The Annals of Statistics
2024, Vol. 52, No. 1, 78–102
https://doi.org/10.1214/23-AOS2296

STARTREK: COMBINATORIAL VARIABLE SELECTION WITH FALSE DISCOVERY RATE CONTROL

BY LU ZHANG[1,a] AND JUNWEI LU[2,b]

[1]*Department of Statistics, Harvard University,* [a]*luzhang.r2@gmail.com*
[2]*Department of Biostatistics, Harvard University,* [b]*junweilu@hsph.harvard.edu*

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. Compared to the Gaussian comparison bound via the Kolmogorov distance established by Chernozhukov, Chetverikov and Kato (*Ann. Statist.* **42** (2014) 1787–1818), 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.

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 [55, 60, 68]. Scientists and practitioners aim to understand the underlying network structure behind large-scale datasets. For a high-dimensional random vector $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_d) \in \mathbb{R}^d$, 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 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. We denote the associated weight matrix by $\boldsymbol{\Theta}$ with $\boldsymbol{\Theta}_{jk}$ being the weight on the edge between j and k. Many existing works in the literature seek to learn the structure of $\mathcal{G}$ via estimating the weight matrix $\boldsymbol{\Theta}$. For example, [9, 27, 41, 58, 63, 65, 67, 71, 84] 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, that is, constructing hypothesis tests and confidence intervals on the network edges;

Received September 2021; revised December 2022.
MSC2020 subject classifications. Primary 62H15, 62H22; secondary 60F99.
Key words and phrases. Graphical models, multiple testing, false discovery rate control, combinatorial inference, Gaussian multiplier bootstrap, comparison bounds.

see [11, 12, 24, 26, 29, 33, 66, 81]. Recently, simultaneously testing multiple hypotheses on edges of the graphical models have received increasing attention [10, 25, 43, 46, 78, 79].

Most of the aforementioned works formulate the testing problems based on continuous parameters and local properties. For example, [46] 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 [31, 42]. In gene co-expression network analysis, identifying central regulators/hub genes [50, 51, 83] 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 [64, 69, 77] 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:

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

based on i.i.d. samples $\boldsymbol{X}_1, \ldots \boldsymbol{X}_n \overset{i.i.d.}{\sim} \boldsymbol{X} \in \mathbb{R}^d$. 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:

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

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 [54], 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 [4, 34–37, 46, 75, 76, 78, 79, 86]. 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 [5–7]. 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 familywise error rate (FWER) control, relies on the nonasymptotic bounds of the Kolmogorov distance between the true distribution of the maximum statistics and the Gaussian multiplier bootstrap approximation, which is established in [16]. 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 [14]. 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 $|\frac{\mathbb{P}(\|U\|_\infty > t)}{\mathbb{P}(\|V\|_\infty > t)} - 1|$ for certain range of t. Compared to the Kolmogorov distance $\sup_{t\in\mathbb{R}} |\mathbb{P}(\|U\|_\infty > t) - \mathbb{P}(\|V\|_\infty > t)|$ [18], 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 [73] into an ordinary differential inequality such that the relative error can be controlled via the Grönwall's inequality [28]. To control the Δ_0 rate, the anticoncentration inequality of Gaussian maxima developed in [18] is no longer sufficient, and we establish a new type of anticoncentration inequality for the derivatives of the soft-max of high-dimensional Gaussian vectors. The existing works on the Cramér-type comparison bounds such as [14, 48, 49] does not cover the high-dimensional maximum statistics. Therefore, their techniques cannot 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 [5–7] 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 nontrivial for high-dimensional hypothesis testing problems concerning continuous parameters [4, 35–37, 75, 76, 86], 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 [2, 3, 13, 20, 21, 80]. 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 [34, 45–47, 78, 79]. Among them, [46] 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 [39] provides a confidence interval for the number of spiked eigenvalues in a covariance matrix. Reference [38] focuses on estimating the number of communities in a network and yields confidence lower bounds. References [54, 61] 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. Reference [70] 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. 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 7, we demonstrate StarTrek's performance through empirical simulations and a real data application.

1.3. *Notation.* 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 $A \ominus B$, that is, $A \ominus B = (A \setminus B) \cup (B \setminus A)$;

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 $C > 0$. 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 $\epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}(X_n - a > \epsilon) = 0$. Given a random variable Z, we define its ψ_ℓ-norm for $\ell \ge 1$ as $\|Z\|_{\psi_\ell} = \sup_{p\ge 1} p^{-1/\ell} (\mathbb{E}|Z|^p)^{1/p}$. 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 $\|\mathbf{A}\|_0 = \sum_{i\in[d_1], j\in[d_2]} \mathbb{1}(\mathbf{A}_{ij} \ne 0)$. 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 $\mathcal{V}_1 \ne \mathcal{V}_2$, 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, based on the n data samples $\boldsymbol{X}_1, \dots \boldsymbol{X}_n \overset{i.i.d.}{\sim} \boldsymbol{X} \in \mathbb{R}^d$. The selection problem is equivalent to simultaneously testing d hypotheses:

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

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:

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

where $\mathcal{H}_0 = \{j \mid \text{degree of node } j < k_\tau\}$. Given the data $\boldsymbol{X}_1, \dots \boldsymbol{X}_n$ from the graphical model, 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, that is, whose degree is larger than or equal to k_τ.

2.1. *StarTrek filter.* Letting $\mathbf{\Theta}_j$ be the jth row of $\mathbf{\Theta}$ and $\mathbf{\Theta}_{j,-j}$ be the vector $\mathbf{\Theta}_j$ excluding its jth element, we formulate the testing problem for each single node as below:

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

To test the above hypothesis, we need some estimator of the weight matrix $\mathbf{\Theta}$. In the Gaussian graphical model, it is natural to use the estimator of a precision matrix. In the bipartite graph (multiple response model), an estimated parameter matrix will suffice. Denote this generic

Algorithm 1 Skip-down Method in [54] (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 $|E_t^c|\geq k_\tau$ or $\mathcal{R}=\varnothing$
Output: $\psi_{j,\alpha}=1$ if $|E_t^c|\geq k$ and $\psi_{j,\alpha}=0$ otherwise.

estimator by $\widetilde{\mathbf{\Theta}}$ (without causing confusion in notation), and the maximum test statistic over a given subset E of $\mathcal{V}\times\mathcal{V}$ will be

$$T_E := \max_{(j,k)\in E} \sqrt{n}|\widetilde{\mathbf{\Theta}}_{jk}| \tag{2.2}$$

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 the skip-down procedure can be used, which was originally proposed in [54] for testing a family of monotone graph invariants. When applied to the specific degree testing problem, it leads to Algorithm 1.

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 hub nodes as possible with the FDR remaining controlled under the nominal level q. For example, if the BHq procedure [6] is considered, $\widehat{\alpha}$ can be defined as follows:

$$\widehat{\alpha}=\sup\left\{\alpha\in(0,1): \frac{\alpha d}{\max\{1,\sum_{j\in[d]}\psi_{j,\alpha}\}}\leq q\right\}. \tag{2.3}$$

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 if it only involves estimating k_τ different quantiles of some maximum statistics per node, which is more efficient than the Skip-down procedure [54] in terms of computation. We shall note that Algorithm 2 is equivalent to running the BHq procedure with Algorithm 1: rejecting H_{0j} if $\psi_{j,\widehat{\alpha}}=1$, $j\in[d]$, where $\widehat{\alpha}$ is defined by (2.3) and the test $\psi_{j,\alpha}$ is defined by Algorithm 1; see the proof at the beginning of Appendix A in the Supplementary Material [85]. Without causing confusion,

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 $\{|\widetilde{\mathbf{\Theta}}_{j\ell}|:\ell\neq j\}$ as $|\widetilde{\mathbf{\Theta}}_{j,(1)}|\geq|\widetilde{\mathbf{\Theta}}_{j,(2)}|\geq\cdots\geq|\widetilde{\mathbf{\Theta}}_{j,(d-1)}|$, where $|\widetilde{\mathbf{\Theta}}_{j,(\ell)}|$ is the ℓth largest entry. Compute $\alpha_j=\max_{1\leq s\leq k_\tau}\widehat{c}^{-1}(\sqrt{n}|\widetilde{\mathbf{\Theta}}_{j,(s)}|, E_j^{(s)})$ where $E_j^{(s)}:=\{(j,\ell):\ell\neq j, |\widetilde{\mathbf{\Theta}}_{j\ell}|\leq|\widetilde{\mathbf{\Theta}}_{j,(s)}|\}$.
end for
Order α_j as $\alpha_{(1)}\leq\alpha_{(2)}\leq\cdots\leq\alpha_{(d)}$ and set $\alpha_{(0)}=0$, and let $j_{\max}=\max\{0\leq j\leq d: \alpha_{(j)}\leq qj/d\}$.
Output: $S=\{j:\alpha_j\leq\alpha_{(j_{\max})}\}$ if $j_{\max}>0$; $S=\varnothing$ otherwise.

we will refer to the BHq procedure with Algorithm 1 simply by Algorithm 1. Intuitively, Algorithm 2 directly acts the hub node selection by conducting the BHq adjustment to the p-values of all d nodes but Algorithm 1 has to specify a significance level first to test the degree for each node and then search for an appropriate significance level for all nodes. We conduct a numerical comparison for the two methods in Section 7.3.

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 $\mathbf{\Sigma}^U = (\sigma^U_{jk})_{1\le j,k\le d}$, $\mathbf{\Sigma}^V = (\sigma^V_{jk})_{1\le j,k\le d}$. 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 $\Delta_0 := \|\mathbf{\Sigma}^U - \mathbf{\Sigma}^V\|_0 = \sum_{j,k\in[d]} \mathbb{1}(\sigma^U_{jk} \neq \sigma^V_{jk})$. 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*

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

for some constant $C_0 > 0$.

REMARK 3.2. We can actually prove a more general form (see Theorem B.3 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 $A(\Delta_\infty) = M_1 \log d\Delta_\infty^{1/2}\exp(M_2\log^2 d\Delta_\infty^{1/2})$ with the constants M_1, M_2 only depending on the variance terms $\min_{1\le j\le d}\{\sigma^U_{jj}, \sigma^V_{jj}\}$, $\max_{1\le j\le d}\{\sigma^U_{jj}, \sigma^V_{jj}\}$ and M_3 being a universal constant.

When applying Theorem 3.1 to Gaussian multiplier bootstrap, Δ_∞ actually controls the maximum differences between the true covariance matrix and the empirical covariance matrix, where $\Delta_\infty = O_P(\sqrt{\log d/n})$. The proof of the theorem can be found in Appendix B.1. Compared with the proof of Kolmogorov distance results in [16, 17], the key innovation in our proof of the Cramér-type Gaussian comparison bounds is a contraction mapping inequality. In specific, denote the Slepian interpolation $W(s) = \sqrt{s}U + \sqrt{1-s}V$, $s \in [0,1]$ and the tail probability of maxima $Q_t(s) = \mathbb{P}(||W(s)||_\infty > t)$. Our proof shows that $R_t(s) = Q_t(s)/Q_t(0) - 1$ has the following key inequality:

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

where AB and A are only depending on Δ_∞. By Grönwall's inequality [28], 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 especially in the applications of graphical models, 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. Namely, 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 3.3 (CCB with elementwise ℓ_0-norm difference). *Assume the Gaussian random vectors U and V have unit variances, that is, $\sigma^U_{jj} = \sigma^V_{jj} = 1$, $j \in [d]$ and there exists some $\sigma_0 < 1$ such that $|\sigma^V_{jk}| \le \sigma_0$, $|\sigma^U_{jk}| \le \sigma_0$ for any $j \ne k$. Suppose there exists a disjoint $\mathfrak{p}$-partition of nodes $\bigcup_{\ell=1}^{\mathfrak{p}} \mathcal{C}_\ell = [d]$ such that $\sigma^U_{jk} = \sigma^V_{jk} = 0$ when $j \in \mathcal{C}_\ell$ and $k \in \mathcal{C}_{\ell'}$ for some $\ell \ne \ell'$. We have*

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

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 nonhub nodes, we can show $\Delta_0 = O(1)$. In Theorem 3.3, 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.6 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 the 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 $U = (X_1, X_2, Z, \ldots, Z) \in \mathbb{R}^d$, $V = (Y_1, Y_2, Z, \ldots, Z) \in \mathbb{R}^d$, where $\operatorname{corr}(X_1, X_2) = 0.9$, $\operatorname{corr}(Y_1, Y_2) = 0$ 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 $\mathfrak{p} = 2$.

The proof of Theorem 3.3 can be found in Appendix B.2. Our main technical innovation is to establish a new type of anticoncentration bound for "derivatives" of Gaussian maxima. Different from the anticoncentration inequalities in [17] bounding the maxima of Slepian interpolation $\mathbb{E}[\mathbb{1}(t - \epsilon \le ||W(s)||_\infty \le t + \epsilon)]$, we are able to further bound its derivatives:

$$\mathbb{E}\big[\big|\partial_j \partial_k \varphi\big(W(s)\big)\big| \cdot \mathbb{1}\big(t - \epsilon \le \|W(s)\|_\infty \le t + \epsilon\big)\big] \lesssim \frac{\mathbb{P}(\|V\|_\infty > t)(\log d)^2}{\epsilon \beta \mathfrak{p}}, \tag{3.3}$$

where φ is the some smooth approximation of the maxima with the parameter β measuring the level of approximation. The above anticoncentration bound is nonuniform and has only a logarithm dependence on the dimension d. It provides a relatively sharp characterization when t is large and the graph is not highly connected (i.e., $\mathfrak{p}$ is 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: $\boldsymbol{\Theta}$ 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 the 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 $\boldsymbol{\Theta}$ indicates whether a given predictor is

nonnull or not for a given response variable. In the case of a 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 nonnull for all response variables [22, 62]. This section problem is columnwise 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 rowwise 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 nonnull predictors. We will call this type of response variables as 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:

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

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 $\mathbf{\Theta}$, and we want to select columns of the parameter matrix, which has at least k_τ nonzero entries, that is, select nodes with a 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 [36, 37]. 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

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

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 $\widetilde{\mathbf{\Theta}}^{\mathrm{d}} := (\widetilde{\mathbf{\Theta}}_1^{\mathrm{d}}, \ldots, \widetilde{\mathbf{\Theta}}_{d_1}^{\mathrm{d}})^\top$, 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 works studying the asymptotic distribution of the debiased Lasso estimator. Among them, the results in [36] (when translated into our multitask regression setup) imply, for each response variable $\boldsymbol{Y}^{(j)}$, $j \in [d_1]$,

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

under proper assumptions. Additionally, with a natural probabilistic model of the design matrix, the bias term can be shown to be $\|\Xi\|_\infty = O(\frac{s \log d_2}{\sqrt{n}})$ with high probability. As discussed in [36], 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 [36, 74], given by the following joint optimization problem:

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

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 $T_E = \max_{(j,k)\in E}\sqrt{n}|\widetilde{\boldsymbol{\Theta}}^{\mathrm{d}}_{jk}|$ 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:

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

Indeed, under proper scaling conditions, we can show that, that is, as $n, d \to \infty$,

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

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, that is, (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 [36]:

- Let $\boldsymbol{\Sigma} = \mathbb{E}[\boldsymbol{X}_1\boldsymbol{X}_1^\top] \in \mathbb{R}^{d_2\times d_2}$ 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 sub-Gaussian rows, with zero mean and sub-Gaussian 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. This is when the Cramér-type comparison bound for Gaussian maxima established in Section 3 play its role. In addition, the 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 nonnull parameter coefficients all exceed certain thresholds, and thus characterizes the overall signal strength. As mentioned at the beginning, the application of the StarTrek filter in this section is less involved than that in Gaussian graphical models. The major simplification comes from how the multitask regression problem with linear models gets set up: the response vector $\boldsymbol{Y}_i \in \mathbb{R}^{d_1}$ follows $\boldsymbol{Y}_i = \boldsymbol{\Theta}\boldsymbol{X}_i + \boldsymbol{E}_i$, where $\boldsymbol{E}_i \sim \mathcal{N}(0, \mathbf{D}_{d_1\times d_1})$ and $\boldsymbol{X}_i \perp\!\!\!\perp \boldsymbol{E}_i$; thus those d_1 response variables are independent, conditional on the covariate $\boldsymbol{X}_i$. Such conditional independence carries over to the statistics of testing each response (node) in our proposed method. In Gaussian graphical models, the test statistics for each node unavoidably have complicated dependence structures, due to the nature of this combinatorial selection problem and how the nodes are connected to each other. To this end, we will introduce some quantity (see (5.3)) to measure the dependence level of the graph and characterize how such dependence affects the validity of our StarTrek filter (see (5.5) in Assumption 5.1). 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*

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

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 [34, 46, 47, 49, 78, 79].

5. Discovering hub nodes in Gaussian graphical models. This section focuses on the hub node selection problem on Gaussian graphical models where $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \overset{\text{i.i.d.}}{\sim} N_d(0, \boldsymbol{\Sigma})$. Let the weight matrix be the precision matrix $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}$. Given some estimator $\widehat{\boldsymbol{\Theta}}$ (e.g., the graphical Lasso (GLasso) estimator [27] or the CLIME estimator [9]), we consider the following one-step estimator $\{\widehat{\boldsymbol{\Theta}}^{\mathrm{d}}_e\}_{e\in\mathcal{V}\times\mathcal{V}}$:

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

where $\mathbf{e}_k$ denotes the kth canonical basis in $\mathbb{R}^d$. Then we use its standardized version $\{\widetilde{\boldsymbol{\Theta}}^{\mathrm{d}}_e\}_{e\in\mathcal{V}\times\mathcal{V}}$ as the input of Algorithm 2. 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 the certain subset E. The quantiles are approximated using the Gaussian multiplier bootstrap [16]: given the Gaussian multipliers $\xi_i \overset{\text{i.i.d.}}{\sim} N(0,1)$, we compute

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

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

Reference [16] shows that the above quantile approximation is accurate enough for FWER control in modern high-dimensional simultaneous testing problems. Their results are based on the control of the nonasymptotic bounds in a Kolmogorov distance sense. Reference [54] 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 2.1. Compared with the results in [16], 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 3.3. In a bit more detail, computing the maximum test statistic for testing node actually involves the whole graph, resulting in complicated dependence among the test statistics. The nondifferentiability of the maximum function makes it very difficult to track this dependence. Also, note that this type of difficulty cannot 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, [46, 78, 79] 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 nonhub response variables in Section 4. Similarly, the set of nonhub 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:

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

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 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 nonhub nodes. Therefore, $|S|$ actually measures the dependence level via checking how a pair of nonhub nodes interact through other nodes. References [10, 46] 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 a 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 two 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 nonhub nodes are only connected to the hub node, while in (d), the edges are evenly distributed and each node is connected to its two nearest neighbors. 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 7.

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. Similarly, as in Section 4, we define ρ to measure the signal strength, that is, $\rho = |\mathcal{B}|/d$, where $\mathcal{B} := \{j \in \mathcal{H}_0^c : \forall k \in$

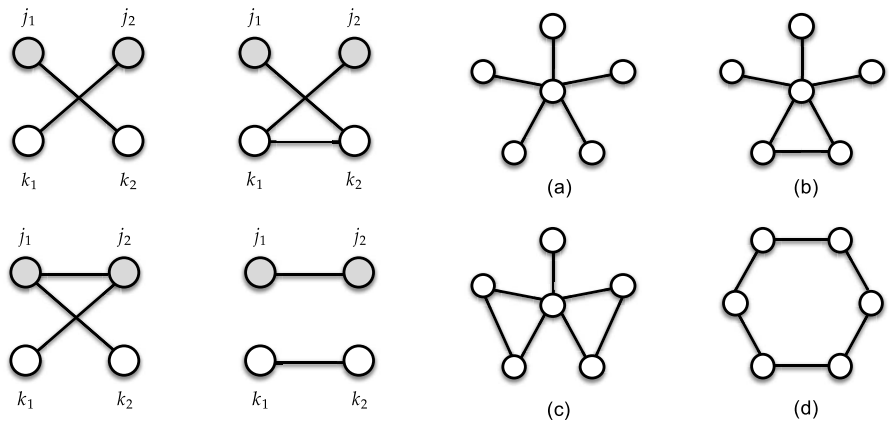

FIG. 1. *Left panel*: *a graphical demonstration of the definition of S via four examples of a* 4*-vertex graph*; *Right panel*: *four different graph patterns with* 6 *vertices*. *Calculating* $|S|$ *yields* $10, 15, 24, 51$ *for* (*a*), (*b*), (*c*), (*d*), *respectively.*

$\text{supp}(\mathbf{\Theta}_j), |\mathbf{\Theta}_{jk}| > c\sqrt{\log d/n}\}$. 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 [40] and the Cramér-type Gaussian comparison bound established in Theorem 3.1. 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 [16]. 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 nonhub 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. When $|S|/d_0^2$ is small, the graph is allowed to be more connected. When there exist more subgraphs, we allow higher correlations between the nonhub 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 3.3. 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* $(n, d) \to \infty$,

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

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 7.2. 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 [34, 46, 47, 78, 79] focus on the discovery of continuous signals and their tools are not applicable to the problem here.

6. StarTrek for general graphical models. Sections 4 and 5 apply the StarTrek filter to two concrete examples: Gaussian graphical models and multitask regression and provide FDR results with explicit assumptions. In this section, we will discuss how to generalize the results in Theorem 5.2 to the general graphical models.

Recall that in Section 2.1, we denote $\mathbf{\Theta}$ as the weight matrix of the general graphical models, that is, $\mathbf{\Theta}_e \neq 0$ if and only if $e \in \mathcal{E}$ where $\mathcal{E}$ is the edge set, and $\widetilde{\mathbf{\Theta}}$ is the generic estimator

of $\mathbf{\Theta}$. Since Algorithm 2 requires the quantile of the maximal statistics in (2.2), we need the estimator $\widetilde{\mathbf{\Theta}}$ to be asymptotically Gaussian. In specific, for each $e \in \mathcal{V} \times \mathcal{V}$, there exist n i.i.d. mean zero random vectors $\boldsymbol{Y}_1(e), \ldots, \boldsymbol{Y}_n(e)$ such that $\widetilde{\mathbf{\Theta}}_e - \mathbf{\Theta}_e = n^{-1} \sum_{i=1}^n \boldsymbol{Y}_i(e) + o_P(1/\sqrt{n})$. We can then estimate the quantile of $T_E = \max_{e \in E} \sqrt{n} |\widetilde{\mathbf{\Theta}}_e - \mathbf{\Theta}_e|$ by some Gaussian multiplier bootstrap statistic $T_E^{\mathcal{B}}$, that is,

$$\widehat{c}(\alpha, E) = \inf\{t \in \mathbb{R} : \mathbb{P}_\xi(T_E^{\mathcal{B}} \le t) \ge 1 - \alpha\}. \tag{6.1}$$

In specific, similar to the assumptions in Theorem 3.2 of [16], we need the following general assumptions on $\widetilde{\mathbf{\Theta}}$.

ASSUMPTION 6.1. There exist $\zeta_1 \ge 0$ and $\zeta_2 \ge 0$ such that for any edge set $E \subseteq \mathcal{V} \times \mathcal{V}$, we have

$$\mathbb{P}(|T_E - T_{0E}| > \zeta_1) < \zeta_2, \quad \text{where } T_{0E} = \max_{e \in E} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{Y}_i(e) \right|,$$

$$\mathbb{P}(\mathbb{P}_\xi(|T_E^{\mathcal{B}} - T_{0E}^{\mathcal{B}}| > \zeta_1) > \zeta_2) < \zeta_2, \quad \text{where } T_{0E}^{\mathcal{B}} = \max_{e \in E} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{Y}_i(e) \xi_i \right|,$$

where the i.i.d. mean zero random vectors $\boldsymbol{Y}_1(e), \ldots, \boldsymbol{Y}_n(e) \sim \boldsymbol{Y}(e)$ satisfy $\min_{e \in \mathcal{V} \times \mathcal{V}} \mathbb{E}[\boldsymbol{Y}^2(e)] > c$, $\max_{e \in \mathcal{V} \times \mathcal{V}} \|\boldsymbol{Y}(e)\|_{\psi_1} \le C$ and $\max_{e \in \mathcal{V} \times \mathcal{V}} \|\boldsymbol{Y}^2(e)\|_{\psi_1} \le C\sqrt{n/\log d}$ for some positive constants c and C. The multiplier variables $\xi_i \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$ are independent from the data and $\mathbb{P}_\xi$ is the measure only on $\xi_1, \ldots, \xi_n$.

For the multitask regression problem, we verify the above assumption in Lemma A.7 in the Supplementary Material [85]. As for the Gaussian graphical models, the assumption is verified in the proof of Lemma C.4 in the Supplementary Material [85]. The assumption has also been validated under other graphical models. For example, [82] proved the assumption holds for the exponential family pairwise graphical models, which include nonnegative Gaussian, conditionally specified mixed graphical models, exponential square-root graphical model, etc.

Similar to (5.3), for the general case, we also need to define a dependency set S as

$$S = \{(j_1, j_2, k_1, k_2) : j_1, j_2 \in \mathcal{H}_0, j_1 \ne j_2, \mathbf{\Theta}_{j_1 k_1} = \mathbf{\Theta}_{j_2 k_2} = 0, \\ \operatorname{Cov}(\boldsymbol{Y}((j_1, k_1)), \boldsymbol{Y}((j_2, k_2))) \ne 0\}.$$

We also impose the cardinality of S similar to Assumption 5.1 for general graphical models.

ASSUMPTION 6.2. Suppose that $\mathbf{\Theta} \in \mathcal{U}(M, s, r_0)$ and the following scaling condition holds:

$$\frac{\log d}{\rho}\left(\frac{(\log d)^{19/6}}{n^{1/6}} + \frac{(\log d)^{11/6}}{\rho^{1/3} n^{1/6}} + \zeta_1 \log d + \frac{\zeta_2}{\rho}\right) + \zeta_2 d^4 + \frac{\log d}{\rho d_0} + \frac{(\log d)^2 |S|}{\rho d_0^2} = o(1),$$

where the definitions of d_0 and ρ are the same as in Section 5.

We then have the following theorem on the FDR control for the general graphical models.

THEOREM 6.3. *Under Assumptions 6.1 and 6.2, the StarTrek procedure in Algorithm 2 with the generic estimator $\widetilde{\mathbf{\Theta}}$ and the approximated quantiles (6.1) satisfies the following: as $(n, d) \to \infty$,*

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

The proof can be found in Appendix A.5.

7. Numerical results. In this section, we conduct simulation studies to complement the main theoretical claims of the paper and demonstrate the empirical performance of our method. Section 7.1 presents numerical results of applying the StarTrek filter to the multitask regression problem in Section 4. Section 7.2 focuses on the Gaussian graphical models studied in Section 5. In Section 7.3, we numerically compare our method to the grid search based on the skip-down method and demonstrate the computational advantages of the StarTrek filter. We also study the power performance of our approach against three competitor testing methods.

7.1. *Simulations for multitask regression.* Section 4 considers the application of the StarTrek filter to the multitask regression problem and provides theoretical results on FDP/FDR control. Here, we conduct some simulation studies. The synthetic datasets are generated from the multitask regression model described in (4.1). We sample the covariates from a Gaussian autoregressive model of order 1 (AR(1)) and choose the noise variance to be 1 for all responses. Now we describe how to generate the parameter matrix. First, the number of nonzero coefficients for each row is independently uniformly sampled from the integers between 0 and 20. Then the locations of nonzero coefficients are independently uniformly drawn from among the covariates. Finally, the values of nonzero coefficients are taking uniform random signs and identical magnitudes of 1. Throughout the simulated examples, we fix the number of responses and the number of covariates ($d_1 = d_2 = 300$) and vary the sample size n and the autocorrelation coefficient of the AR(1) design. We also run the selection procedure under two choices of the nominal FDR level, that is, $q \in \{0.1, 0.2\}$. Given the sparsity level of each response is uniformly distributed over integers between 0 and 20, we choose the threshold k_τ for determining hub responses to be 19, which is roughly the upper 10% quantile of the sparsity level's distribution. To run the StarTrek filter, we exactly follow the procedures described in Section 4 to calculate the test statistics and the approximated quantiles. The involving estimation steps are based on [36, 74].

Table 1 shows that the FDRs of the StarTrek filter are all well controlled below the nominal levels for different sample sizes and autocorrelation coefficient $\bar{\sigma}$. From Table 2, we find that the power of the proposed method increases as the sample size grows and decreases as the covariates become more dependent (i.e., with higher autocorrelations).

7.2. *Simulations for Gaussian graphical models.* In this section, we provide simulations results for Section 5. 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 (subgraph), the way of assigning edges for different graph types will be explained below in detail. After determining the adjacency

TABLE 1
Empirical FDR for the multitask regression problem. We set $d_1 = d_2 = 300$ *and* $n = 150, 200, 250$ *and* 300. *The autocorrelaiton* $\bar{\sigma}$ *varies from* 0.3 *to* 0.7

	$q = 0.1$				$q = 0.2$			
n	150	200	250	300	150	200	250	300
$\bar{\sigma} = 0.3$	0.0656	0.0416	0.0162	0.0184	0.0991	0.0676	0.0269	0.0355
$\bar{\sigma} = 0.4$	0.0638	0.0355	0.0144	0.0158	0.1006	0.0577	0.0252	0.0300
$\bar{\sigma} = 0.5$	0.0554	0.0376	0.0177	0.0179	0.0827	0.0532	0.0253	0.0349
$\bar{\sigma} = 0.6$	0.0525	0.0316	0.0144	0.0155	0.0762	0.0516	0.0257	0.0270
$\bar{\sigma} = 0.7$	0.0406	0.0454	0.0233	0.0224	0.0557	0.0662	0.0464	0.0385

TABLE 2
Empirical power for the multitask regression. We set $d_1 = d_2 = 300$ *and* $n = 150, 200, 250$ *and* 300. *The autocorrelaiton* $\bar{\sigma}$ *is chosen from* 0.3 *to* 0.7

	$q = 0.1$				$q = 0.2$			
n	150	200	250	300	150	200	250	300
$\bar{\sigma} = 0.3$	0.8902	1.0000	1.0000	1.0000	0.9206	1.0000	1.0000	1.0000
$\bar{\sigma} = 0.4$	0.8090	1.0000	1.0000	1.0000	0.8481	1.0000	1.0000	1.0000
$\bar{\sigma} = 0.5$	0.6563	0.9912	1.0000	1.0000	0.7081	0.9953	1.0000	1.0000
$\bar{\sigma} = 0.6$	0.4058	0.9549	0.9976	1.0000	0.4590	0.9622	0.9990	1.0000
$\bar{\sigma} = 0.7$	0.1215	0.7119	0.9678	0.9965	0.1578	0.7621	0.9778	0.9995

matrix of the graph, we follow [87] 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 (subgraph) for four different graph patterns.

- *Hub graph.* We randomly pick one node as the hub node of the subgraph, then the rest of the nodes are made to connect with this hub node. There is no edge between the nonhub nodes.
- *Random graph.* This is the Erdős–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 (5.1). 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. All results (of FDR and power) are averaged over 64 independent replicates.

As we can see from Table 3, the FDRs of the StarTrek filter for different types of graph are well controlled below the nominal levels. In the hub graph, the FDRs are relatively small but the power is still pretty good. A similar phenomenon for the multiple edge testing problem is observed [46]. 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.

TABLE 3
Empirical FDR for Gaussian graphical models

$d = 300$	$q = 0.1$			$q = 0.2$		
n	200	300	400	200	300	400
			$p = 20$			
hub	0.0000	0.0007	0.0000	0.0018	0.0016	0.0015
random	0.0186	0.0329	0.0438	0.0438	0.0727	0.0851
scale-free	0.0091	0.0243	0.0259	0.0265	0.0480	0.0579
knn	0.0103	0.0288	0.0345	0.0275	0.0648	0.0736
			$p = 30$			
hub	0.0012	0.0017	0.0000	0.0031	0.0039	0.0036
random	0.0464	0.0498	0.0478	0.0874	0.0969	0.0911
scale-free	0.0205	0.0326	0.0271	0.0414	0.0602	0.0580
knn	0.0216	0.0475	0.0431	0.0551	0.0909	0.0883

The power performance of the StarTrek filter is shown in Table 4. 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 hub nodes in general due to the way of constructing the graphs, and we find that the power is higher. Among different types of graphs, the power in the hub graph and scale-free graph is higher than that in the random and knn graph since the latter two are relatively denser and have more complicated topological structures.

In Figure 2, 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 that the height of the short blue solid 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 that 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

TABLE 4
Empirical power for Gaussian graphical models

$d = 300$	$q = 0.1$			$q = 0.2$		
n	200	300	400	200	300	400
			$p = 20$			
hub	0.6789	0.9406	0.9812	0.7727	0.9609	0.9867
random	0.3445	0.7734	0.9390	0.4637	0.8413	0.9592
scale-free	0.4799	0.8050	0.9347	0.5549	0.8479	0.9545
knn	0.1337	0.5689	0.8381	0.2254	0.6913	0.8920
			$p = 30$			
hub	0.6861	0.9242	0.9736	0.7497	0.9405	0.9810
random	0.5136	0.8728	0.9741	0.6027	0.9085	0.9842
scale-free	0.6296	0.8975	0.9778	0.7060	0.9230	0.9842
knn	0.2442	0.7036	0.8990	0.3396	0.7799	0.9335

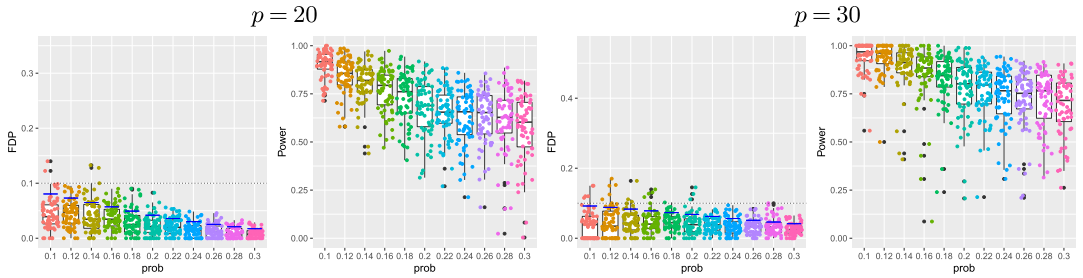

FIG. 2. *FDP and power plots for the StarTrek filter in the random graph. The connecting probability is varied on the x-axis. The number of samples n is chosen to be* 300 *and the number of connected components p equals* 20 *and* 30. *The nominal FDR level is set to be $q = 0.1$; the short blue solid 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.*

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.

7.3. *Comparison with the grid search based on the skip-down method.* In this section, we empirically compare the performance of two methods: our StarTrek filter (Algorithm 2) and the BHq procedure with Algorithm 1 (referred simply as Algorithm 1 without causing confusion). To implement Algorithm 1, we estimate $\widehat{\alpha}$ in (2.3) by a grid search of the suprema on a evenly spaced grid over $(0, q)$ with 7 spacing sizes: $0.1, 0.05, 0.02, 0.01, 0.005, 0.002, 0.001$. Note that the smaller spacing sizes correspond to higher granularity levels. We can see that the computation complexity of Algorithm 2 is $O(dk_\tau + d^2 \log d)$ and the time complexity is $O(d^2 k_\tau / g)$ for Algorithm 1 where g is the grid spacing size. In additional to the computational differences, we shall note that the selected hub node set from the grid search method must be a subset of that from the StarTrek filter, and as the grid becomes sufficiently granular (i.e., the gird spacing size becomes sufficiently small), the selected hub node sets from the two methods will be the same. To illustrate such points in our empirical comparison, we will additionally compute a relative version of FDR and power. Specifically, we compute the FDR and power for both methods but treat the selected hub node set from the StarTrek filter as the true hub node set. We follow Section 7.2 to generate the synthetic data and consider exactly the same settings in Table 3. The results are then visualized in Figure 3. First, we see that the relative FDR is always 0 and the relative power approaches 1 as the grid spacing size decreases to 0, which illustrates the equivalence of two algorithms. In terms of power and computational performances, we find that StarTrek filter achieves higher power than Algorithm 1 when the granularity level is coarse. By making the grid sufficiently granular, the grid search method can attain comparable power but cost much longer computational time than the StarTrek filter, and hence demonstrate the superiority of our proposed method.

7.4. *Comparison with other testing procedures.* This section compares the performance of the StarTrek filter against some other testing procedures. Three competitor methods are considered:

1. Method 1 computes the p-values with respect to testing $\mathbf{\Theta}_{jk} = 0$ for all the $(d^2 - d)/2$ pairs of (j, k). Then it adopts the canonical FWER control method to select the significant edges and count the selected edges for each row/column to determine whether each node is selected to be a hub node.
2. Method 2 computes all the p-values as in Method 1, but changes the way of applying FWER control adjustment. For each node j, it applies the Bonferroni procedure to the $d - 1$

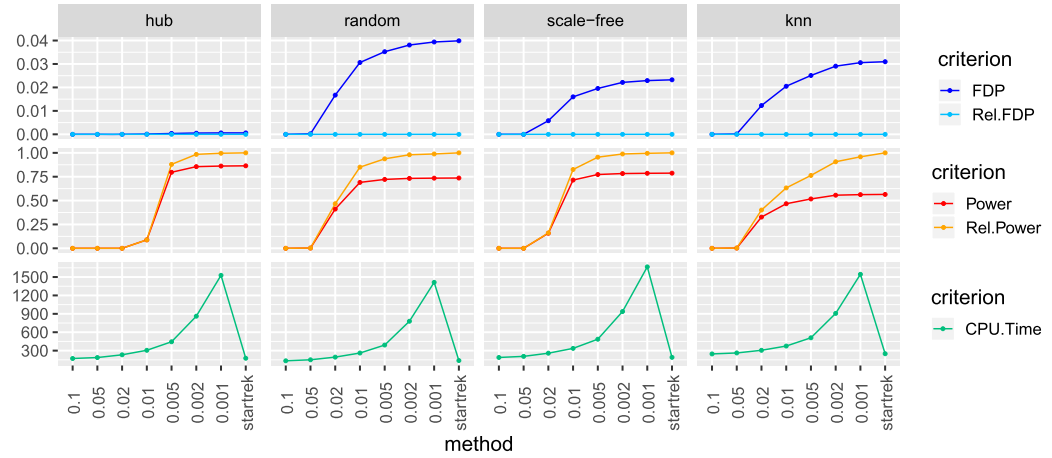

FIG. 3. *Empirical comparison of the StarTrek filter with the grid search based on the skip-down method. On the x-axis, we have the grid method with* 7 *granularity levels and our proposed StarTrek filter. The top panel compares FDR and relative FDR; the middle panel compares power and relative power; the bottom panel compares the CPU time (in seconds). Four columns correspond to four different graph patterns. For each graph pattern, we average those criterion quantities over* 6 *different settings (i.e.,* 3 *choices of* n *and* 2 *choices of* p*) where* 64 *independent replicates are simulated for each setting. The FDR nominal level* q *is* 0.1.

p-values corresponding to the jth column of the precision matrix, resulting in the node p-value. Then the BHq procedure is further applied to these node p-values to select the hub nodes.

3. Method 3 utilizes the node p-values computed from Algorithm 2 but applies the BY procedure [7] instead of the BHq procedure [6].

All three competitor methods are more conservative than the StarTrek filter since they are either only based on continuous edge testing procedures or not adapting to the complex dependence structures. We follow Section 7.2 to generate the synthetic data and consider exactly the same settings in Table 3, which involve 3 choices of the sample size n, 2 choices of p and 4 different types of graph patterns.

In Figure 4, we visualize the performances of the StarTrek filter against the above three competitor methods in terms of FDR and power. To understand how the set of selected hub nodes produced from each competitor method is similar/different to that from the StarTrek filter, we also calculate a relative version of FDR and power similarly as in Section 7.3 and the Jaccard index [32]. We find that our proposed StarTrek filter is less conservative and more powerful than all the three competitor methods, among which Method 1 is the most conservative method and Method 3 has the most similar selected hub node set to the StarTrek filter.

7.5. *Application to gene expression data.* We also apply our method to the Genotype-Tissue Expression (GTEx) data studied in [52]. 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 nondiseased 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.

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 for female individuals. The original count data is log-transformed and

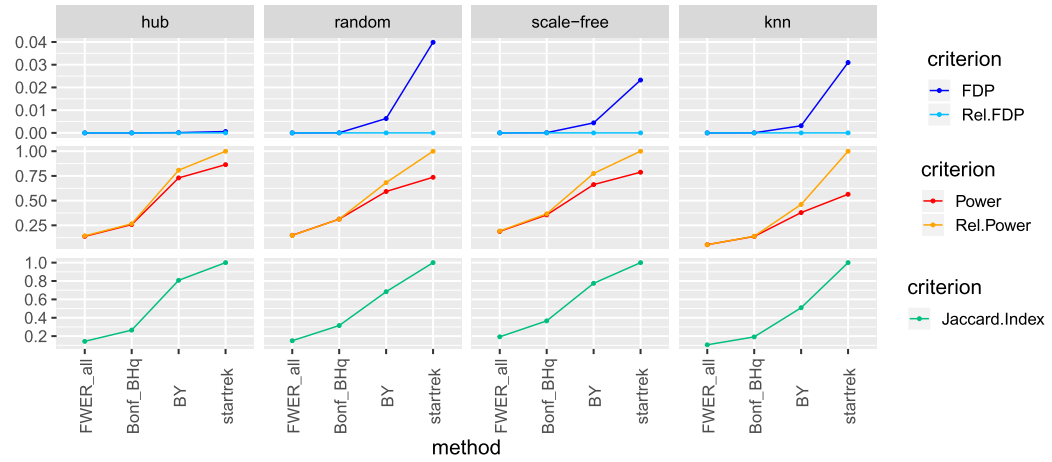

FIG. 4. *Empirical comparison of the StarTrek filter with the three competitor methods. On the x-axis, we have Method* 1 (*FWER_all*), *Method* 2 (*Bonf_BHq*), *Method* 3 (*BY*) *and our proposed StarTrek filter. The top panel compares FDR and relative FDR*; *the middle panel compares power and relative power*; *the bottom panel compares the Jaccard index* (*with respect to the set of selected hub nodes by the StarTrek filter*). *Four columns correspond to four different graph patterns. For each graph pattern, we average those criterion quantities over* 6 *different settings* (*i.e.*, 3 *choices of n and* 2 *choices of p*) *where* 64 *independent replicates are simulated for each setting. The FDR nominal level q is* 0.1.

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.

Figure 5 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 6, the results for the male and female dataset agree with each other except that the number of selected hub genes using the 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 [1,

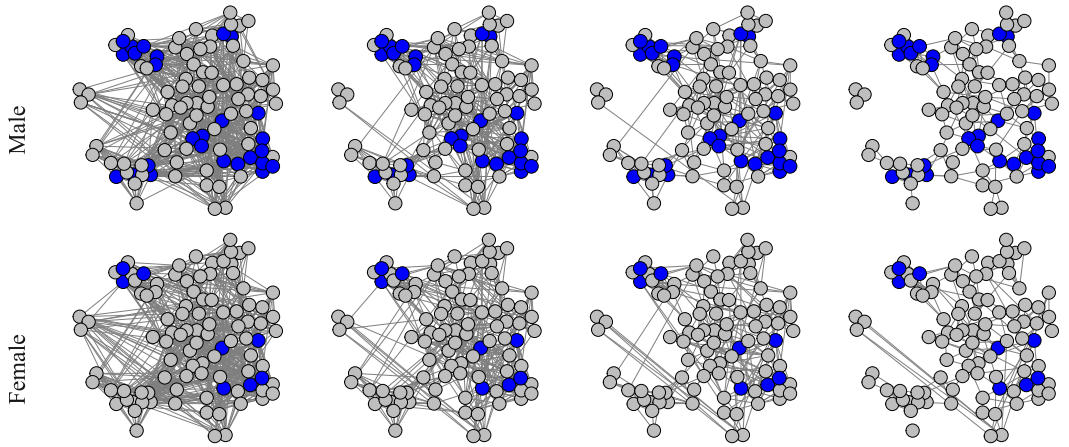

FIG. 5. *The above graphs are based the estimated precision matrices* (*the left two plots*). *The adjacency matrices of the other six plots are based on the standardized estimated precision matrices but thresholded at* 0.025, 0.05, 0.075, *respectively. Blue vertices represent the selected hub genes.*

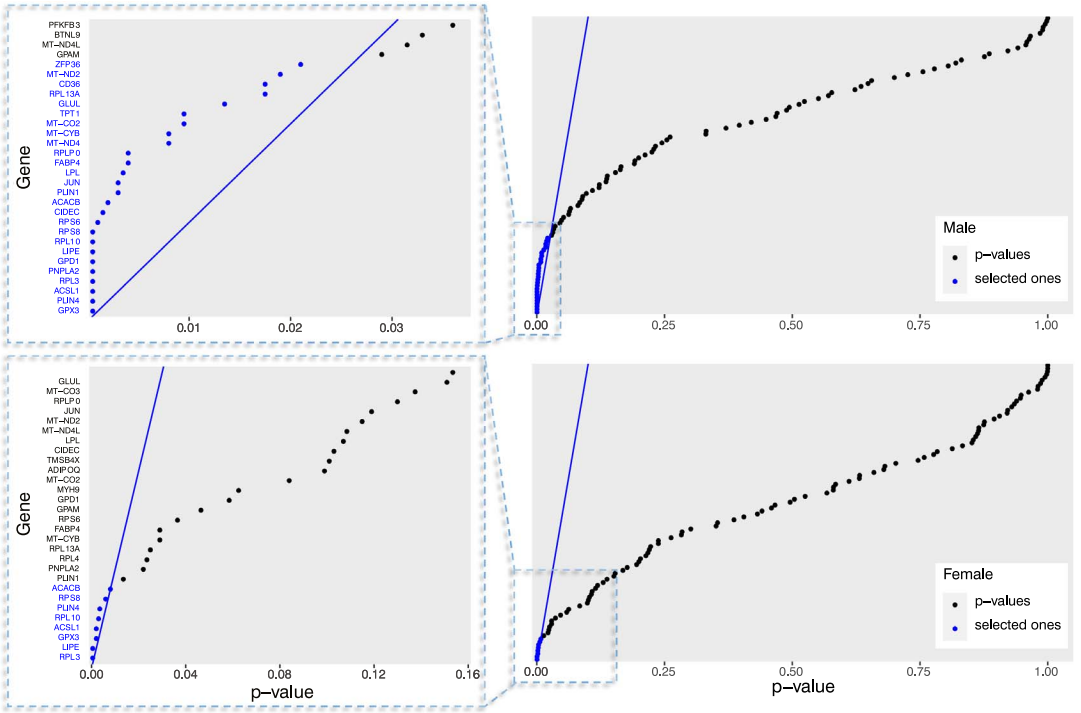

FIG. 6. *Plots of the sorted p-values* $(\alpha_j, j \in [d])$ *in Algorithm 2. Those blue points correspond to selected hub genes. The blue line is the rejection line of the BHq procedure. The coordinates of the plots are flipped. We abbreviate the names of the* 100 *genes and only show selected ones with blue colored text.*

8, 15, 30, 44, 53, 56, 57, 59, 72]. 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 a valuable prognosis and treatment signature.

8. Discussions. In this paper, we have proposed a novel method to select the hub nodes in the graph with degrees larger than a certain thresholding level. To show the validity of the method, we prove Cramér-type Gaussian comparison bounds with two types of covariance matrix differences and Cramér-type deviation results of the Gaussian multiplier bootstrap procedure. The extension of our results to other bootstrap methods is interesting for future research. In specific, [23] generalizes the Kolmogorov distance results of the Gaussian multiplier bootstrap [16] to the wild bootstrap and empirical bootstrap by proposing new comparison bounds and anticoncentration inequalities. Their techniques have the potential to be extended to the Cramér-type deviation bounds in the future. Moreover, [19] showed a faster rate of Kolmogorov distance consistency of the Gaussian multiplier bootstrap and it could be extended to Cramér-type deviation bounds to improve the rates in our paper as well.

Funding. The authors are grateful for the support of NSF DMS-1916211, NIH R35 CA220523, NIH R01 ES32418 and NIH U01CA209414.

SUPPLEMENTARY MATERIAL

Supplement to "StarTrek: Combinatorial variable selection with false discovery rate control". (DOI: 10.1214/23-AOS2296SUPP; .pdf). Supplementary information.

REFERENCES

[1] 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. *Oncol. Lett.* **18** 5821–5830.

[2] BARBER, R. F. and CANDÈS, E. J. (2015). Controlling the false discovery rate via knockoffs. *Ann. Statist.* **43** 2055–2085. MR3375876 https://doi.org/10.1214/15-AOS1337

[3] BARBER, R. F. and CANDÈS, E. J. (2019). A knockoff filter for high-dimensional selective inference. *Ann. Statist.* **47** 2504–2537. MR3988764 https://doi.org/10.1214/18-AOS1755

[4] BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Rev. Econ. Stud.* **81** 608–650. MR3207983 https://doi.org/10.1093/restud/rdt044

[5] BENJAMINI, Y. (2010). Discovering the false discovery rate. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 405–416. MR2758522 https://doi.org/10.1111/j.1467-9868.2010.00746.x

[6] BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. MR1325392

[7] BENJAMINI, Y. and YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165–1188. MR1869245 https://doi.org/10.1214/aos/1013699998

[8] BLEIN, S., BARJHOUX, L., GENESIS INVESTIGATORS, DAMIOLA, F., DONDON, M.-G., EON-MARCHAIS, S., MARCOU, M., CARON, O., LORTHOLARY, A. et al. (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.

[9] CAI, T., LIU, W. and LUO, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.* **106** 594–607. MR2847973 https://doi.org/10.1198/jasa.2011.tm10155

[10] CAI, T., LIU, W. and XIA, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J. Amer. Statist. Assoc.* **108** 265–277. MR3174618 https://doi.org/10.1080/01621459.2012.758041

[11] CAI, T. T. and MA, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19** 2359–2388. MR3160557 https://doi.org/10.3150/12-BEJ455

[12] CAI, T. T. and ZHANG, A. (2016). Inference for high-dimensional differential correlation matrices. *J. Multivariate Anal.* **143** 107–126. MR3431422 https://doi.org/10.1016/j.jmva.2015.08.019

[13] CANDÈS, E., FAN, Y., JANSON, L. and LV, J. (2018). Panning for gold: 'model-X' knockoffs for high dimensional controlled variable selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 551–577. MR3798878 https://doi.org/10.1111/rssb.12265

[14] CHANG, J., SHAO, Q.-M. and ZHOU, W.-X. (2016). Cramér-type moderate deviations for Studentized two-sample U-statistics with applications. *Ann. Statist.* **44** 1931–1956. MR3546439 https://doi.org/10.1214/15-AOS1375

[15] 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.

[16] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. MR3161448 https://doi.org/10.1214/13-AOS1161

[17] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Anti-concentration and honest, adaptive confidence bands. *Ann. Statist.* **42** 1787–1818. MR3262468 https://doi.org/10.1214/14-AOS1235

[18] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields* **162** 47–70. MR3350040 https://doi.org/10.1007/s00440-014-0565-9

[19] CHERNOZHUOKOV, V., CHETVERIKOV, D., KATO, K. and KOIKE, Y. (2022). Improved central limit theorem and bootstrap approximations in high dimensions. *Ann. Statist.* **50** 2562–2586. MR4500619 https://doi.org/10.1214/22-aos2193

[20] DAI, C., LIN, B., XING, X. and LIU, J. S. (2022). False discovery rate control via data splitting. *J. Amer. Statist. Assoc.* 1–18. https://doi.org/10.1080/01621459.2022.2060113

[21] DAI, C., LIN, B., XING, X. and LIU, J. S. (2023). A scale-free approach for false discovery rate control in generalized linear models. *J. Amer. Statist. Assoc.* **118** 1551–1565. MR4646583 https://doi.org/10.1080/01621459.2023.2165930

[22] DAI, R. and BARBER, R. (2016). The knockoff filter for FDR control in group-sparse and multitask regression. In *International Conference on Machine Learning* 1851–1859. PMLR.

[23] DENG, H. and ZHANG, C.-H. (2020). Beyond Gaussian approximation: Bootstrap for maxima of sums of independent random vectors. *Ann. Statist.* **48** 3643–3671. MR4185823 https://doi.org/10.1214/20-AOS1946

[24] DING, X. and ZHOU, Z. (2020). Estimation and inference for precision matrices of nonstationary time series. *Ann. Statist.* **48** 2455–2477. MR4134802 https://doi.org/10.1214/19-AOS1894

[25] EISENACH, C., BUNEA, F., NING, Y. and DINICU, C. (2020). High-dimensional inference for cluster-based graphical models. *J. Mach. Learn. Res.* **21** Paper No. 53, 55. MR4095332

[26] 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* 654–663. PMLR.

[27] FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441. https://doi.org/10.1093/biostatistics/kxm045

[28] GRONWALL, T. H. (1919). Note on the derivatives with respect to a parameter of the solutions of a system of differential equations. *Ann. of Math.* (2) **20** 292–296. MR1502565 https://doi.org/10.2307/1967124

[29] GU, Q., CAO, Y., NING, Y. and LIU, H. (2015). Local and global inference for high dimensional nonparanormal graphical models. Preprint. Available at arXiv:1502.02347.

[30] HELLWIG, B., MADJAR, K., EDLUND, K., MARCHAN, R., CADENAS, C., HEIMES, A.-S., ALMSTEDT, K., LEBRECHT, A., SICKING, I. et al. (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. https://doi.org/10.1371/journal.pone.0167585

[31] 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* 561–565. IEEE.

[32] JACCARD, P. (1901). Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. *Bull. Soc. Vaud. Sci. Nat.* **37** 241–272.

[33] JANKOVÁ, J. and VAN DE GEER, S. (2017). Honest confidence regions and optimality in high-dimensional precision matrix estimation. *TEST* **26** 143–162. MR3613609 https://doi.org/10.1007/s11749-016-0503-5

[34] JAVANMARD, A. and JAVADI, H. (2019). False discovery rate control via debiased lasso. *Electron. J. Stat.* **13** 1212–1253. MR3935848 https://doi.org/10.1214/19-ejs1554

[35] 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*) 1427–1434. IEEE.

[36] JAVANMARD, A. and MONTANARI, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15** 2869–2909. MR3277152

[37] JAVANMARD, A. and MONTANARI, A. (2014). Hypothesis testing in high-dimensional regression under the Gaussian random design model: Asymptotic theory. *IEEE Trans. Inf. Theory* **60** 6522–6554. MR3265038 https://doi.org/10.1109/TIT.2014.2343629

[38] JIN, J., KE, Z. T., LUO, S. and WANG, M. (2023). Optimal Estimation of the Number of Network Communities. *J. Amer. Statist. Assoc.* **118** 2101–2116.

[39] KE, Z. T., MA, Y. and LIN, X. (2023). Estimation of the number of spiked eigenvalues in a covariance matrix by bulk eigenvalue matching analysis. *J. Amer. Statist. Assoc.* **118** 374–392.

[40] KUCHIBHOTLA, A. K., MUKHERJEE, S. and BANERJEE, D. (2021). High-dimensional CLT: Improvements, non-uniform extensions and large deviations. *Bernoulli* **27** 192–217. MR4177366 https://doi.org/10.3150/20-BEJ1233

[41] LAM, C. and FAN, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. Statist.* **37** 4254–4278. MR2572459 https://doi.org/10.1214/09-AOS720

[42] LEE, R. K.-W., HOANG, T.-A. and LIM, E.-P. (2019). Discovering hidden topical hubs and authorities across multiple online social networks. *IEEE Trans. Knowl. Data Eng.* **33** 70–84.

[43] LI, J. and MAATHUIS, M. H. (2021). GGM knockoff filter: False discovery rate control for Gaussian graphical models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 534–558. MR4294543 https://doi.org/10.1111/rssb.12430

[44] LI, Y., GIORGI, E. E., BECKMAN, K. B., CABERTO, C., KAZMA, R., LUM-JONES, A., HAIMAN, C. A., MARCHAND, L. L., STRAM, D. O. et al. (2019). Association between mitochondrial genetic variation and breast cancer risk: The multiethnic cohort. *PLoS ONE* **14** 1–14. https://doi.org/10.1371/journal.pone.0222284

[45] LIU, M., XIA, Y., CHO, K. and CAI, T. (2021). Integrative high dimensional multiple testing with heterogeneity under data sharing constraints. *J. Mach. Learn. Res.* **22** Paper No. 126, 26. MR4279777

[46] Liu, W. (2013). Gaussian graphical model estimation with false discovery rate control. *Ann. Statist.* **41** 2948–2978. MR3161453 https://doi.org/10.1214/13-AOS1169
[47] Liu, W. and Luo, S. (2014). Hypothesis testing for high-dimensional regression models.
[48] 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. *Ann. Statist.* **38** 1913–1935. MR2662363 https://doi.org/10.1214/09-AOS774
[49] Liu, W. and Shao, Q.-M. (2014). Phase transition and regularized bootstrap in large-scale t-tests with false discovery rate control. *Ann. Statist.* **42** 2003–2025. MR3262475 https://doi.org/10.1214/14-AOS1249
[50] Liu, Y., Gu, H.-Y., Zhu, J., Niu, Y.-M., Zhang, C. and Guo, G.-L. (2019). Identification of hub genes and key pathways associated with bipolar disorder based on weighted gene co-expression network analysis. *Front. Physiol.* **10** 1081.
[51] Liu, Y., Yi, Y., Wu, W., Wu, K. and Zhang, W. (2019). Bioinformatics prediction and analysis of hub genes and pathways of three types of gynecological cancer. *Oncol. Lett.* **18** 617–628.
[52] Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F. et al. (2013). The genotype-tissue expression (GTEx) project. *Nat. Genet.* **45** 580–585.
[53] 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 Int.* **20** 1–15.
[54] Lu, J., Neykov, M. and Liu, H. (2017). Adaptive inferential method for monotone graph invariants. Preprint. Available at arXiv:1707.09114.
[55] 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.
[56] 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. *Sci. Rep.* **9** 1–15.
[57] Marino, N., German, R., Rao, X., Simpson, E., Liu, S., Wan, J., Liu, Y., Sandusky, G., Jacobsen, M. et al. (2020). Upregulation of lipid metabolism genes in the breast prior to cancer diagnosis. *npj Breast Cancer* **6** 1–13.
[58] Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. MR2278363 https://doi.org/10.1214/009053606000000281
[59] 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. *Oxid. Med. Cell. Longev.* **2014**.
[60] Newman, M. E., Watts, D. J. and Strogatz, S. H. (2002). Random graph models of social networks. *Proc. Natl. Acad. Sci. USA* **99** 2566–2572.
[61] Neykov, M., Lu, J. and Liu, H. (2019). Combinatorial inference for graphical models. *Ann. Statist.* **47** 795–827. MR3909951 https://doi.org/10.1214/17-AOS1650
[62] Obozinski, G., Taskar, B. and Jordan, M. (2006). Multi-task feature selection. Statistics Department, UC Berkeley, Tech. Rep. **2** 2.
[63] Peng, J., Zhou, N. and Zhu, J. (2009). Partial correlation estimation by joint sparse regression models. *J. Amer. Statist. Assoc.* **104** 735–746. MR2541591 https://doi.org/10.1198/jasa.2009.0126
[64] 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. https://doi.org/10.1016/j.neuron.2013.07.035
[65] Ravikumar, P., Wainwright, M. J., Raskutti, G. and Yu, B. (2011). High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electron. J. Stat.* **5** 935–980. MR2836766 https://doi.org/10.1214/11-EJS631
[66] Ren, Z., Sun, T., Zhang, C.-H. and Zhou, H. H. (2015). Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *Ann. Statist.* **43** 991–1026. MR3346695 https://doi.org/10.1214/14-AOS1286
[67] Rothman, A. J., Bickel, P. J., Levina, E. and Zhu, J. (2008). Sparse permutation invariant covariance estimation. *Electron. J. Stat.* **2** 494–515. MR2417391 https://doi.org/10.1214/08-EJS176
[68] Rubinov, M. and Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage* **52** 1059–1069. https://doi.org/10.1016/j.neuroimage.2009.10.003
[69] Shaw, P., Kabani, N. J., Lerch, J. P., Eckstrand, K., Lenroot, R., Gogtay, N., Greenstein, D., Clasen, L., Evans, A. et al. (2008). Neurodevelopmental trajectories of the human cerebral cortex. *J. Neurosci.* **28** 3586–3594. https://doi.org/10.1523/JNEUROSCI.5309-07.2008
[70] Shen, S. and Lu, J. (2023). Combinatorial-probabilistic trade-off: P-values of community property test in the stochastic block models. *IEEE Trans. Inf. Theory* **69** 6605–6618. MR4650324

[71] SHEN, X., PAN, W. and ZHU, Y. (2012). Likelihood-based selection and sharp parameter estimation. *J. Amer. Statist. Assoc.* **107** 223–232. MR2949354 https://doi.org/10.1080/01621459.2011.645783

[72] SIROIS, I., AGUILAR-MAHECHA, A., LAFLEUR, J., FOWLER, E., VU, V., SCRIVER, M., BUCHANAN, M., CHABOT, C., RAMANATHAN, A. et al. (2019). A unique morphological phenotype in chemoresistant triple-negative breast cancer reveals metabolic reprogramming and PLIN4 expression as a molecular vulnerability. *Mol. Cancer Res.* **17** 2492–2507. https://doi.org/10.1158/1541-7786.MCR-19-0264

[73] SLEPIAN, D. (1962). The one-sided barrier problem for Gaussian noise. *Bell Syst. Tech. J.* **41** 463–501. MR0133183 https://doi.org/10.1002/j.1538-7305.1962.tb02419.x

[74] SUN, T. and ZHANG, C.-H. (2012). Scaled sparse linear regression. *Biometrika* **99** 879–898. MR2999166 https://doi.org/10.1093/biomet/ass043

[75] SUR, P. and CANDÈS, E. J. (2019). A modern maximum-likelihood theory for high-dimensional logistic regression. *Proc. Natl. Acad. Sci. USA* **116** 14516–14525. MR3984492 https://doi.org/10.1073/pnas.1810420116

[76] VAN DE GEER, S., BÜHLMANN, P., RITOV, Y. and DEZEURE, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42** 1166–1202. MR3224285 https://doi.org/10.1214/14-AOS1221

[77] VAN DEN HEUVEL, M. P. and SPORNS, O. (2013). Network hubs in the human brain. *Trends Cogn. Sci.* **17** 683–696. https://doi.org/10.1016/j.tics.2013.09.012

[78] 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. MR3371002 https://doi.org/10.1093/biomet/asu074

[79] 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. *J. Amer. Statist. Assoc.* **113** 328–339. MR3803468 https://doi.org/10.1080/01621459.2016.1251930

[80] XING, X., ZHAO, Z. and LIU, J. S. (2023). Controlling false discovery rate using Gaussian mirrors. *J. Amer. Statist. Assoc.* **118** 222–241. MR4571118 https://doi.org/10.1080/01621459.2021.1923510

[81] YANG, Z., NING, Y. and LIU, H. (2018). On semiparametric exponential family graphical models. *J. Mach. Learn. Res.* **19** Paper No. 57, 59. MR3899759

[82] YU, M., GUPTA, V. and KOLAR, M. (2020). Simultaneous inference for pairwise graphical models with generalized score matching. *J. Mach. Learn. Res.* **21** Paper No. 91, 51. MR4119159

[83] 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.

[84] YUAN, M. and LIN, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* **94** 19–35. MR2367824 https://doi.org/10.1093/biomet/asm018

[85] ZHANG, L. and LU, J. (2024). Supplement to "StarTrek: Combinatorial variable selection with false discovery rate control." https://doi.org/10.1214/23-AOS2296SUPP

[86] ZHAO, Q., SUR, P. and CANDÈS, E. J. (2022). The asymptotic distribution of the MLE in high-dimensional logistic models: Arbitrary covariance. *Bernoulli* **28** 1835–1861. MR4411513 https://doi.org/10.3150/21-bej1401

[87] ZHAO, T., LIU, H., ROEDER, K., LAFFERTY, J. and WASSERMAN, L. (2012). The `huge` package for high-dimensional undirected graph estimation in R. *J. Mach. Learn. Res.* **13** 1059–1062. MR2930633