ||Z||_\infty \le t+\epsilon)\right].$$

We define $Z^\dagger = (Z_k)_{k\in\mathcal{E}^\dagger}$ where

$$\mathcal{E}^\dagger = \{j\}\cup\mathcal{E}_G\cup\{k\in[d] : |\widetilde{\sigma}^U_{jk}| \le \sigma_0, \max_{m\in\mathcal{E}_G}\{|\widetilde{\sigma}^U_{mk}|\}\le\sigma_0\}. \tag{B.52}$$

Under the condition of Theorem 3.2, we have $|[d]\setminus\mathcal{E}^\dagger| \le |\{(j,k) : j\ne k, |\widetilde{\sigma}^U_{jk}|\} > \sigma_0\}| \le b_0$ for some constant b_0. Note we can write $1 = \mathbb{1}(||Z^\dagger||_\infty = ||Z||_\infty) + \sum_{k\in[d]\setminus\mathcal{E}^\dagger}\mathbb{1}(|Z_k| = ||Z||_\infty)$. Then we have

$$\frac{\mathrm{II}_1}{\mathbb{P}(||V||_\infty > t)} \le \frac{\mathrm{II}_1^\dagger}{\mathbb{P}(||V||_\infty > t)} + b_0\cdot\max_{k\in[d]\setminus\mathcal{E}^\dagger}\frac{\mathrm{II}_1^{(k)}}{\mathbb{P}(||V||_\infty > t)}, \tag{B.53}$$

where $\mathrm{II}_1^\dagger$ and $\mathrm{II}_1^{(k)}$ are defined as

$$\begin{aligned}\mathrm{II}_1^\dagger &:= \mathbb{E}\left[e^{-\beta(||Z^\dagger||_\infty-|Z_j|)}\cdot\mathbb{1}(||Z^\dagger||_\infty > ||G||_\infty, ||Z^\dagger||_\infty > Z_j)\cdot\mathbb{1}(t-\epsilon\le ||Z^\dagger||_\infty\le t+\epsilon)\right],\\ \mathrm{II}_1^{(k)} &:= \mathbb{E}\left[e^{-\beta(|Z_k|-|Z_j|)}\cdot\mathbb{1}(|Z_k| > ||G||_\infty, |Z_k| > Z_j)\cdot\mathbb{1}(t-\epsilon\le |Z_k|\le t+\epsilon)\right].\end{aligned} \tag{B.54}$$

Denote the conditional density function of $||Z^\dagger||_\infty \mid Z_j = z_j, G = g$ by $f^\dagger_{g,z_j}(u)$. Then we apply exactly the same derivations as in Lemma B.5 (except that $f^\dagger_{g,z_j}(u)$ is bounded using Lemma B.13 instead of Lemma B.7) and obtain the following bound

$$\frac{\mathrm{II}_1^\dagger}{\mathbb{P}(||V||_\infty > t)} \le \frac{C'\epsilon\log d}{\beta\mathfrak{p}}. \tag{B.55}$$

Regarding the term $\mathrm{II}_1^{(k)}$, we follow the same derivations as in the beginning of the proof of Lemma B.5. Specifically, we have

$$\begin{aligned}\mathrm{II}_1^{(k)} &= \mathbb{E}\left[e^{-\beta(|Z_k|-|Z_j|)}\cdot\mathbb{1}(|Z_k| > ||G||_\infty, |Z_k| > Z_j)\cdot\mathbb{1}(t-\epsilon\le |Z_k|\le t+\epsilon)\right]\\ &= \mathbb{E}\left[e^{\beta|Z_j|}\cdot\mathbb{1}(||G||_\infty\le t+\epsilon, |Z_j|\le t+\epsilon)\left(\int_{t-\epsilon}^{t+\epsilon} f_{Z_j}(u)e^{-\beta u}\mathbb{1}(u > ||G||_\infty, u > |Z_j|)du\right)\right],\end{aligned} \tag{B.56}$$

where $f_{Z_j,G}(u)$ denotes the conditional density of Z_k given Z_j, G. Recall the construction of G described in the proof of Theorem 3.2, we have for any $k\ne j, k\notin\mathcal{E}_G = \{m\in[d] : Z_m = G_\ell \text{ for some } \ell\in[\mathfrak{p}]\}$, there exists at most one $m\in\{j\}\cup\mathcal{E}_G$, such that Z_k and Z_m belong to the same component. Denote that random variable by Z_{m_0}, then $f_{Z_j,G}(u)$ is just the conditional density function of Z_k given Z_{m_0}. Since Z follows a multivariate Gaussian distribution, we can immediately figure out the expression of the conditional density $f_{Z_{m_0}}(u)$ and simply derive a bound

$$
\begin{aligned}
f_{Z_j,G}(u) = f_{Z_{m_0}}(u) &\le \frac{1}{\sqrt{2\pi \mathrm{Var}\left(Z_k \mid Z_{m_0}\right)}} \\
&= \frac{1}{\sqrt{2\pi(\sigma_{kk} - \sigma_{km_0}^2/\sigma_{m_0m_0}))}} \\
&= \frac{1}{\sqrt{2\pi\sigma_{kk}}} \cdot \frac{1}{1 - \sigma_{km_0}^2/(\sigma_{kk}\sigma_{m_0m_0})} \\
&\le \frac{1}{\sqrt{2\pi a_0}} \cdot \frac{1}{1 - \sigma_{km_0}^2/(\sigma_{kk}\sigma_{m_0m_0})},
\end{aligned} \tag{B.57}
$$

where $\sigma_{kk} = \mathrm{Var}\left(Z_k\right), \sigma_{m_0m_0} = \mathrm{Var}\left(Z_{m_0}\right), \sigma_{km_0} = \mathrm{Cov}\left(Z_k, Z_{m_0}\right)$ and we use the fact that $\sigma_{kk} = \mathrm{Var}\left(Z_k\right) = \sigma^U_{kk} \ge a_0$ (under the general variance assumption). Note $Z = \sqrt{s}U + \sqrt{1-s}V$, then we have $\sigma^2_{km_0} = (\mathrm{Cov}\left(Z_k, Z_{m_0}\right))^2 = (s\sigma^U_{km_0} + (1-s)\sigma^V_{km_0})^2$ where $m_0 \in \{j\} \cup \mathcal{E}_G$. Since $|\widetilde{\sigma}^U_{km_0}| \le 1$ by definition and $|\widetilde{\sigma}^V_{km_0}| \le \sigma_0$ under the assumption of Theorem 3.2, we have

$$
(s\sigma^U_{km_0} + (1-s)\sigma^V_{km_0})^2/(\sigma_{kk}\sigma_{m_0m_0}) = (s\widetilde{\sigma}^U_{km_0} + (1-s)\widetilde{\sigma}^V_{km_0})^2 \le (s + (1-s)\sigma_0)^2. \tag{B.58}
$$

Now we obtain a upper bound on the conditional density function $f_{Z_j,G}(u)$ based on (B.57) and (B.58). Combining this bound and following the same derivations as in Lemma B.5 to deal with the term in (B.56), we establish the upper bound on the term $\mathrm{II}_1^{(k)}/\mathbb{P}(||V||_\infty > t)$ for any $k \in [d] \setminus \mathcal{E}^\dagger$,

$$
\frac{\mathrm{II}_1^{(k)}}{\mathbb{P}(||V||_\infty > t)} \le \frac{C'\epsilon \log d}{\beta \mathfrak{p}} \cdot \frac{1}{\sqrt{1 - (s + (1-s)\sigma_0)^2}}. \tag{B.59}
$$

Combining (B.53), (B.54),(B.55) with (B.59), we derive the bound in (B.51). $\square$

Lemma B.13. Recall that the density function of the conditional distribution of $||Z^\dagger||_\infty \mid \{Z_j = z_j, G = g\}$ is denoted by $f^\dagger_{g,z_j}(z)$ where $Z^\dagger$ is defined in .... Suppose $\epsilon > 0$, when $0 \le t \le C_0\sqrt{\log d}$ for some constant $C_0 > 0$ and $|z_j|, ||g||_\infty \le t + \epsilon$, we have

$$
f^\dagger_{g,z_j}(z) \le C\sqrt{\log d}, \quad \forall\ z \in (\max\{|z_j|, ||g||_\infty\}, t + \epsilon]. \tag{B.60}
$$

where the finite constant C depends on a_0 and $\sigma_0 < 1$.

Proof of Lemma B.13. Following exactly the same derivations as in Lemma B.7 (up to (B.45)), we have

$$
f^\dagger_{g,z_j}(z) \le \frac{6}{\underline{\sigma_{\cdot j}}} \left(\frac{2\bar{\rho}_j(1+c_0) + 1}{\underline{\sigma_{\cdot j}}} C_1\sqrt{\log d} + C_2\sqrt{\log d} \right), \tag{B.61}
$$

for any $z \in (\max\{|z_j|, ||g||_\infty\}, t+\epsilon]$, where C_1, C_2 are some constants. First, $\bar{\rho}_j$ is defined in (B.37). Simply, we have

$$
\bar{\rho}_j \le \max_{k \ne j} \frac{|\sigma_{jk}|}{\sigma_{jj}} \le \frac{\max_j \sigma^U_{jj}}{\min_j \sigma^U_{jj}} \le \frac{a_1}{a_0}
$$

under the general variance assumption. Recall the construction of G described in the proof of Theorem 3.2, we have for any $k \ne j, k \notin \mathcal{E}_G = \{m \in [d] : Z_m = G_\ell \text{ for some } \ell \in [\mathfrak{p}]\}$, there is at most one $m \in \mathcal{E}_G$, such that Z_k and Z_m belong to the same component. Then we have

$$
\sum_{m \in \mathcal{E}_G} \mathbb{1}(\sigma_{km} \ne 0) \le 1,
$$

hence $c_0 = 1$ by definition. Also note by the definition of $Z^\dagger$ and $\mathcal{E}^\dagger$ in (B.52), for any $k \in \mathcal{E}^\dagger, k \neq j, k \neq \mathcal{E}_G$, we have

$$\max\{|\widetilde{\sigma}_{jk}^U|, |\widetilde{\sigma}_{jk}^V|\} \leq \sigma_0, \quad \max_{m\in\mathcal{E}_G}\{|\widetilde{\sigma}_{mk}^U|, |\widetilde{\sigma}_{mk}^V|\} \leq \sigma_0 \tag{B.62}$$

under the assumption of Theorem 3.2. We will take advantage of this together with the above property of G to derive a bound on $\underline{\sigma_{\cdot j}}$. Similarly as in (B.37), we have $\underline{\sigma_{\cdot j}^2} := \min_{k\in\mathcal{E}_X} \mathrm{Var}\,(Z_k \,|\, Z_j, G)$ with $\mathcal{E}_X := \{k \in \mathcal{E}^\dagger : k \neq j, k \notin \mathcal{E}_G\}$. For each $k \in \mathcal{E}_X$, we have it can at most belong to the same component as one of $\{j\} \cup \mathcal{E}_G$, due to the property of G. Then we have

$$\begin{aligned}
\mathrm{Var}\,(Z_k \,|\, Z_j, G) &\geq \min\{\mathrm{Var}\,(Z_k \,|\, Z_j)\,, \min_{m\in\mathcal{E}_G}\{\mathrm{Var}\,(Z_k \,|\, Z_m)\}\} \\
&= \sigma_{kk}\cdot \min\{1-\sigma_{jk}^2/(\sigma_{jj}\sigma_{kk}), \min_{m\in\mathcal{E}_G}\{1-\sigma_{mk}^2/(\sigma_{mm}\sigma_{kk})\}\} \\
&\geq a_0\cdot \min\{1-\sigma_{jk}^2/(\sigma_{jj}\sigma_{kk}), \min_{m\in\mathcal{E}_G}\{1-\sigma_{mk}^2/(\sigma_{mm}\sigma_{kk})\}\}.
\end{aligned} \tag{B.63}$$

since (Z_j, G) are all independent and $\sigma_{kk} = \mathrm{Var}\,(Z_k) = \sigma_{kk}^U \geq a_0$ (under the general variance assumption). Recall the definition of $Z = \sqrt{s}U + \sqrt{1-s}V$, we have

$$|\sigma_{mk}|/\sqrt{\sigma_{mm}\sigma_{kk}} = |\mathrm{Cov}\,(Z_k, Z_m)\,|/\sqrt{\sigma_{mm}\sigma_{kk}} = |s\widetilde{\sigma}_{mk}^U + (1-s)\widetilde{\sigma}_{mk}^V| \leq \sigma_0, \quad \forall\ s\in[0,1], \tag{B.64}$$

when $m \in \{j\} \cup \mathcal{E}_G$. This is due to (B.62). Then we can derive a bound on $1/\underline{\sigma_{\cdot j}^2}$, i.e.,

$$\begin{aligned}
\frac{1}{\underline{\sigma_{\cdot j}^2}} &= \frac{1}{\min_{k\in\mathcal{E}_X}\mathrm{Var}\,(Z_k \,|\, Z_j, G)} \\
&\leq \frac{1}{a_0 \min_{k\in\mathcal{E}_X}\min\{1-\sigma_{jk}^2/(\sigma_{jj}\sigma_{kk}), \min_{m\in\mathcal{E}_G}\{1-\sigma_{mk}^2/(\sigma_{mm}\sigma_{kk})\}\}} \\
&\leq \frac{1}{a_0(1-\sigma_0^2)},
\end{aligned}$$

where the first inequality holds by (B.63) and the second equality holds by (B.64). Combining the above bound with (B.61), we finally establish (B.60) for some finite constant C. □

C Ancillary propositions for FDR control

Throughout this section, we introduce some new notations. For a given mean zero random vector $\boldsymbol{Y} \in \mathbb{R}^d$ with positive semi-definite covariance matrix $\boldsymbol{\Sigma}^Y := \mathbb{E}\left[\boldsymbol{Y}\boldsymbol{Y}^\top\right] \in \mathbb{R}^{d\times d}$, we denote its Gaussian counterpart by $\boldsymbol{Z} \in \mathbb{R}^d$ (i.e., $\mathbb{E}\left[\boldsymbol{Z}\right] = \boldsymbol{0}$ and its covariance matrix $\mathbb{E}\left[\boldsymbol{Z}\boldsymbol{Z}^\top\right] := \boldsymbol{\Sigma}^Z$ equals $\boldsymbol{\Sigma}^Y = (\sigma_{jk}^Y)_{1\leq j,k\leq d}$). Consider n i.i.d. copies of $\boldsymbol{Y}$, denoted by $\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_n \in \mathbb{R}^d$. We define the maximum $T_{\boldsymbol{Y}}$ and $T_{\boldsymbol{Z}}$ as below,

$$T_{\boldsymbol{Y}} := \left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{Y}_i\right\|_\infty, \quad T_{\boldsymbol{Z}} := ||\boldsymbol{Z}||_\infty, \tag{C.1}$$

where $q(\alpha; T_{\boldsymbol{Y}})$ and $q(\alpha; T_{\boldsymbol{Z}})$ ($\alpha \in [0,1]$) are the corresponding upper quantile functions. Define the Gaussian multiplier bootstrap counterpart as

$$T_{\boldsymbol{W}} := \left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{Y}_i\xi_i\right\|_\infty, \tag{C.2}$$

where $\xi_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0,1)$ and are independent from $\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_n$. Let $q_\xi(\alpha; T_{\boldsymbol{W}})$ be the conditional quantile of $T_{\boldsymbol{W}}$, then we have $\mathbb{P}_\xi\left(T_{\boldsymbol{W}} \geq q_\xi(\alpha; T_{\boldsymbol{W}})\right) = \alpha$. Note that we use the ξ subscript to remind ourselves that the probability measure is induced by the multiplier random variables $\{\xi_i\}_{i=1}^n$ conditional on $\{\boldsymbol{Y}_i\}_{i=1}^n$. And we have the covariance matrix of $\frac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{Y}_i \xi_i$ (conditional on $\{\boldsymbol{Y}_i\}_{i=1}^n$) equals $\boldsymbol{\Sigma}^W := \frac{1}{n} \sum_{i=1}^n \boldsymbol{Y}_i \boldsymbol{Y}_i^\top$. Denote $\Delta_\infty = ||\boldsymbol{\Sigma}^Z - \boldsymbol{\Sigma}^W||_\infty$, which measures the maximal differences between the true covariance matrix $\boldsymbol{\Sigma}^Z$ and the sample version $\boldsymbol{\Sigma}^W$.

C.1 Cramér-type deviation bounds for the Gaussian multiplier bootstrap

Based on the Cramér-type Gaussian comparison bound in Theorem 3.1, the Cramér-type approximation bound (Kuchibhotla et al., 2021), the maximal inequalities and a careful treatment to the comparison of quantiles, we will establish the Cramér-type deviation bounds for the Gaussian multiplier bootstrap (CGMB) in this section.

Proposition C.1 (CGMB)**.** *Assuming the covariance matrix $\boldsymbol{\Sigma}^Y$ satisfies $0 < c_1 \leq \sigma_{jj}^Y \leq c_2 < \infty$, for any $j \in [d]$ and $\boldsymbol{Y}$ satisfies the tail condition that $\max_{1 \leq i \leq n} \max_{1 \leq j \leq p} ||\boldsymbol{Y}_{ij}||_{\psi_1} \leq K_3$ for some constants c_1, c_2, K_3, under the scaling condition $(\log ed)^3 (\log(ed+n))^{56/3}/n = o(1)$, we have the following bound,*

$$\sup_{\alpha \in [\alpha_L, 1]} \left| \frac{\mathbb{P}(T_{\boldsymbol{Y}} > q_\xi(\alpha; T_{\boldsymbol{W}}))}{\mathbb{P}(T_{\boldsymbol{Z}} > q(\alpha; T_{\boldsymbol{Z}}))} - 1 \right| = O\left(\frac{(\log d)^{11/6}}{n^{1/6} \alpha_L^{1/3}} + \frac{(\log d)^{19/6}}{n^{1/6}} \right), \tag{C.3}$$

where α_L satisfies $q(\alpha_L; T_{\boldsymbol{Z}}) = O\left(\sqrt{\log d}\right)$ and $\frac{\log^{11} d}{n \alpha_L} = O(1)$.

The proof can be found in Appendix C.2. In practice, there are many situations where the relevant statistics come from the maxima of approximated averages. In particular, the test statistics in our node selection problem can not be directly expressed as maxima of scaled averages, but can be approximated by a $T_{\boldsymbol{Y}}$-like term with the approximation error suitably controlled. Therefore, we also prove an extended version of Proposition C.1. Suppose the statistics of interest and its Gaussian multiplier bootstrap counterpart, denoted by T and $T^{\mathcal{B}}$ respectively, can be approximated by $T_{\boldsymbol{Y}}$ (defined in (C.1)) and $T_{\boldsymbol{W}}$ (defined in (C.2)). The quantile functions $q(\alpha; T)$ and $q_\xi(\alpha; T^{\mathcal{B}})$ are defined correspondingly.

Proposition C.2 (CGMB with approximation)**.** *Under the same conditions as in Proposition C.1 and the additional assumption about the differences between the maximum statistics:*

$$\mathbb{P}(|T - T_{\boldsymbol{Y}}| > \zeta_1) < \zeta_2, \tag{C.4}$$

$$\mathbb{P}(\mathbb{P}_\xi(|T^{\mathcal{B}} - T_{\boldsymbol{W}}| > \zeta_1) > \zeta_2) < \zeta_2, \tag{C.5}$$

where $\zeta_1, \zeta_2 \geq 0$ characterize the approximation error and satisfy $\zeta_1 \log d = O(1), \zeta_2 = O(\alpha_L)$, we have the following Cramér-type deviation bound

$$\sup_{\alpha \in [\alpha_L, 1]} \left| \frac{\mathbb{P}(T > q_\xi(\alpha; T^{\mathcal{B}}))}{\mathbb{P}(T_{\boldsymbol{Z}} > q(\alpha; T_{\boldsymbol{Z}}))} - 1 \right| = \eta(d, n, \zeta_1, \zeta_2, \alpha_L), \tag{C.6}$$

where $\eta(d, n, \zeta_1, \zeta_2, \alpha_L) = O\left(\frac{(\log d)^{19/6}}{n^{1/6}} + \frac{(\log d)^{11/6}}{n^{1/6} \alpha_L^{1/3}} + \zeta_1 \log d + \frac{\zeta_2}{\alpha_L} \right)$.

C.2 Proof of Proposition C.1

Before proving Proposition C.1, we present Lemma C.1. It bounds the conditional quantile $q_\xi(\alpha; T_{\boldsymbol{W}})$ in terms of the quantile $q(\alpha; T_{\boldsymbol{Z}})$ of the Gaussian maxima $T_{\boldsymbol{Z}}$ when the maximal covariance matrix differences are controlled. In the proof of Lemma C.1, we apply the Cramér-type comparison bound (3.1), which is establised in Theorem 3.1. To simplify the notation, we denote the bound $C_1(\log d)^{5/2}\Delta_\infty^{1/2}$ in (3.1) by $\pi(\Delta_\infty)$, where the constant C_1 only depends on $\min_{1\le j\le d}\{\sigma_{jj}^U, \sigma_{jj}^V\}$, $\max_{1\le j\le d}\{\sigma_{jj}^U, \sigma_{jj}^V\}$.

Lemma C.1. Suppose δ satisfies $(\log d)^5\delta = O(1)$. On the event $\{\Delta_\infty \le \delta\}$, we have

$$q_\xi(\alpha; T_{\boldsymbol{W}}) \ge q\Big(\frac{\alpha}{1-\pi(\delta)}; T_{\boldsymbol{Z}}\Big), \tag{C.7}$$

$$q_\xi(\alpha; T_{\boldsymbol{W}}) \le q\Big(\frac{\alpha}{1+\pi(\delta)}; T_{\boldsymbol{Z}}\Big). \tag{C.8}$$

Proof of Lemma C.1. On the event $\{\Delta_\infty \le \delta\}$, we have $(\log d)^5\Delta_\infty \le (\log d)^5\delta = O(1)$, then by applying Theorem 3.1 to $\boldsymbol{Z}$ and $\boldsymbol{W}$, we obtain the following,

$$\sup_{0\le t\le C_0\sqrt{\log d}}\left|\frac{\mathbb{P}_\xi(T_{\boldsymbol{W}} > t)}{\mathbb{P}(T_{\boldsymbol{Z}} > t)} - 1\right| \le \pi(\delta).$$

Therefore we have

$$\mathbb{P}_\xi\left(T_{\boldsymbol{W}} \ge q\Big(\frac{\alpha}{1-\pi(\delta)}; T_{\boldsymbol{Z}}\Big)\right) \ge \mathbb{P}\left(T_{\boldsymbol{Z}} \ge q\Big(\frac{\alpha}{1-\pi(\delta)}; T_{\boldsymbol{Z}}\Big)\right)\cdot(1-\pi(\delta)) = \alpha,$$

when t satisfies $0 \le t \le C_0\sqrt{\log d}$. Then $q_\xi(\alpha; T_{\boldsymbol{W}}) \ge q\Big(\frac{\alpha}{1-\pi(\delta)}; T_{\boldsymbol{Z}}\Big)$ immediately follows, i.e., (C.7) holds. Similarly, on the event $\{\Delta_\infty \le \delta\}$, we have

$$\mathbb{P}_\xi\left(T_{\boldsymbol{W}} \ge q\Big(\frac{\alpha}{1+\pi(\delta)}; T_{\boldsymbol{Z}}\Big)\right) \le \mathbb{P}\left(T_{\boldsymbol{Z}} \ge q\Big(\frac{\alpha}{1+\pi(\delta)}; T_{\boldsymbol{Z}}\Big)\right)\cdot(1+\pi(\delta)) = \alpha.$$

Thus $q_\xi(\alpha; T_{\boldsymbol{W}}) \le q\Big(\frac{\alpha}{1+\pi(\delta)}; T_{\boldsymbol{Z}}\Big)$, i.e., (C.8) holds. □

Proof of Proposition C.1. By the triangle inequality, we have

$$\left|\frac{\mathbb{P}(T_{\boldsymbol{Y}} > q_\xi(\alpha; T_{\boldsymbol{W}}))}{\mathbb{P}(T_{\boldsymbol{Z}} > q(\alpha; T_{\boldsymbol{Z}}))} - 1\right| \le \underbrace{\left|\frac{\mathbb{P}(T_{\boldsymbol{Y}} > q(\alpha; T_{\boldsymbol{Z}}))}{\mathbb{P}(T_{\boldsymbol{Z}} > q(\alpha; T_{\boldsymbol{Z}}))} - 1\right|}_{\text{I}} + \underbrace{\frac{|\mathbb{P}(T_{\boldsymbol{Y}} > q_\xi(\alpha; T_{\boldsymbol{W}})) - \mathbb{P}(T_{\boldsymbol{Y}} > q(\alpha; T_{\boldsymbol{Z}}))|}{\mathbb{P}(T_{\boldsymbol{Z}} > q(\alpha; T_{\boldsymbol{Z}}))}}_{\text{II}}. \tag{C.9}$$

Regarding the first term I, we will directly apply Corollary 5.1 in Kuchibhotla et al. (2021). Specifically, we verify the tail assumption on $\boldsymbol{Y}$ and the condition on the quantile that $q(\alpha; T_{\boldsymbol{Z}}) \le q(\alpha_L; T_{\boldsymbol{Z}}) = O\left(\sqrt{\log d}\right)$ when $\alpha \in [\alpha_L, 1]$. Then we obtain the following bound

$$\text{I} = \left|\frac{\mathbb{P}(T_{\boldsymbol{Y}} > q(\alpha; T_{\boldsymbol{Z}}))}{\mathbb{P}\left(T_{\boldsymbol{Z}} > q(\alpha; T_{\boldsymbol{Z}})\right)} - 1\right| = O\left(\frac{(\log d)^{19/6}}{n^{1/6}}\right). \tag{C.10}$$

Regarding the second term II, we write it as

$$\begin{aligned}
\mathrm{II} &= \frac{1}{\alpha}\left|\mathbb{P}(T_{\boldsymbol{Y}} > q_\xi(\alpha; T_{\boldsymbol{W}})) - \mathbb{P}(T_{\boldsymbol{Y}} > q(\alpha; T_{\boldsymbol{Z}}))\right| \\
&\leq \frac{1}{\alpha}\mathbb{P}(\{T_{\boldsymbol{Y}} > q_\xi(\alpha; T_{\boldsymbol{W}})\} \ominus \{T_{\boldsymbol{Y}} > q(\alpha; T_{\boldsymbol{Z}})\}) \\
&= \frac{1}{\alpha}\Big(\mathbb{P}(T_{\boldsymbol{Y}} > q_\xi(\alpha; T_{\boldsymbol{W}}), T_{\boldsymbol{Y}} \leq q(\alpha; T_{\boldsymbol{Z}})) + \mathbb{P}(T_{\boldsymbol{Y}} \leq q_\xi(\alpha; T_{\boldsymbol{W}}), T_{\boldsymbol{Y}} > q(\alpha; T_{\boldsymbol{Z}}))\Big) \\
&\leq \frac{1}{\alpha}\mathbb{P}(T_{\boldsymbol{Y}} > q_\xi(\alpha; T_{\boldsymbol{W}}), T_{\boldsymbol{Y}} \leq q(\alpha; T_{\boldsymbol{Z}}), \Delta_\infty \leq \delta) \\
&\quad + \frac{1}{\alpha}\mathbb{P}(T_{\boldsymbol{Y}} \leq q_\xi(\alpha; T_{\boldsymbol{W}}), T_{\boldsymbol{Y}} > q(\alpha; T_{\boldsymbol{Z}}), \Delta_\infty \leq \delta) + \frac{2\mathbb{P}(\Delta_\infty > \delta)}{\alpha},
\end{aligned}$$

where the first inequality holds by the definition of the symmetric difference; recall the symmetric difference between A and B is defined as $A \ominus B = (A \setminus B) \cup (B \setminus A))$. Remark that we will give the explicit choice of δ later in the proof. Now we apply Lemma C.1 (whose condition will be verified in (C.16)) and further bound II as,

$$\begin{aligned}
\mathrm{II} &\leq \frac{1}{\alpha}\Big(\mathbb{P}(T_{\boldsymbol{Y}} \geq q(\frac{\alpha}{1-\pi(\delta)}; T_{\boldsymbol{Z}}), T_{\boldsymbol{Y}} \leq q(\alpha; T_{\boldsymbol{Z}})) \\
&\qquad + \mathbb{P}(T_{\boldsymbol{Y}} \leq q(\frac{\alpha}{1+\pi(\delta)}; T_{\boldsymbol{Z}}), T_{\boldsymbol{Y}} > q(\alpha; T_{\boldsymbol{Z}}))\Big) + \frac{2\mathbb{P}(\Delta_\infty > \delta)}{\alpha} \\
&= \frac{1}{\alpha}\,\mathbb{P}\left(q(\frac{\alpha}{1-\pi(\delta)}; T_{\boldsymbol{Z}}) \leq T_{\boldsymbol{Y}} \leq q(\frac{\alpha}{1+\pi(\delta)}; T_{\boldsymbol{Z}})\right) + \frac{2\mathbb{P}(\Delta_\infty > \delta)}{\alpha} \qquad \text{(C.11)} \\
&\leq \frac{1}{\alpha}\,\mathbb{P}\left(q(\frac{\alpha}{1-\pi(\delta)}; T_{\boldsymbol{Z}}) \leq T_{\boldsymbol{Z}} \leq q(\frac{\alpha}{1+\pi(\delta)}; T_{\boldsymbol{Z}})\right) + \frac{2\mathbb{P}(\Delta_\infty > \delta)}{\alpha} + \mathrm{III} \\
&= \frac{2\pi(\delta)}{1-\pi^2(\delta)} + \frac{2\mathbb{P}(\Delta_\infty > \delta)}{\alpha} + \mathrm{III}, \qquad \text{(C.12)}
\end{aligned}$$

where the term III in the second inequality is defined as,

$$\mathrm{III} := \frac{1}{\alpha}\left|\mathbb{P}\left(q(\frac{\alpha}{1-\pi(\delta)}; T_{\boldsymbol{Z}}) \leq T_{\boldsymbol{Y}} \leq q(\frac{\alpha}{1+\pi(\delta)}; T_{\boldsymbol{Z}})\right) - \mathbb{P}\left(q(\frac{\alpha}{1-\pi(\delta)}; T_{\boldsymbol{Z}}) \leq T_{\boldsymbol{Z}} \leq q(\frac{\alpha}{1+\pi(\delta)}; T_{\boldsymbol{Z}})\right)\right|.$$

Below we further rewrite III as

$$\mathrm{III} = \frac{1}{\alpha}\left|\frac{\alpha}{1-\pi(\delta)} \cdot \mathrm{III}_1 - \frac{\alpha}{1+\pi(\delta)} \cdot \mathrm{III}_2\right|,$$

with $\mathrm{III}_1, \mathrm{III}_2$ defined as

$$\begin{aligned}
\mathrm{III}_1 &= \frac{\mathbb{P}\left(T_{\boldsymbol{Y}} > q(\frac{\alpha}{1-\pi(\delta)}; T_{\boldsymbol{Z}})\right) - \mathbb{P}\left(T_{\boldsymbol{Z}} > q(\frac{\alpha}{1-\pi(\delta)}; T_{\boldsymbol{Z}})\right)}{\mathbb{P}\left(T_{\boldsymbol{Z}} > q(\frac{\alpha}{1-\pi(\delta)}; T_{\boldsymbol{Z}})\right)}, \\
\mathrm{III}_2 &= \frac{\mathbb{P}\left(T_{\boldsymbol{Y}} > q(\frac{\alpha}{1+\pi(\delta)}; T_{\boldsymbol{Z}})\right) - \mathbb{P}\left(T_{\boldsymbol{Z}} > q(\frac{\alpha}{1+\pi(\delta)}; T_{\boldsymbol{Z}})\right)}{\mathbb{P}\left(T_{\boldsymbol{Z}} > q(\frac{\alpha}{1+\pi(\delta)}; T_{\boldsymbol{Z}})\right)}.
\end{aligned}$$

Thus by applying Corollary 5.1 of Kuchibhotla et al. (2021) to $\mathrm{III}_1, \mathrm{III}_2$ similarly as in (C.10), we have the following bound on III,

$$\mathrm{III} = O\left(\frac{(\log d)^{19/6}}{n^{1/6}}\right). \qquad \text{(C.13)}$$

Combining (C.12) and (C.13) yields the following bound,

$$\begin{aligned}
\text{II} &\le \frac{1}{\alpha}\mathbb{P}(\{T_{\boldsymbol{Y}} > q_\xi(\alpha; T_{\boldsymbol{W}})\} \ominus \{T_{\boldsymbol{Y}} > q(\alpha; T_{\boldsymbol{Z}})\}) \\
&\le \frac{C(\log d)^{19/6}}{n^{1/6}} + C_0'\pi(\delta) + \frac{C''\,\mathbb{P}\left(\Delta_\infty > \delta\right)}{\alpha} \\
&\le \frac{C(\log d)^{19/6}}{n^{1/6}} + C'(\log d)^{5/2}\delta^{1/2} + \frac{C''\,\mathbb{E}\left[\Delta_\infty\right]}{\delta\alpha} \\
&= O\left(\left(\frac{\mathbb{E}\left[\Delta_\infty\right]\log^5 d}{\alpha}\right)^{1/3} + \frac{(\log d)^{19/6}}{n^{1/6}}\right),
\end{aligned} \tag{C.14}$$

where the second inequality holds due to the definition of $\pi(\delta)$ and Markov's inequality, the last line holds by choosing δ to be $(\mathbb{E}\left[\Delta_\infty\right])^{2/3}/(\alpha^{1/3}(\log d)^{5/3})$. We will bound the term $\mathbb{E}\left[\Delta_\infty\right]$ using Lemma C.1 in Chernozhukov et al. (2013). Specifically, under the stated tail assumption on $\boldsymbol{Y}$, the condition (E.1) of Lemma C.1 in Chernozhukov et al. (2013) is satisfied; see Comment 2.2 in Chernozhukov et al. (2013). Thus we have

$$\mathbb{E}\left[\Delta_\infty\right] \le \sqrt{\frac{B_n^2 \log d}{n}} \vee \frac{B_n^2(\log(dn))^2(\log d)}{n}, \tag{C.15}$$

where B_n equals some constant C which does not depend on n. As promised previously, we verify the assumption of Lemma C.1 for our choice of δ. Specifically, for $\delta = (\mathbb{E}\left[\Delta_\infty\right])^{2/3}/(\alpha^{1/3}(\log d)^{5/3})$, we have $(\log d)^5\delta$ satisfies the following

$$(\log d)^5\delta \le \frac{(\log d)^5(\mathbb{E}\left[\Delta_\infty\right])^{2/3}}{\alpha_L^{1/3}(\log d)^{5/3}} = \left(\frac{\log^{11} d}{n\alpha_L}\right)^{1/3} = O(1), \tag{C.16}$$

under the stated condition on α_L. Finally, when $\alpha \in [\alpha_L, 1]$, we combine (C.9), (C.10), (C.14) with (C.15), then establish (C.3), i.e.,

$$\sup_{\alpha\in[\alpha_L,1]}\left|\frac{\mathbb{P}(T_{\boldsymbol{Y}} > q_\xi(\alpha; T_{\boldsymbol{W}}))}{\mathbb{P}(T_{\boldsymbol{Z}} > q(\alpha; T_{\boldsymbol{Z}}))} - 1\right| = O\left(\frac{(\log d)^{11/6}}{n^{1/6}\alpha_L^{1/3}} + \frac{(\log d)^{19/6}}{n^{1/6}}\right).$$

□

C.3 Proof of Proposition C.2

Before proving Proposition C.2, we need to present a simple lemma. It translates the approximation error ζ_1, ζ_2 into the bounds on the quantiles. And its proof is quite straightforward thus omitted.

Lemma C.2. Under the assumption in (C.5), we have, for $\alpha \in (0, 1)$,

$$\mathbb{P}(q_\xi(\alpha; T^{\mathcal{B}}) \le q_\xi(\alpha + \zeta_2; T_{\boldsymbol{W}}) + \zeta_1) \ge 1 - \zeta_2,$$
$$\mathbb{P}(q_\xi(\alpha; T^{\mathcal{B}}) \ge q_\xi(\alpha - \zeta_2; T_{\boldsymbol{W}}) - \zeta_1) \ge 1 - \zeta_2.$$

Proof of Proposition C.2. By the triangle inequality, we have

$$\left|\frac{\mathbb{P}(T > q_\xi(\alpha; T^{\mathcal{B}}))}{\mathbb{P}(T_{\boldsymbol{Z}} > q(\alpha; T_{\boldsymbol{Z}}))} - 1\right| \le \underbrace{\left|\frac{\mathbb{P}(T_{\boldsymbol{Y}} > q(\alpha; T_{\boldsymbol{Z}}))}{\mathbb{P}(T_{\boldsymbol{Z}} > q(\alpha; T_{\boldsymbol{Z}}))} - 1\right|}_{\text{I}} + \underbrace{\frac{|\mathbb{P}(T > q_\xi(\alpha; T^{\mathcal{B}})) - \mathbb{P}(T_{\boldsymbol{Y}} > q(\alpha; T_{\boldsymbol{Z}}))|}{\mathbb{P}(T_{\boldsymbol{Z}} > q(\alpha; T_{\boldsymbol{Z}}))}}_{\text{II}}. \tag{C.17}$$

Note that (C.10) in the proof of Proposition C.1 immediately gives the bound on I, i.e.,

$$\mathrm{I} = O\Big(\frac{(\log d)^{19/6}}{n^{1/6}}\Big). \tag{C.18}$$

Regarding the term II, we have

$$\begin{aligned}
\mathrm{II} &= \frac{1}{\alpha}\left|\mathbb{P}(T > q_\xi(\alpha; T^{\mathcal{B}})) - \mathbb{P}(T_{\boldsymbol{Y}} > q(\alpha; T_{\boldsymbol{Z}}))\right| \\
&\le \frac{1}{\alpha}\left|\mathbb{P}(\{T > q_\xi(\alpha; T^{\mathcal{B}})\} \ominus \{T_{\boldsymbol{Y}} > q(\alpha; T_{\boldsymbol{Z}})\})\right| \\
&= \frac{1}{\alpha}\mathbb{P}(T > q_\xi(\alpha; T^{\mathcal{B}}), T_{\boldsymbol{Y}} \le q(\alpha; T_{\boldsymbol{Z}})) + \frac{1}{\alpha}\mathbb{P}(T \le q_\xi(\alpha; T^{\mathcal{B}}), T_{\boldsymbol{Y}} > q(\alpha; T_{\boldsymbol{Z}})).
\end{aligned} \tag{C.19}$$

To bound the two terms in (C.19), first notice that on the event $|T - T_{\boldsymbol{Y}}| > \zeta_1$, we have

$$\{T > q_\xi(\alpha; T^{\mathcal{B}}), T_{\boldsymbol{Y}} \le q(\alpha; T_{\boldsymbol{Z}})\} \subset \{T_{\boldsymbol{Y}} > q_\xi(\alpha; T^{\mathcal{B}}) - \zeta_1, T_{\boldsymbol{Y}} \le q(\alpha; T_{\boldsymbol{Z}})\}.$$

Then under the assumption in (C.4), i.e., $\mathbb{P}(|T - T_{\boldsymbol{Y}}| > \zeta_1) < \zeta_2$, we obtain

$$\mathbb{P}(T > q_\xi(\alpha; T^{\mathcal{B}}), T_{\boldsymbol{Y}} \le q(\alpha; T_{\boldsymbol{Z}})) \le \mathbb{P}(T_{\boldsymbol{Y}} > q_\xi(\alpha; T^{\mathcal{B}}) - \zeta_1, T_{\boldsymbol{Y}} \le q(\alpha; T_{\boldsymbol{Z}})) + \zeta_2.$$

Applying such strategies to the second term in (C.19) similarly, we get the following,

$$\begin{aligned}
\mathrm{II} &\le \mathrm{II}_1 + \mathrm{II}_2 + \frac{2\zeta_2}{\alpha}, \quad \text{where} \qquad (C.20)\\
\mathrm{II}_1 &:= \frac{1}{\alpha}\mathbb{P}(T_{\boldsymbol{Y}} > q_\xi(\alpha; T^{\mathcal{B}}) - \zeta_1, T_{\boldsymbol{Y}} \le q(\alpha; T_{\boldsymbol{Z}})), \\
\mathrm{II}_2 &:= \frac{1}{\alpha}\mathbb{P}(T_{\boldsymbol{Y}} \le q_\xi(\alpha; T^{\mathcal{B}}) + \zeta_2, T_{\boldsymbol{Y}} > q(\alpha; T_{\boldsymbol{Z}})).
\end{aligned}$$

Under the assumption (C.5), by Lemma C.2, we have

$$\begin{aligned}
\mathbb{P}(q_\xi(\alpha; T^{\mathcal{B}}) \le q_\xi(\alpha + \zeta_2; T_{\boldsymbol{W}}) + \zeta_1) &\ge 1 - \zeta_2, \\
\mathbb{P}(q_\xi(\alpha; T^{\mathcal{B}}) \ge q_\xi(\alpha - \zeta_2; T_{\boldsymbol{W}}) - \zeta_1) &\ge 1 - \zeta_2.
\end{aligned}$$

Hence we can bound $\mathrm{II}_1, \mathrm{II}_2$ as below,

$$\begin{aligned}
\mathrm{II}_1 &\le \frac{1}{\alpha}\mathbb{P}\left(T_{\boldsymbol{Y}} > q_\xi(\alpha - \zeta_2; T_{\boldsymbol{W}}) - 2\zeta_1, T_{\boldsymbol{Y}} \le q(\alpha; T_{\boldsymbol{Z}})\right) + \frac{\zeta_2}{\alpha}, \\
\mathrm{II}_2 &\le \frac{1}{\alpha}\mathbb{P}\left(T_{\boldsymbol{Y}} \le q_\xi(\alpha + \zeta_2; T_{\boldsymbol{W}}) + 2\zeta_1, T_{\boldsymbol{Y}} > q(\alpha; T_{\boldsymbol{Z}})\right) + \frac{\zeta_2}{\alpha}.
\end{aligned}$$

Now we will use the strategy of deriving (C.11) in the proof of Proposition C.1, i.e., apply Lemma C.1, then we have,

$$\begin{aligned}
\mathrm{II}_1 &\le \frac{1}{\alpha}\mathbb{P}\left(T_{\boldsymbol{Y}} > q\Big(\frac{\alpha - \zeta_2}{1 - \pi(\delta)}; T_{\boldsymbol{Z}}\Big) - 2\zeta_1, T_{\boldsymbol{Y}} \le q(\alpha; T_{\boldsymbol{Z}})\right) + \frac{\mathbb{P}\left(\Delta_\infty > \delta\right)}{\alpha} + \frac{\zeta_2}{\alpha}, \\
\mathrm{II}_2 &\le \frac{1}{\alpha}\mathbb{P}\left(T_{\boldsymbol{Y}} \le q\Big(\frac{\alpha + \zeta_2}{1 + \pi(\delta)}; T_{\boldsymbol{Z}}\Big) + 2\zeta_1, T_{\boldsymbol{Y}} > q(\alpha; T_{\boldsymbol{Z}})\right) + \frac{\mathbb{P}\left(\Delta_\infty > \delta\right)}{\alpha} + \frac{\zeta_2}{\alpha}.
\end{aligned}$$

Combining the above two inequalities with (C.20), we have

$$\mathrm{II} \le \mathrm{III} + \frac{2\mathbb{P}\left(\Delta_\infty > \delta\right)}{\alpha} + \frac{4\zeta_2}{\alpha}, \tag{C.21}$$

where III is defined as below,

$$
\begin{aligned}
\mathrm{III} := \ & \frac{1}{\alpha}\left|\mathbb{P}\left(T_{\boldsymbol{Y}} > q\Big(\frac{\alpha-\zeta_2}{1-\pi(\delta)};T_{\boldsymbol{Z}}\Big)-2\zeta_1\right)-\mathbb{P}\left(T_{\boldsymbol{Y}} > q\Big(\frac{\alpha+\zeta_2}{1+\pi(\delta)};T_{\boldsymbol{Z}}\Big)+2\zeta_1\right)\right| \\
= \ & \frac{1}{\alpha}\Bigg|\mathbb{P}\left(T_{\boldsymbol{Y}} > q\Big(\frac{\alpha-\zeta_2}{1-\pi(\delta)};T_{\boldsymbol{Z}}\Big)-2\zeta_1\right)-\mathbb{P}\left(T_{\boldsymbol{Z}} > q\Big(\frac{\alpha-\zeta_2}{1-\pi(\delta)};T_{\boldsymbol{Z}}\Big)-2\zeta_1\right) \\
& -\mathbb{P}\left(T_{\boldsymbol{Y}} > q\Big(\frac{\alpha+\zeta_2}{1+\pi(\delta)};T_{\boldsymbol{Z}}\Big)+2\zeta_1\right)+\mathbb{P}\left(T_{\boldsymbol{Z}} > q\Big(\frac{\alpha+\zeta_2}{1+\pi(\delta)};T_{\boldsymbol{Z}}\Big)+2\zeta_1\right) \\
& +\mathbb{P}\left(T_{\boldsymbol{Z}} > q\Big(\frac{\alpha-\zeta_2}{1-\pi(\delta)};T_{\boldsymbol{Z}}\Big)-2\zeta_1\right)-\mathbb{P}\left(T_{\boldsymbol{Z}} > q\Big(\frac{\alpha+\zeta_2}{1+\pi(\delta)};T_{\boldsymbol{Z}}\Big)+2\zeta_1\right)\Bigg| \\
\leq \ & \mathrm{III}_1+\mathrm{III}_2+\mathrm{III}_3.
\end{aligned}
$$

The last line comes from the triangle inequality, with $\mathrm{III}_1, \mathrm{III}_2, \mathrm{III}_3$ defined as,

$$
\begin{aligned}
\mathrm{III}_1 &:= \frac{1}{\alpha}\left|\mathbb{P}\Big(T_{\boldsymbol{Y}} > q\Big(\frac{\alpha+\zeta_2}{1+\pi(\delta)};T_{\boldsymbol{Z}}\Big)+2\zeta_1\Big)-\mathbb{P}\Big(T_{\boldsymbol{Z}} > q\Big(\frac{\alpha+\zeta_2}{1+\pi(\delta)};T_{\boldsymbol{Z}}\Big)+2\zeta_1\Big)\right|, \\
\mathrm{III}_2 &:= \frac{1}{\alpha}\left|\mathbb{P}\Big(T_{\boldsymbol{Y}} > q\Big(\frac{\alpha-\zeta_2}{1-\pi(\delta)};T_{\boldsymbol{Z}}\Big)-2\zeta_1\Big)-\mathbb{P}\Big(T_{\boldsymbol{Z}} > q\Big(\frac{\alpha-\zeta_2}{1-\pi(\delta)};T_{\boldsymbol{Z}}\Big)-2\zeta_1\Big)\right|, \\
\mathrm{III}_3 &:= \frac{1}{\alpha}\left|\mathbb{P}\Big(T_{\boldsymbol{Z}} > q\Big(\frac{\alpha-\zeta_2}{1-\pi(\delta)};T_{\boldsymbol{Z}}\Big)-2\zeta_1\Big)-\mathbb{P}\Big(T_{\boldsymbol{Z}} > q\Big(\frac{\alpha+\zeta_2}{1+\pi(\delta)};T_{\boldsymbol{Z}}\Big)+2\zeta_1\Big)\right|.
\end{aligned}
$$

We first bound III_3 by the triangle inequality,

$$
\begin{aligned}
\mathrm{III}_3 &= \frac{1}{\alpha}\left|\mathbb{P}\Big(T_{\boldsymbol{Z}} > q\Big(\frac{\alpha-\zeta_2}{1-\pi(\delta)};T_{\boldsymbol{Z}}\Big)-2\zeta_1\Big)-\mathbb{P}\Big(T_{\boldsymbol{Z}} > q\Big(\frac{\alpha+\zeta_2}{1+\pi(\delta)};T_{\boldsymbol{Z}}\Big)+2\zeta_1\Big)\right| \\
&\leq \underbrace{\frac{1}{\alpha}\left|\mathbb{P}\Big(T_{\boldsymbol{Z}} > q\Big(\frac{\alpha+\zeta_2}{1+\pi(\delta)};T_{\boldsymbol{Z}}\Big)+2\zeta_1\Big)-\frac{\alpha+\zeta_2}{1+\pi(\delta)}\right|}_{\mathrm{III}_{31}} \\
&+ \underbrace{\frac{1}{\alpha}\left|\mathbb{P}\Big(T_{\boldsymbol{Z}} > q\Big(\frac{\alpha-\zeta_2}{1-\pi(\delta)};T_{\boldsymbol{Z}}\Big)-2\zeta_1\Big)-\frac{\alpha-\zeta_2}{1-\pi(\delta)}\right|}_{\mathrm{III}_{32}}+\underbrace{\frac{1}{\alpha}\left|\frac{\alpha-\zeta_2}{1-\pi(\delta)}-\frac{\alpha+\zeta_2}{1+\pi(\delta)}\right|}_{\mathrm{III}_{33}}.
\end{aligned}
$$

Note that III_{31} can be rewritten as

$$
\begin{aligned}
\mathrm{III}_{31} &= \frac{\alpha+\zeta_2}{\alpha(1+\pi(\delta))}\cdot\frac{\left|\mathbb{P}\Big(T_{\boldsymbol{Z}} > q\big(\frac{\alpha+\zeta_2}{1+\pi(\delta)};T_{\boldsymbol{Z}}\big)+2\zeta_1\Big)-\mathbb{P}\left(T_{\boldsymbol{Z}} > q\big(\frac{\alpha+\zeta_2}{1+\pi(\delta)};T_{\boldsymbol{Z}}\big)\right)\right|}{\mathbb{P}\left(T_{\boldsymbol{Z}} > q\big(\frac{\alpha+\zeta_2}{1+\pi(\delta)};T_{\boldsymbol{Z}}\big)\right)} && \text{(C.22)} \\
&\leq \frac{\alpha+\zeta_2}{\alpha(1+\pi(\delta))}\cdot K_4\zeta_1\Big(q\big(\frac{\alpha+\zeta_2}{1+\pi(\delta)};T_{\boldsymbol{Z}}\big)+\zeta_1\Big)\leq C\zeta_1\log d, && \text{(C.23)}
\end{aligned}
$$

where the first inequality holds by applying a non-uniform anti-concentration bound. Specifically, we apply the part 3 of Theorem 2.1 in Kuchibhotla et al. (2021) (with $r-\epsilon = q\big(\frac{\alpha+\zeta_2}{1+\pi(\delta)};T_{\boldsymbol{Z}}\big), r+\epsilon = q\big(\frac{\alpha+\zeta_2}{1+\pi(\delta)};T_{\boldsymbol{Z}}\big)+2\zeta_1$) to the Gaussian random vector $\boldsymbol{Z}$. Remark that the term K_3 is a constant only depending on $\min_{1\leq j\leq d}\{\sigma_{jj}^Y\}, \max_{1\leq j\leq d}\{\sigma_{jj}^Y\}$ and the median of Gaussian maxima (up to 2-nd power, hence at most of rate $O(\log d)$). As for the second inequality, under the assumption $\zeta_2 = O(\alpha_L)$, we have $\frac{\zeta_2}{\alpha}\leq\frac{\zeta_2}{\alpha_L} = O(1)$ when $\alpha\in[\alpha_L, 1]$; we also use the fact that $\zeta_1 = O(\sqrt{\log d})$ (which holds under the stated assumption), and $q\big(\frac{\alpha+\zeta_2}{1+\pi(\delta)};T_{\boldsymbol{Z}}\big) = O(\sqrt{\log d})$ (which will be verified later in (C.27)). Thus we show $\mathrm{III}_{31} = O(\zeta_1 \log d)$. Similarly, III_{32} can be bounded as $O(\zeta_1 \log d)$. As for III_{33}, we have

$$\mathrm{III}_{33} = \frac{1}{\alpha}\Big|\frac{\alpha-\zeta_2}{1-\pi(\delta)} - \frac{\alpha+\zeta_2}{1+\pi(\delta)}\Big| \le \frac{2\pi(\delta)}{1-\pi^2(\delta)} + \frac{2\zeta_2}{\alpha(1-\pi^2(\delta))}.$$

Thus by combining the bounds on $\mathrm{III}_{31}, \mathrm{III}_{32}, \mathrm{III}_{33}$, we obtain

$$\mathrm{III}_3 \le \mathrm{III}_{31} + \mathrm{III}_{32} + \mathrm{III}_{33} \le C'\zeta_1 \log d + \frac{2\pi(\delta)}{1-\pi^2(\delta)} + \frac{2\zeta_2}{\alpha(1-\pi^2(\delta))}. \tag{C.24}$$

Regarding the term III_1, we first consider the following,

$$\begin{aligned}
\mathrm{III}_{11} &:= \frac{1}{\alpha}\mathbb{P}\Big(T_{\boldsymbol{Z}} > q(\frac{\alpha+\zeta_2}{1+\pi(\delta)}; T_{\boldsymbol{Z}}) + 2\zeta_1\Big) \\
&\le \frac{1}{\alpha}\mathbb{P}\Big(T_{\boldsymbol{Z}} > q(\frac{\alpha+\zeta_2}{1+\pi(\delta)}; T_{\boldsymbol{Z}})\Big) \cdot \Big(1 + K_4\zeta_1\Big(q(\frac{\alpha+\zeta_2}{1+\pi(\delta)}; T_{\boldsymbol{Z}}) + \zeta_1\Big)\Big) \\
&= \frac{\alpha+\zeta_2}{\alpha(1+\pi(\delta))} \cdot \Big(1 + K_4\zeta_1\Big(q(\frac{\alpha+\zeta_2}{1+\pi(\delta)}; T_{\boldsymbol{Z}}) + \zeta_1\Big)\Big) \\
&\le C'' + C\zeta_1 \log d = O(1),
\end{aligned}$$

where the first inequality holds due to the derivations from (C.22) to (C.23), the second inequality holds due to the last inequality in (C.23) and the stated assumption $\zeta_2 = O(\alpha_L)$. Then we bound III_1 in terms of III_{11} and write

$$\begin{aligned}
\mathrm{III}_1 &= \frac{1}{\alpha}\Big|\mathbb{P}\Big(T_{\boldsymbol{Y}} > q\Big(\frac{\alpha+\zeta_2}{1+\pi(\delta)}; T_{\boldsymbol{Z}}\Big) + 2\zeta_1\Big) - \mathbb{P}\Big(T_{\boldsymbol{Z}} > q\Big(\frac{\alpha+\zeta_2}{1+\pi(\delta)}; T_{\boldsymbol{Z}}\Big) + 2\zeta_1\Big)\Big| \\
&= \mathrm{III}_{11} \cdot \left|\frac{\mathbb{P}\Big(T_{\boldsymbol{Y}} > q\Big(\frac{\alpha+\zeta_2}{1+\pi(\delta)}; T_{\boldsymbol{Z}}\Big) + 2\zeta_1\Big) - \mathbb{P}\Big(T_{\boldsymbol{Z}} > q\Big(\frac{\alpha+\zeta_2}{1+\pi(\delta)}; T_{\boldsymbol{Z}}\Big) + 2\zeta_1\Big)}{\mathbb{P}\Big(T_{\boldsymbol{Z}} > q\Big(\frac{\alpha+\zeta_2}{1+\pi(\delta)}; T_{\boldsymbol{Z}}\Big) + 2\zeta_1\Big)}\right| \\
&\le \mathrm{III}_{11} \cdot \frac{(\log d)^{19/6}}{n^{1/6}} = O\Big(\frac{(\log d)^{19/6}}{n^{1/6}}\Big),
\end{aligned}$$

where the inequality holds by applying Corollary 5.1 in Kuchibhotla et al. (2021) again to $T_{\boldsymbol{Y}}$ as the derivations of (C.10) in the proof of Proposition C.1. The term III_2 can be similarly bounded as III_1. Combining the above bounds on $\mathrm{III}_1, \mathrm{III}_2$ and (C.24) yields the following bound on III,

$$\mathrm{III} \le \frac{C(\log d)^{19/6}}{n^{1/6}} + C'\zeta_1 \log d + \frac{2\pi(\delta)}{1-\pi^2(\delta)} + \frac{2\zeta_2}{\alpha(1-\pi^2(\delta))}. \tag{C.25}$$

By (C.17), (C.18), (C.21) and (C.25), we have, when $\alpha \in [\alpha_L, 1]$,

$$\begin{aligned}
\left|\frac{\mathbb{P}(T > q_\xi(\alpha; T^{\mathcal{B}}))}{\mathbb{P}(T_{\boldsymbol{Z}} > q(\alpha; T_{\boldsymbol{Z}}))} - 1\right| &\le \mathrm{I} + \mathrm{II} \le \mathrm{I} + \mathrm{III} + \frac{2\mathbb{P}(\Delta_\infty > \delta)}{\alpha} + \frac{4\zeta_2}{\alpha} \\
&\le \frac{C(\log d)^{19/6}}{n^{1/6}} + C'\zeta_1 \log d + \frac{C''\zeta_2}{\alpha} + \frac{2\pi(\delta)}{1-\pi^2(\delta)} + \frac{2\mathbb{P}(\Delta_\infty > \delta)}{\alpha} \\
&\le \frac{C(\log d)^{19/6}}{n^{1/6}} + C'\zeta_1 \log d + \frac{C''\zeta_2}{\alpha} + \frac{C(\log d)^{11/6}}{n^{1/6}\alpha_L^{1/3}} \\
&= O\left(\frac{(\log d)^{19/6}}{n^{1/6}} + \frac{(\log d)^{11/6}}{n^{1/6}\alpha_L^{1/3}} + \zeta_1 \log d + \frac{\zeta_2}{\alpha_L}\right).
\end{aligned} \tag{C.26}$$

where the third line holds due to the derivations between (C.13) and (C.16) in the proof of Proposition C.1. Remark by the choice of δ and (C.16), we have $\pi(\delta) = O(1)$. Also note that $\zeta_2 = O(\alpha_L)$, hence we can show

$$q\big(\frac{\alpha + \zeta_2}{1 + \pi(\delta)}; T_{\boldsymbol{Z}}\big) = O(\sqrt{\log d}). \tag{C.27}$$

when $\alpha \in [\alpha_L, 1]$. Hence we are able to verify $q\big(\frac{\alpha+\zeta_2}{1+\pi(\delta)}; T_{\boldsymbol{Z}}\big) = O(\sqrt{\log d})$, as promised when deriving (C.23). Denoting the bound in (C.26) by $\eta(d, n, \zeta_1, \zeta_2, \alpha_L)$, we finally establish (C.6), i.e.,

$$\sup_{\alpha \in [\alpha_L, 1]} \left| \frac{\mathbb{P}(T > q_\xi(\alpha; T^{\mathcal{B}}))}{\mathbb{P}(T_{\boldsymbol{Z}} > q(\alpha; T_{\boldsymbol{Z}}))} - 1 \right| = \eta(d, n, \zeta_1, \zeta_2, \alpha_L).$$

□

D Validity and power analysis of single node testing

In this section, we focus on Lemma A.1 and Lemma C.2. Note that these results are established using the same strategies as Theorem 4.1, Lemma S.1 and Theorem S.7 in Lu et al. (2017). We still present their proofs for completeness.

D.1 Proof of Lemma A.1

Proof. For given node j, we denote $N_{0j} = \{(j, k) : \boldsymbol{\Theta}_{jk} = 0\}$, then $N_{0j}^c = \{(j, k) : |\boldsymbol{\Theta}_{jk}| > 0\}$. First we consider the following event,

$$\mathcal{E} = \Big\{ \min_{e \in N_{0j}^c} \sqrt{n} |\widetilde{\boldsymbol{\Theta}}_e^{\mathrm{d}}| > \widehat{c}(\alpha, E_0) \Big\}, \qquad \text{where } E_0 = \{(j, k) : k \neq j, k \in [d]\}.$$

By the definition of Algorithm 1, we immediately have the rejected edge set in the first iteration can be written as

$$E_1 = \{(j, k) \in E_0 : \sqrt{n} |\widetilde{\boldsymbol{\Theta}}_{jk}^{\mathrm{d}}| > \widehat{c}(\alpha, E_0)\}.$$

Regarding (i) i.e., under the alternative hypothesis $H_{1j} : \|\boldsymbol{\Theta}_{j,-j}\|_0 \geq k_\tau$, we first note $\psi_{j,\alpha} = 1$ on the event $\mathcal{E}$. Also notice that $N_{0j}^c \subseteq E_1$ given $\mathcal{E}$. Then the following bound immediately follows:

$$\mathbb{P}\left(\psi_{j,\alpha} = 1\right) \geq \mathbb{P}\left(\mathcal{E}\right). \tag{D.1}$$

We further derive a lower bound for $\mathbb{P}(\mathcal{E})$ by the triangle inequality:

$$\mathbb{P}\left(\mathcal{E}\right) \geq \mathbb{P}\left(\min_{e \in N_{0j}^c} |\boldsymbol{\Theta}_e^\star| > \frac{\widehat{c}(\alpha, E_0)}{\sqrt{n}} + C_0 \sqrt{\frac{\log d}{n}} \text{ and } ||\widetilde{\boldsymbol{\Theta}}^{\mathrm{d}} - \boldsymbol{\Theta}^\star||_{\max} \leq C_0 \sqrt{\frac{\log d}{n}} \right). \tag{D.2}$$

For any fixed $\alpha \in (0, 1)$, we consider sufficiently large n and d such that $2/d^2 \leq \alpha/2$ and $\mathbb{P}(T_{E_0} > \widehat{c}(\alpha, E_0)) > \alpha/2$. This is possible due to the convergence result of the quantile approximation (D.27). Then, the above result together with (D.29) yields the following:

$$\widehat{c}(\alpha, E_0) \leq C_0 \sqrt{\frac{\log d}{n}} \cdot \sqrt{n} = C_0 \sqrt{\log d}.$$

Choosing the constant in the signal strength condition of Lemma A.1 to be $2C_0$ (i.e., for any $(j,k) \in N_{0j}^c$, $|\mathbf{\Theta}_{jk}| \geq 2C_0\sqrt{\log d/n}$) and applying (D.29), we have

$$\min_{e \in N_{0j}^c} |\mathbf{\Theta}_e^\star| \geq 2C_0\sqrt{\frac{\log d}{n}} \geq \frac{\widehat{c}(\alpha, E_0)}{\sqrt{n}} + C_0\sqrt{\frac{\log d}{n}},$$
$$\mathbb{P}\left(||\widetilde{\mathbf{\Theta}}^{\mathrm{d}} - \mathbf{\Theta}^\star||_{\max} \leq C_0\sqrt{\frac{\log d}{n}}\right) \geq 1 - 2/d.$$

Combining the above two inequalities with (D.1) and (D.2), we have $\mathbb{P}(\psi_{j,\alpha} = 1) \geq \mathbb{P}(\mathcal{E}) > 1 - 2/d$. Therefore, we establish

$$\lim_{(n,d)\to\infty} \mathbb{P}(\psi_{j,\alpha} = 1) = 1.$$

Now we consider (ii), i.e., the case when $\|\mathbf{\Theta}_{j,-j}\|_0 < k_\tau$. Since $\|\mathbf{\Theta}_{j,-j}\|_0 \leq k_\tau - 1$, $\psi_{j,\alpha} = 1$ implies at least one edge in N_{0j} is rejected in Algorithm 1. Suppose the first rejected edge in N_{0j} is (j, k_*) and it is rejected at the t_*-th iteration. Then we have $N_{0j} \subseteq E_{t_*-1}$ and

$$\max_{e \in N_{0j}} \sqrt{n}|\widetilde{\mathbf{\Theta}}_e^{\mathrm{d}} - \mathbf{\Theta}_e^\star| \geq \sqrt{n}|\widetilde{\mathbf{\Theta}}_{jk_*}^{\mathrm{d}} - \mathbf{\Theta}_{jk_*}^\star| \geq \widehat{c}(\alpha, E_{t_*-1}) \geq \widehat{c}(\alpha, N_{0j}), \tag{D.3}$$

where the first inequality holds since $(j, k_*) \subset N_{0j}$, the second inequality holds since $\mathbf{\Theta}_{jk_*}^\star = 0$ and the edge (j, k_*) is rejected at the t_*-th iteration. The last inequality holds simply because $N_{0j} \subseteq E_{t_*-1}$. Therefore by applying Lemma 2.1 with E chosen to be N_{0j}, we have

$$\lim_{(n,d)\to\infty} \mathbb{P}(\psi_{j,\alpha} = 1) \leq \alpha.$$

□

D.2 Proof of Lemma 2.1

We first recall the definition of $\mathcal{U}(M, s, r_0)$ and write down the statement of Lemma 2.1 below.

$$\mathcal{U}(M, s, r_0) = \Big\{\mathbf{\Theta} \in \mathbb{R}^{d\times d} \,\big|\, \lambda_{\min}(\mathbf{\Theta}) \geq 1/r_0, \lambda_{\max}(\mathbf{\Theta}) \leq r_0, \max_{j\in[d]} \|\mathbf{\Theta}_j\|_0 \leq s, \|\mathbf{\Theta}\|_1 \leq M\Big\}. \tag{D.4}$$

Lemma D.1. Suppose that $\mathbf{\Theta} \in \mathcal{U}(M, s, r_0)$. If $(\log(dn))^7/n + s^2(\log dn)^4/n = o(1)$, for any edge set $E \subseteq \mathcal{V} \times \mathcal{V}$, we have for any $\alpha \in [0, 1]$,

$$\lim_{(n,d)\to\infty} \sup_{\mathbf{\Theta}\in\mathcal{U}(M,s,r_0)} \sup_{\alpha\in(0,1)} \left|\mathbb{P}\left(\max_{e\in E} \sqrt{n}|\widetilde{\mathbf{\Theta}}_e^{\mathrm{d}} - \mathbf{\Theta}_e^\star| > \widehat{c}(\alpha, E)\right) - \alpha\right| = 0. \tag{D.5}$$

Throughout the following parts, we will write the standardized one-step estimator explicitly:

$$\widehat{\mathbf{\Theta}}_{jk}^{\mathrm{d}}/\sqrt{\widehat{\mathbf{\Theta}}_{jj}^{\mathrm{d}}\widehat{\mathbf{\Theta}}_{kk}^{\mathrm{d}}}, \quad \text{where } \widehat{\mathbf{\Theta}}_{jk}^{\mathrm{d}} := \widehat{\mathbf{\Theta}}_{jk} - \frac{\widehat{\mathbf{\Theta}}_j^\top\left(\widehat{\mathbf{\Sigma}}\widehat{\mathbf{\Theta}}_k - \mathbf{e}_k\right)}{\widehat{\mathbf{\Theta}}_j^\top\widehat{\mathbf{\Sigma}}_j}.$$

In order to prove (D.5), we need preliminary results on the estimation rates of CLIME estimator. Cai et al. (2011) gives the following theorem. We can also prove the same result for the GLasso estimator (Janková and van de Geer, 2018). Therefore, Lemma D.1 applies for both the CLIME estimator and the GLasso estimator. This also implies that the results in our paper apply to both the CLIME estimator and the GLasso estimator.

Lemma D.2. Suppose $\boldsymbol{\Theta} \in \mathcal{U}(M, s, r_0)$ and we choose the tuning parameter $\lambda \geq CM\sqrt{\log d/n}$ in the CLIME estimator. With probability greater than $1 - c/d^2$, we have the following bounds:

$$||\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}||_{\max} \leq C\sqrt{\frac{\log d}{n}}, ||\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}} - \mathbf{I}||_{\max} \leq CM\sqrt{\frac{\log d}{n}}, \text{ and} \tag{D.6}$$

$$||\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}||_{\max} \leq CM\sqrt{\frac{\log d}{n}}, ||\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}||_1 \leq CM\sqrt{\frac{s^2 \log d}{n}}, \tag{D.7}$$

where C is a universal constant only depending on r_0 in (D.4).

Remark D.1. *Note the first inequality in* (D.6) *directly follows from Equation (26) in Cai et al. (2011), the second inequality follows from the constraint in the CLIME estimator and the third inequality holds due to Theorem 6 in Cai et al. (2011).*

Given a random variable Z, we define its ψ_ℓ-norm for $\ell \geq 1$ as $\|Z\|_{\psi_\ell} = \sup_{p\geq 1} p^{-1/\ell}(\mathbb{E}|Z|^p)^{1/p}$. The following lemma controls the ψ_ℓ-norm of $\boldsymbol{X}$ and gives the lower bound of the variance of the debiased estimator.

Lemma D.3. There exist universal constants c and C only depending on r_0 in (D.4) such that

$$\sup_{\|\mathbf{v}\|_2=1} \|\mathbf{v}^\top \boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}\|_{\psi_2} \leq C \text{ and } \min_{j,k\in[d]} \mathbb{E}[(\boldsymbol{\Theta}_j^\top(\boldsymbol{X}\boldsymbol{X}^\top - \boldsymbol{\Sigma})\boldsymbol{\Theta}_k)^2] \geq c. \tag{D.8}$$

Proof. The first inequality in (D.8) immediately follows since $\mathbf{v}^\top\boldsymbol{\Sigma}^{-1/2}\boldsymbol{X} \sim N(0,1)$ for any $\|\mathbf{v}\|_2 = 1$. Regarding the second inequality, note that $\mathbb{E}[(\boldsymbol{\Theta}_j^\top(\boldsymbol{X}\boldsymbol{X}^\top - \boldsymbol{\Sigma})\boldsymbol{\Theta}_k)^2] = \mathrm{Var}(\boldsymbol{\Theta}_j^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{\Theta}_k)$. Below we calculate the expression of the general form $\mathrm{Var}\left(\mathbf{u}^\top\boldsymbol{X}\boldsymbol{X}^\top\mathbf{v}\right)$. Specifically, we apply Isserlis' theorem (Isserlis, 1918) to deal with the moments of Gaussian random variables. For any deterministic vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$, Isserlis' theorem says

$$\begin{aligned}
\mathrm{Var}(\mathbf{u}^\top\boldsymbol{X}\boldsymbol{X}^\top\mathbf{v}) &= \mathbb{E}[(\mathbf{u}^\top\boldsymbol{X})^2(\mathbf{v}^\top\boldsymbol{X})^2] - (\mathbb{E}[\mathbf{u}^\top\boldsymbol{X}\mathbf{v}^\top\boldsymbol{X}])^2 \\
&= \mathbb{E}[(\mathbf{u}^\top\boldsymbol{X})^2]\mathbb{E}[(\mathbf{v}^\top\boldsymbol{X})^2] + (\mathbb{E}[\mathbf{u}^\top\boldsymbol{X}\mathbf{v}^\top\boldsymbol{X}])^2 \\
&= (\mathbf{u}^\top\boldsymbol{\Sigma}\mathbf{u}^\top)(\mathbf{v}^\top\boldsymbol{\Sigma}\mathbf{v}^\top) + (\mathbf{u}^\top\boldsymbol{\Sigma}\mathbf{v}^\top)^2.
\end{aligned}$$

Therefore, we obtain the following,

$$\mathbb{E}[(\boldsymbol{\Theta}_j^\top(\boldsymbol{X}\boldsymbol{X}^\top - \boldsymbol{\Sigma})\boldsymbol{\Theta}_k)^2] = (\boldsymbol{\Theta}_j^\top\boldsymbol{\Sigma}\boldsymbol{\Theta}_j^\top)(\boldsymbol{\Theta}_k^\top\boldsymbol{\Sigma}\boldsymbol{\Theta}_k^\top) + (\boldsymbol{\Theta}_j^\top\boldsymbol{\Sigma}\boldsymbol{\Theta}_k^\top)^2 = \boldsymbol{\Theta}_{jj}\boldsymbol{\Theta}_{kk} + \boldsymbol{\Theta}_{jk}^2 \geq 1/r_0^2,$$

where the last inequality holds since $\lambda_{\min}(\boldsymbol{\Theta}) \geq 1/r_0$ when $\boldsymbol{\Theta} \in \mathcal{U}(M, s, r_0)$. ∎

Now we are ready to prove Lemma 2.1. Note the proof of this lemma follows a similar idea as the one used in Proposition 3.1 of Neykov et al. (2019). Since Lemma 2.1 involves the standardized version of the one-step estimator in Neykov et al. (2019), we still present the detailed proof for completeness.

Proof of Lemma 2.1. To approximate

$$T_E := \max_{(j,k)\in E} \sqrt{n}\left|(\widehat{\boldsymbol{\Theta}}_{jk}^{\mathrm{d}}/\sqrt{\widehat{\boldsymbol{\Theta}}_{jj}^{\mathrm{d}}\widehat{\boldsymbol{\Theta}}_{kk}^{\mathrm{d}}} - \boldsymbol{\Theta}_{jk}/\sqrt{\boldsymbol{\Theta}_{jj}\boldsymbol{\Theta}_{jk}})\right|, \tag{D.9}$$

by the multiplier bootstrap process

$$T_E^{\mathcal{B}} := \max_{(j,k)\in E} \frac{1}{\sqrt{n\ \widehat{\boldsymbol{\Theta}}_{jj}\widehat{\boldsymbol{\Theta}}_{kk}}}\left|\sum_{i=1}^{n}\widehat{\boldsymbol{\Theta}}_j^\top(\boldsymbol{X}_i\boldsymbol{X}_i^\top\widehat{\boldsymbol{\Theta}}_k - \mathbf{e}_k)\xi_i\right|, \tag{D.10}$$

we define two intermediate processes

$$\breve{T}_E := \max_{(j,k)\in E} \left| \frac{1}{\sqrt{n\ \mathbf{\Theta}_{jj}\mathbf{\Theta}_{kk}}} \sum_{i=1}^{n} \mathbf{\Theta}_j^\top (\boldsymbol{X}_i \boldsymbol{X}_i^\top \mathbf{\Theta}_k - \mathbf{e}_k) \right|, \tag{D.11}$$

$$\breve{T}_E^{\mathcal{B}} := \max_{(j,k)\in E} \left| \frac{1}{\sqrt{n\ \mathbf{\Theta}_{jj}\mathbf{\Theta}_{kk}}} \sum_{i=1}^{n} \mathbf{\Theta}_j^\top (\boldsymbol{X}_i \boldsymbol{X}_i^\top \mathbf{\Theta}_k - \mathbf{e}_k)\xi_i \right|. \tag{D.12}$$

The strategy of proving this lemma is to verify the three conditions in Corollary 3.1 of Chernozhukov et al. (2013):

(a) $\min_{j,k} \mathbb{E}[(\mathbf{\Theta}_j^\top(\boldsymbol{X}\boldsymbol{X}^\top\mathbf{\Theta}_k - \mathbf{e}_k))^2] > c$ and $\max_{j,k\in[d]} ||\mathbf{\Theta}_j^\top(\boldsymbol{X}\boldsymbol{X}^\top\mathbf{\Theta}_k - \mathbf{e}_k)||_{\psi_1} \le C$ for some positive constants c and C;

(b) $\mathbb{P}(|T_E - \breve{T}_E| > \zeta_1) < \zeta_2$ holds for some $\zeta_1, \zeta_2 > 0$;

(c) And $\mathbb{P}(\mathbb{P}_\xi(|T_E^{\mathcal{B}} - \breve{T}_E^{\mathcal{B}}| > \zeta_1 \mid \{\boldsymbol{X}_i\}_{i=1}^n) > \zeta_2) < \zeta_2$ holds for $\zeta_1\sqrt{\log d} + \zeta_2 = o(1)$.

Notice that in Chernozhukov et al. (2013), the original conditions require the last scaling to be $\zeta_1\sqrt{\log d} + \zeta_2 = o(n^{-c_1})$ for some c_1. This is because they pursue a stronger result that $|\mathbb{P}(T_E > \widehat{c}(\alpha, E)) - \alpha| = O(n^{-c_1})$. Since we do not emphasize on the polynomial decaying in our result, we only require $\zeta_1\sqrt{\log d} + \zeta_2 = o(1)$.

We start by checking the first condition (a). Lemma D.3 immediately implies the first part. By the second condition in (D.8), we have $\|\boldsymbol{X}_j\boldsymbol{X}_k - \mathbb{E}[\boldsymbol{X}_j\boldsymbol{X}_k]\|_{\psi_1} \le C$. By the definition of the ψ-norms, we have

$$\begin{aligned}\max_{j,k\in[d]} \|\mathbf{\Theta}_j^\top(\boldsymbol{X}_i\boldsymbol{X}_i^\top\mathbf{\Theta}_k - \mathbf{e}_k)\|_{\psi_1} &\le r_0^2 \|(\boldsymbol{X}_j\boldsymbol{X}_k - \mathbb{E}[\boldsymbol{X}_j\boldsymbol{X}_k])\|_{\psi_1} \\ &\le r_0^2 \sup_{\|\mathbf{v}\|_2=1} \|\mathbf{v}^\top\boldsymbol{X}\boldsymbol{X}^\top\mathbf{v} - \mathbb{E}[\mathbf{v}^\top\boldsymbol{X}\boldsymbol{X}^\top\mathbf{v}]\|_{\psi_1} = O(1).\end{aligned}$$

Regarding the condition (b), we check by bounding the difference $|T_E - \breve{T}_E|$. Recall the one-step estimator

$$\widehat{\mathbf{\Theta}}^{\mathrm{d}}_{jk} = \widehat{\mathbf{\Theta}}_{jk} - \frac{\widehat{\mathbf{\Theta}}_j^\top\left(\widehat{\mathbf{\Sigma}}\widehat{\mathbf{\Theta}}_k - \mathbf{e}_k\right)}{\widehat{\mathbf{\Theta}}_j^\top\widehat{\mathbf{\Sigma}}_j},$$

and plug it into T_E. Then we have the following bound,

$$\begin{aligned}|T_E - \breve{T}_E| &= \left| \max_{(j,k)\in E} \sqrt{n}\cdot\left(\frac{\widehat{\mathbf{\Theta}}^{\mathrm{d}}_{jk}}{\sqrt{\widehat{\mathbf{\Theta}}^{\mathrm{d}}_{jj}\widehat{\mathbf{\Theta}}^{\mathrm{d}}_{kk}}} - \frac{\mathbf{\Theta}_{jk}}{\sqrt{\mathbf{\Theta}_{jj}\mathbf{\Theta}_{jk}}}\right) - \max_{(j,k)\in E}\frac{\sqrt{n}}{\sqrt{\mathbf{\Theta}_{jj}\mathbf{\Theta}_{kk}}}\mathbf{\Theta}_j^\top(\widehat{\mathbf{\Sigma}}\mathbf{\Theta}_k - \mathbf{e}_k)\right| \\ &\le \frac{\mathrm{I}_1\mathrm{I}_2}{\min_{(j,k)\in E}\sqrt{\mathbf{\Theta}_{jj}\mathbf{\Theta}_{kk}}} + \frac{\mathrm{I}_3}{\min_{(j,k)\in E}\sqrt{\widetilde{\mathbf{\Theta}}^{\mathrm{d}}_{jj}\widetilde{\mathbf{\Theta}}^{\mathrm{d}}_{kk}}},\end{aligned} \tag{D.13}$$

where $\mathrm{I}_1 = \max_{(j,k)\in E}|\widehat{\mathbf{\Theta}}^{\mathrm{d}}_{jj}\widehat{\mathbf{\Theta}}^{\mathrm{d}}_{kk} - \mathbf{\Theta}_{jj}\mathbf{\Theta}_{kk}|$, $\mathrm{I}_2 = \max_{(j,k)\in E}|\sqrt{n}\cdot\mathbf{\Theta}_j^\top(\widehat{\mathbf{\Sigma}}\mathbf{\Theta}_k - \mathbf{e}_k|$ and

$$\mathrm{I}_3 = \max_{(j,k)\in E}\left|\sqrt{n}(\widehat{\mathbf{\Theta}}^{\mathrm{d}}_{jk} - \mathbf{\Theta}_{jk}) - \sqrt{n}\cdot\mathbf{\Theta}_j^\top\left(\widehat{\mathbf{\Sigma}}\mathbf{\Theta}_k - \mathbf{e}_k\right)\right|.$$

Note I_1 can be bounded using Lemma D.4, i.e.,

$$\mathrm{I}_1 = \max_{(j,k)\in E} |\widehat{\boldsymbol{\Theta}}^{\mathrm{d}}_{jj}\widehat{\boldsymbol{\Theta}}^{\mathrm{d}}_{kk} - \boldsymbol{\Theta}_{jj}\boldsymbol{\Theta}_{kk}| \le 2M\big\|\widehat{\boldsymbol{\Theta}}^{\mathrm{d}} - \boldsymbol{\Theta}\big\|_{\max} \le CM^2\sqrt{\frac{\log d}{n}}, \tag{D.14}$$

with probability $1-1/d^2$. As for the term I_2, we have

$$\begin{aligned}\mathrm{I}_2 = \max_{(j,k)\in E}\Big|\sqrt{n}\boldsymbol{\Theta}_j^\top\big(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Theta}_k - \mathbf{e}_k\big|\ &= \max_{(j,k)\in E}\sqrt{n}\Big|\boldsymbol{\Theta}_j^\top\big(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\big)\boldsymbol{\Theta}_k\Big| \\ &\le \sqrt{n}M^2\big\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\big\|_{\max} \le CM^2\sqrt{\log d}.\end{aligned} \tag{D.15}$$

Denote $\breve{\boldsymbol{\Theta}}_k = (\widehat{\boldsymbol{\Theta}}_{k1},\ldots,\widehat{\boldsymbol{\Theta}}_{k(j-1)},\boldsymbol{\Theta}_{kj},\widehat{\boldsymbol{\Theta}}_{k(j+1)},\ldots,\widehat{\boldsymbol{\Theta}}_{kd})^\top \in \mathbb{R}^d$. To deal with the term I_3, we first rewrite the following

$$\sqrt{n}(\widehat{\boldsymbol{\Theta}}^{\mathrm{d}}_{jk} - \boldsymbol{\Theta}_{jk}) = -\sqrt{n}\cdot\frac{\widehat{\boldsymbol{\Theta}}_j^\top\big(\widehat{\boldsymbol{\Sigma}}\breve{\boldsymbol{\Theta}}_k - \mathbf{e}_k\big)}{\widehat{\boldsymbol{\Theta}}_j^\top\widehat{\boldsymbol{\Sigma}}_j}, \tag{D.16}$$

then quantify $\sqrt{n}\widehat{\boldsymbol{\Theta}}_j^\top\big(\widehat{\boldsymbol{\Sigma}}\breve{\boldsymbol{\Theta}}_k - \mathbf{e}_k^\top\big)$. Notice that

$$\sqrt{n}\cdot\widehat{\boldsymbol{\Theta}}_j^\top\big(\widehat{\boldsymbol{\Sigma}}\breve{\boldsymbol{\Theta}}_k - \mathbf{e}_k^\top\big) = \underbrace{\sqrt{n}\cdot\widehat{\boldsymbol{\Theta}}_j^\top\big(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Theta}_k - \mathbf{e}_k^\top\big)}_{\mathrm{II}_1} + \underbrace{\sqrt{n}\cdot\widehat{\boldsymbol{\Theta}}_j^\top\widehat{\boldsymbol{\Sigma}}\big(\breve{\boldsymbol{\Theta}}_k - \boldsymbol{\Theta}_k\big)}_{\mathrm{II}_2}. \tag{D.17}$$

Further we expand II_1 as

$$\mathrm{II}_1 = \underbrace{\sqrt{n}\cdot\boldsymbol{\Theta}_j^\top\big(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Theta}_k - \mathbf{e}_k\big)}_{\mathrm{II}_{11}} + \underbrace{\sqrt{n}\cdot\big(\widehat{\boldsymbol{\Theta}}_j^\top - \boldsymbol{\Theta}_j^\top\big)\big(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Theta}_k - \mathbf{e}_k\big)}_{\mathrm{II}_{12}}, \tag{D.18}$$

where II_{11} can be rewritten as $\mathrm{II}_{11} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{\Theta}_j^\top(\boldsymbol{X}_i\boldsymbol{X}_i^\top\boldsymbol{\Theta}_k - \mathbf{e}_k)$. We bound $|\mathrm{II}_{12}|$ as

$$|\mathrm{II}_{12}| = \sqrt{n}\cdot\big(\widehat{\boldsymbol{\Theta}}_j - \boldsymbol{\Theta}_j\big)^\top\big(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\big)\boldsymbol{\Theta}_k \le \sqrt{n}\cdot\big\|\widehat{\boldsymbol{\Theta}}_j - \boldsymbol{\Theta}_j\big\|_1\big\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\big\|_{\max}\|\boldsymbol{\Theta}_k\|_1. \tag{D.19}$$

According to Lemma D.2, (D.19) yields that

$$\max_{j,k\in[d]}|\mathrm{II}_{12}| \lesssim M^2\frac{s\log d}{\sqrt{n}}, \tag{D.20}$$

with probability $1-1/d^2$. By Hölder's inequality and Lemma D.2, we finally obtain the bound on II_2:

$$\max_{j,k\in[d]}|\mathrm{II}_2| \le \sqrt{n}\cdot\max_{j,k\in[d]}\|\widehat{\boldsymbol{\Theta}}_j^\top\widehat{\boldsymbol{\Sigma}}_{-j}\|_\infty\big\|\widehat{\boldsymbol{\Theta}}_k - \boldsymbol{\Theta}_k\big\|_1 \lesssim M^2\frac{s\log d}{\sqrt{n}}, \tag{D.21}$$

with probability $1-1/d^2$. Therefore, we conclude that by (D.20) and (D.21), with probability $1-1/d^2$, the following holds:

$$\max_{j,k\in[d]}\sqrt{n}\cdot\Big|\widehat{\boldsymbol{\Theta}}_j^\top\big(\widehat{\boldsymbol{\Sigma}}\breve{\boldsymbol{\Theta}}_k - \mathbf{e}_k^\top\big) - \boldsymbol{\Theta}_j^\top\Big(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Theta}_k - \mathbf{e}_k^\top\Big)\Big| \lesssim M^2\frac{s\log d}{\sqrt{n}}. \tag{D.22}$$

Lemma D.2 also implies

$$\max_{j\in[d]}|\widehat{\boldsymbol{\Theta}}_j^\top\widehat{\boldsymbol{\Sigma}}_j - 1| \le \max_{j\in[d]}||\widehat{\boldsymbol{\Theta}}_j^\top\widehat{\boldsymbol{\Sigma}} - \mathbf{e}_j||_\infty \lesssim M\sqrt{\frac{\log d}{n}}. \tag{D.23}$$

Combining (D.17), (D.18) with (D.22) and (D.23), for sufficiently large d, n, we have, with probability $1-1/d^2$, the following holds:

$$
\begin{aligned}
\mathrm{I}_3 &\le \max_{(j,k)\in E} \sqrt{n}\Big|\frac{\widehat{\boldsymbol{\Theta}}_j^\top(\widehat{\boldsymbol{\Sigma}}\breve{\boldsymbol{\Theta}}_k - \mathbf{e}_k)}{\widehat{\boldsymbol{\Theta}}_j^\top\widehat{\boldsymbol{\Sigma}}_j} - \boldsymbol{\Theta}_j^\top(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Theta}_k - \mathbf{e}_k)\Big| \\
&\le \max_{(j,k)\in E}\big(2\sqrt{n}|\widehat{\boldsymbol{\Theta}}_j^\top\widehat{\boldsymbol{\Sigma}}_j - 1|\cdot|\boldsymbol{\Theta}_j^\top(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})\boldsymbol{\Theta}_k|\big) + 2\max_{(j,k)\in E}|\widehat{\boldsymbol{\Theta}}_j^\top(\widehat{\boldsymbol{\Sigma}}\breve{\boldsymbol{\Theta}}_k - \mathbf{e}_k) - \boldsymbol{\Theta}_j^\top(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Theta}_k - \mathbf{e}_k)| \\
&\le 2M\sqrt{n}\max_{j\in[d]}|\widehat{\boldsymbol{\Theta}}_j^\top\widehat{\boldsymbol{\Sigma}}_j - 1|\cdot||\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}||_{\max} + 2\max_{j,k\in[d]}(|I_{12}|+|I_2|) \lesssim M^2\frac{s\log d}{\sqrt{n}}, \qquad \text{(D.24)}
\end{aligned}
$$

where the second inequality uses $|x/(1+\delta)-y| \le 2|y\delta| + 2|x-y|$ for any $|\delta|<1/2$. Therefore, combining (D.13), (D.14),(D.15) with (D.24) and the fact $\min_{(j,k)\in E}\sqrt{\boldsymbol{\Theta}_{jj}\boldsymbol{\Theta}_{kk}} \ge \lambda_{\min}(\boldsymbol{\Theta}) \ge 1/r_0$ (as $\boldsymbol{\Theta}\in\mathcal{U}(M,s,r_0)$), we obtain the following:

$$
\mathbb{P}(|T_E - \breve{T}_E| > \zeta_1) < \zeta_2, \qquad \text{(D.25)}
$$

where $\zeta_1 = s\log d/\sqrt{n}$ and $\zeta_2 = 1/d^2$; thus the condition (b) is verified. Also note that $\zeta_1\sqrt{\log d} + \zeta_2 = s(\log d)^{3/2}/\sqrt{n} + 1/d^2 = o(1)$ holds under the stated scaling condition of Lemma 2.1.

Regarding the third condition (c), we bound the difference between $T_E^{\mathcal{B}}$ and $\breve{T}_E^{\mathcal{B}}$ as

$$
|T_E^{\mathcal{B}} - \breve{T}_E^{\mathcal{B}}| \le \max_{(j,k)\in E}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\Big(\frac{\widehat{\boldsymbol{\Theta}}_j^\top}{\sqrt{\widehat{\boldsymbol{\Theta}}_{jj}\widehat{\boldsymbol{\Theta}}_{kk}}}(\boldsymbol{X}_i\boldsymbol{X}_i^\top\widehat{\boldsymbol{\Theta}}_k - \mathbf{e}_k) - \frac{\boldsymbol{\Theta}_j^\top}{\sqrt{\boldsymbol{\Theta}_{jj}\boldsymbol{\Theta}_{kk}}}(\boldsymbol{X}_i\boldsymbol{X}_i^\top\boldsymbol{\Theta}_k - \mathbf{e}_k)\Big)\xi_i\Big|
$$

Conditioning on the data $\{\boldsymbol{X}_i\}_{i=1}^n$, the right hand side of the above inequality is a suprema of a Gaussian process. Therefore, we need to bound the following conditional variance

$$
\max_{(j,k)\in E}\frac{1}{n}\sum_{i=1}^n\Big[\frac{\widehat{\boldsymbol{\Theta}}_j^\top}{\sqrt{\widehat{\boldsymbol{\Theta}}_{jj}\widehat{\boldsymbol{\Theta}}_{kk}}}(\boldsymbol{X}_i\boldsymbol{X}_i^\top\widehat{\boldsymbol{\Theta}}_k - \mathbf{e}_k) - \frac{\boldsymbol{\Theta}_j^\top}{\sqrt{\boldsymbol{\Theta}_{jj}\boldsymbol{\Theta}_{kk}}}(\boldsymbol{X}_i\boldsymbol{X}_i^\top\boldsymbol{\Theta}_k - \mathbf{e}_k)\Big]^2
$$

Note the summand (for each i) can be bounded by

$$
2\frac{\mathrm{III}_1\mathrm{III}_2}{\min_{(j,k)\in E}\boldsymbol{\Theta}_{jj}\boldsymbol{\Theta}_{kk}} + 2\frac{\mathrm{III}_3}{\min_{(j,k)\in E}\widehat{\boldsymbol{\Theta}}_{jj}\widehat{\boldsymbol{\Theta}}_{kk}}
$$

where $\mathrm{III}_1, \mathrm{III}_2$ and III_3 are defined and bounded as below:

$$
\begin{aligned}
\mathrm{III}_1 &:= \max_{(j,k)\in E}|\widehat{\boldsymbol{\Theta}}_{jj}\widehat{\boldsymbol{\Theta}}_{kk} - \boldsymbol{\Theta}_{jj}\boldsymbol{\Theta}_{kk}|^2 = \mathrm{II}_1^2 \le \left(CM^2\sqrt{\frac{\log d}{n}}\right)^2 \\
\mathrm{III}_2 &:= \max_{(j,k)\in E}[\boldsymbol{\Theta}_j^\top(\boldsymbol{X}_i\boldsymbol{X}_i^\top\boldsymbol{\Theta}_k - \mathbf{e}_k)]^2 = \max_{(j,k)\in E}[\boldsymbol{\Theta}_j^\top(\boldsymbol{X}_i\boldsymbol{X}_i^\top - \boldsymbol{\Sigma})\boldsymbol{\Theta}_k]^2 \\
&\le \Big[M^2\max_i\|\boldsymbol{X}_i\boldsymbol{X}_i^\top - \boldsymbol{\Sigma}\|_{\max}\Big]^2 \\
\mathrm{III}_3 &= \max_{(j,k)\in E}\Big|\widehat{\boldsymbol{\Theta}}_j^\top(\boldsymbol{X}_i\boldsymbol{X}_i^\top\widehat{\boldsymbol{\Theta}}_k - \mathbf{e}_k) - \boldsymbol{\Theta}_j^\top(\boldsymbol{X}_i\boldsymbol{X}_i^\top\boldsymbol{\Theta}_k - \mathbf{e}_k)\Big|^2 \\
&\lesssim \Big[2M\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_1\max_i\|\boldsymbol{X}_i\boldsymbol{X}_i^\top - \boldsymbol{\Sigma}\|_{\max}\Big]^2.
\end{aligned}
$$

According to Lemma D.3, we have with probability $1-1/d^2$, $\max_i \|\boldsymbol{X}_i\boldsymbol{X}_i^\top - \boldsymbol{\Sigma}\|_{\max} \le C\sqrt{\log(dn)}$. Therefore, the event

$$\mathcal{E} = \Big\{ \max_{(j,k)\in E} \frac{1}{n}\sum_{i=1}^n \Big[\frac{\widehat{\boldsymbol{\Theta}}_j^\top}{\sqrt{\widehat{\boldsymbol{\Theta}}_{jj}\widehat{\boldsymbol{\Theta}}_{kk}}}(\boldsymbol{X}_i\boldsymbol{X}_i^\top\widehat{\boldsymbol{\Theta}}_k - \mathbf{e}_k) - \frac{\boldsymbol{\Theta}_j^\top}{\sqrt{\boldsymbol{\Theta}_{jj}\boldsymbol{\Theta}_{kk}}}(\boldsymbol{X}_i\boldsymbol{X}_i^\top\boldsymbol{\Theta}_k - \mathbf{e}_k)\Big]^2 \le CM^2\frac{(s\log(dn))^2}{n}\Big\}$$

satisfies $\mathbb{P}(\mathcal{E}^c) < 1/d^2$. Therefore, by the maximal inequality, under the event $\mathcal{E}$, we have

$$\begin{aligned}&\mathbb{E}\left[\max_{(j,k)\in E}\frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(\frac{\widehat{\boldsymbol{\Theta}}_j^\top}{\sqrt{\widehat{\boldsymbol{\Theta}}_{jj}\widehat{\boldsymbol{\Theta}}_{kk}}}(\boldsymbol{X}_i\boldsymbol{X}_i^\top\widehat{\boldsymbol{\Theta}}_k - \mathbf{e}_k) - \frac{\boldsymbol{\Theta}_j^\top}{\sqrt{\boldsymbol{\Theta}_{jj}\boldsymbol{\Theta}_{kk}}}(\boldsymbol{X}_i\boldsymbol{X}_i^\top\boldsymbol{\Theta}_k - \mathbf{e}_k)\Big)\xi_i \mid \{\boldsymbol{X}_i\}_{i=1}^n\right] \\ \lesssim\ & M^2\frac{(s\log dn)\sqrt{\log d}}{\sqrt{n}}.\end{aligned}$$

Applying Borell's inequality, we have with probability $1-1/d^2$,

$$\mathbb{P}\left(\max_{(j,k)\in E}\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\frac{\widehat{\boldsymbol{\Theta}}_j^\top(\boldsymbol{X}_i\boldsymbol{X}_i^\top\widehat{\boldsymbol{\Theta}}_k - \mathbf{e}_k)}{\sqrt{\widehat{\boldsymbol{\Theta}}_{jj}\widehat{\boldsymbol{\Theta}}_{kk}}} - \frac{\boldsymbol{\Theta}_j^\top(\boldsymbol{X}_i\boldsymbol{X}_i^\top\boldsymbol{\Theta}_k - \mathbf{e}_k)}{\sqrt{\boldsymbol{\Theta}_{jj}\boldsymbol{\Theta}_{kk}}}\right)\xi_i > C\sqrt{\frac{s^2\log^4 dn}{n}} \mid \{\boldsymbol{X}_i\}_{i=1}^n\right) \le 1/d^2.$$

This implies that

$$\mathbb{P}\left(\mathbb{P}_\xi(|T_E^{\mathcal{B}} - \breve{T}_E^{\mathcal{B}}| > \sqrt{(s^2\log^4 dn)/n}) > 1/d^2\right) < 1/d^2.$$

Now we can verify the condition (c) by showing

$$\mathbb{P}(\mathbb{P}_\xi(|T_E^{\mathcal{B}} - \breve{T}_E^{\mathcal{B}}| > \zeta_1 \mid \{\boldsymbol{X}_i\}_{i=1}^n) > \zeta_2) < \zeta_2, \tag{D.26}$$

where $\zeta_1 = s(\log d)^2/\sqrt{n}$, $\zeta_2 = 1/d^2$ and the condition $\zeta_1\sqrt{\log d} + \zeta_2 = s(\log d)^{3/2}/\sqrt{n} + 1/d^2 = o(1)$ holds under the stated scaling condition of Lemma 2.1. Therefore, by Corollary 3.1 of Chernozhukov et al. (2013), we have

$$\lim_{(n,d)\to\infty} |\mathbb{P}(T_E > \widehat{c}(\alpha, E)) - \alpha| = 0. \tag{D.27}$$

And it holds for any edge set E, thus the proof is complete. □

Lemma D.4. Under the same conditions as Lemma 2.1, we have

$$\mathbb{P}\Big(\max_{j,k\in[d]} |\widehat{\boldsymbol{\Theta}}_{jk}^{\mathrm{d}} - \boldsymbol{\Theta}_{jk}| > C_0\sqrt{\frac{\log d}{n}}\Big) < \frac{2}{d^2}, \tag{D.28}$$

for some constant $C_0 > 0$.

Proof. By (D.16) and (D.24), we have with probability $1-1/d^2$,

$$\max_{j,k\in[d]} |\widehat{\boldsymbol{\Theta}}_{jk}^{\mathrm{d}} - \boldsymbol{\Theta}_{jk} + \boldsymbol{\Theta}_j^\top(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Theta}_k - \mathbf{e}_k)| \le C_1\frac{s\log d}{n}.$$

By Lemma D.3 and $\|\boldsymbol{\Theta}\|_2 \le r_0$, we have $\|\boldsymbol{\Theta}_j^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{\Theta}_k\|_{\psi_1} \le C_2 r_0^2$. Applying the maximal inequality (Lemma 2.2.2 in Van Der Vaart and Wellner (1996)), we have for some constant $C_3 > 0$

$$\begin{aligned}&\mathbb{P}\Big(\max_{j,k\in[d]} |\boldsymbol{\Theta}_j^\top(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Theta}_k - \mathbf{e}_k)| > C_3 r_0^2\sqrt{\frac{\log d}{n}}\Big) \\ \le\ &\mathbb{P}\Big(\max_{j,k\in[d]}\Big|\frac{1}{n}\sum_{i=1}^n(\boldsymbol{\Theta}_j^\top\boldsymbol{X}_i\boldsymbol{X}_i^\top\boldsymbol{\Theta}_k - \mathbb{E}[\boldsymbol{\Theta}_j^\top\boldsymbol{X}_i\boldsymbol{X}_i^\top\boldsymbol{\Theta}_k)]\Big| > C_3 r_0^2\sqrt{\frac{\log d}{n}}\Big) \le 1/d^2.\end{aligned}$$

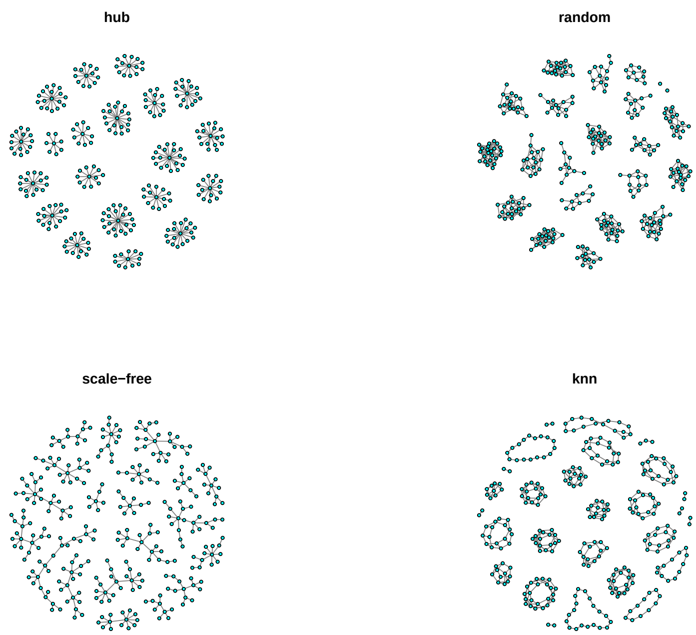
With $C_0 = C_1 + C_3$, (D.28) is proved. And it is not hard to show a similar result for the standardized one-step estimator also holds, i.e.,

$$\mathbb{P}\Big(\max_{j,k\in[d]} |\widetilde{\boldsymbol{\Theta}}^{\mathrm{d}}_{jk} - \boldsymbol{\Theta}^{\star}_{jk}| > C_0'\sqrt{\frac{\log d}{n}}\Big) < \frac{2}{d^2} \tag{D.29}$$

for some constant $C_0' > 0$. □

E Tables and plots deferred from the main paper

E.1 Graph pattern demonstration



E.2 Tables of $q\frac{d_0}{d}$

E.3 Supplementary FDP and Power plots

Table 3: $q\frac{d_0}{d}$

$d = 300$	$q = 0.1$			$q = 0.2$		
n	200	300	400	200	300	400
			$p = 20$			
hub	0.0930	0.0930	0.0930	0.1870	0.1870	0.1870
random	0.0620	0.0610	0.0600	0.1230	0.1220	0.1200
scale-free	0.0810	0.0810	0.0810	0.1620	0.1630	0.1620
knn	0.0680	0.0700	0.0690	0.1360	0.1390	0.1390
			$p = 30$			
hub	0.0900	0.0900	0.0900	0.1800	0.1800	0.1800
random	0.0810	0.0810	0.0810	0.1620	0.1620	0.1620
scale-free	0.0810	0.0810	0.0810	0.1620	0.1620	0.1610
knn	0.0730	0.0750	0.0740	0.1460	0.1510	0.1480

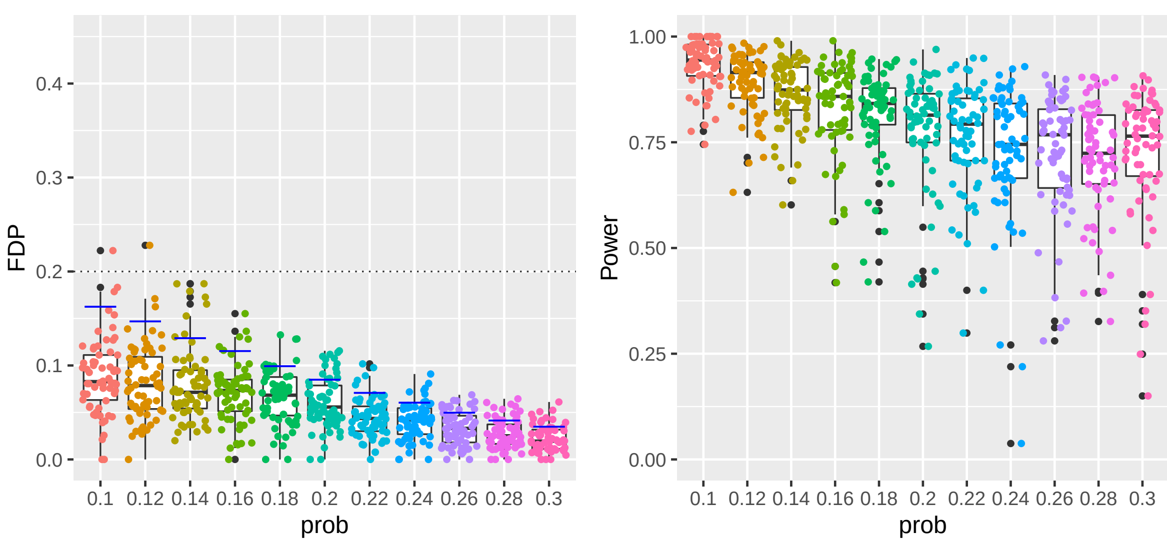


Figure 6: FDP and power plots with $p = 20$ and the nominal FDR level $q = 0.2$. The other setups are the same as Figure 3.

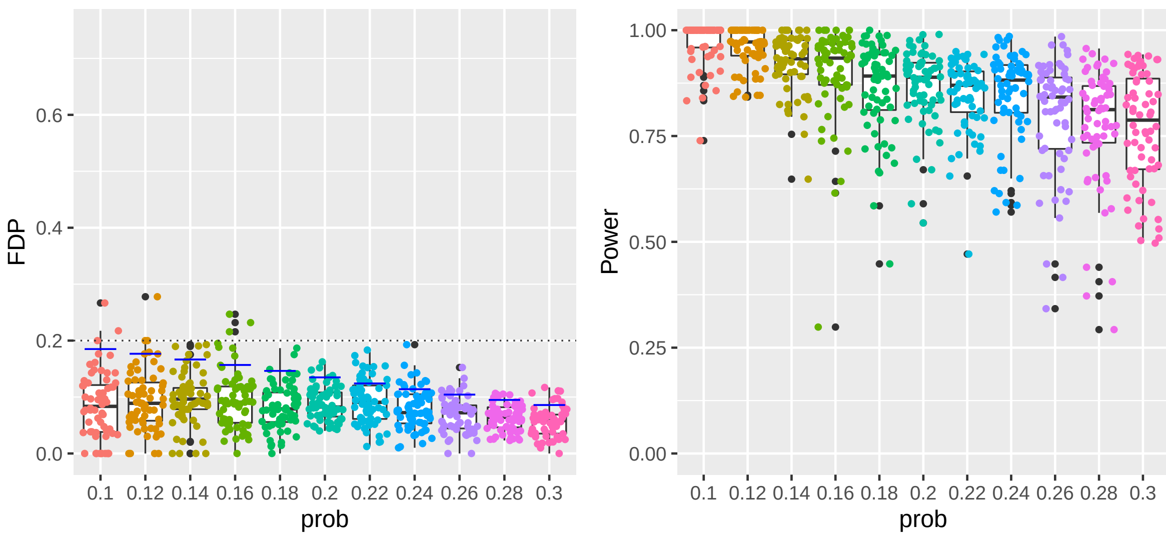


Figure 7: FDP and power plots with $p = 30$ and the nominal FDR level $q = 0.2$. The other setups are the same as Figure 3.