The Annals of Statistics
2019, Vol. 47, No. 2, 795–827
https://doi.org/10.1214/17-AOS1650
© Institute of Mathematical Statistics, 2019

COMBINATORIAL INFERENCE FOR GRAPHICAL MODELS

BY MATEY NEYKOV[1], JUNWEI LU[1] AND HAN LIU[2]

Carnegie Mellon University, Princeton University and Northwestern University

We propose a new family of combinatorial inference problems for graphical models. Unlike classical statistical inference where the main interest is point estimation or parameter testing, combinatorial inference aims at testing the global structure of the underlying graph. Examples include testing the graph connectivity, the presence of a cycle of certain size, or the maximum degree of the graph. To begin with, we study the information-theoretic limits of a large family of combinatorial inference problems. We propose new concepts including structural packing and buffer entropies to characterize how the complexity of combinatorial graph structures impacts the corresponding minimax lower bounds. On the other hand, we propose a family of novel and practical structural testing algorithms to match the lower bounds. We provide numerical results on both synthetic graphical models and brain networks to illustrate the usefulness of these proposed methods.

1. Introduction. Graphical model is a powerful tool for modeling complex relationships among many random variables. A central theme of graphical model research is to infer the structure of the underlying graph based on observational data. Though significant progress has been made, existing works mainly focus on estimating the graph [Meinshausen and Bühlmann (2006), Liu, Lafferty and Wasserman (2009), Ravikumar et al. (2011), Cai, Liu and Luo (2011)] or testing the existence of a single edge [Janková and van de Geer (2015), Ren et al. (2015), Neykov et al. (2018b), Gu et al. (2015)].

In this paper, we consider a new inferential problem: testing the combinatorial structure of the underlying graph. Examples include testing the graph connectivity, cycle presence or assessing the maximum degree of the graph. Unlike classical inference which aims at testing a set of Euclidean parameters, combinatorial inference aims to test some global structural properties and requires the development of new methodology. As for methodological development, this paper mainly considers the Gaussian graphical model (though our method is applicable to the more general semiparametric exponential family graphical models and

Received August 2016; revised August 2017.

[1]These authors contributed equally to this work.

[2]Supported by NSF Grant DMS-1454377-CAREER; NSF Grant IIS1546482-BIGDATA; NIH Grant R01MH102339; NSF Grant IIS1408910; NIH Grant R01GM083084.

MSC2010 subject classifications. 62F03, 62F04, 62H15.

Key words and phrases. Graph structural inference, minimax testing, uncertainty assessment, multiple hypothesis testing, post-regularization inference.

elliptical copula graphical models): Let $\boldsymbol{X} = (X_1, \ldots, X_d)^T \sim N_d(\mathbf{0}, (\boldsymbol{\Theta}^*)^{-1})$ be a d-dimensional Gaussian random vector with precision matrix $\boldsymbol{\Theta}^* = (\Theta^*_{jk})$. Let $G^* = G(\boldsymbol{\Theta}^*) := (\overline{V}, E^*)$ be an undirected graph, where $\overline{V} = \{1, \ldots, d\}$ and an edge $(j, k) \in E^*$ if and only if $\Theta^*_{jk} \neq 0$. It is well known that G^* has the pairwise Markov property, that is, $(j, k) \notin E^*$ if and only if X_j and X_k are conditionally independent given the remaining variables. In a combinatorial inference problem, our goal is to test whether G^* has certain global structural properties based on n random samples $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. Specifically, let $\mathcal{G}$ be the set of all graphs over the vertex set $\overline{V}$ and $(\mathcal{G}_0, \mathcal{G}_1)$ be a pair of nonoverlapping subsets of $\mathcal{G}$. We assume all the graphs in $\mathcal{G}_1$ have a property (e.g., connectivity) while the graphs in $\mathcal{G}_0$ do not have this property. Such a pair $(\mathcal{G}_0, \mathcal{G}_1)$ is called a sub-decomposition of $\mathcal{G}$. Our goal is to test the hypothesis $\mathbf{H}_0 : G^* \in \mathcal{G}_0$ versus $\mathbf{H}_1 : G^* \in \mathcal{G}_1$. We provide several concrete examples below.

Connectivity. A graph is connected if and only if there exists a path connecting each pair of its vertices. To test connectivity, we set $\mathcal{G}_0 = \{G \in \mathcal{G} \mid G \text{ is disconnected}\}$ and $\mathcal{G}_1 = \{G \in \mathcal{G} \mid G \text{ is connected}\}$. Under the Gaussian graphical model, this is equivalent to testing whether the variables can be partitioned into at least two independent sets.

Cycle presence. Sometimes it is of interest to test whether the underlying graph is a forest. In this example we let $\mathcal{G}_0 = \{G \in \mathcal{G} \mid G \text{ is a forest}\}$ and $\mathcal{G}_1 = \{G \in \mathcal{G} \mid G \text{ contains a cycle}\}$. If a graph is a forest, it can be easily visualized on a two-dimensional plane.

Maximum degree. Another relevant question is to test whether the maximum degree of the graph is less than or equal to some integer $s_0 \in \mathbb{N}$ versus the alternative that the maximum degree is at least $s_1 \in \mathbb{N}$, where $s_0 < s_1$. Define the sub-decomposition $\mathcal{G}_0 = \{G \mid d_{\max}(G) \leq s_0\}$ and $\mathcal{G}_1 = \{G \mid d_{\max}(G) \geq s_1\}$, respectively.

While our ultimate goal is to test whether $G^* \in \mathcal{G}_0$ versus $G^* \in \mathcal{G}_1$, our access to G^* is only through the random samples $\{\boldsymbol{X}_i\}_{i=1}^n$. Under Gaussian models, we can translate the original problem of testing graphs to testing the precision matrix:

$$\mathbf{H}_0 : \boldsymbol{\Theta}^* \in \mathcal{S}_0 \quad \text{vs} \quad \mathbf{H}_1 : \boldsymbol{\Theta}^* \in \mathcal{S}_1. \tag{1.1}$$

In (1.1), $\mathcal{S}_0, \mathcal{S}_1 \subset \mathcal{M}(s)$ are two sets of precision matrices such that for all $\boldsymbol{\Theta} \in \mathcal{S}_0, \mathcal{S}_1$ we have $G(\boldsymbol{\Theta}) \in \mathcal{G}_0, \mathcal{G}_1$, respectively, and $\mathcal{M}(s)$ is defined as

$$\mathcal{M}(s) = \Big\{\boldsymbol{\Theta} \mid \boldsymbol{\Theta} = \boldsymbol{\Theta}^T, C^{-1} \leq \boldsymbol{\Theta} \leq C, \|\boldsymbol{\Theta}\|_1 \leq L, \max_{j \in [d]} \|\boldsymbol{\Theta}_{*j}\|_0 \leq s\Big\} \tag{1.2}$$

for some constants $1 \leq C \leq L$. The inequalities $C^{-1} \leq \boldsymbol{\Theta} \leq C$ in (1.2) are meant in a "positive-semidefinite sense," that is, the minimum and maximum eigenvalues of $\boldsymbol{\Theta}$ are assumed to be bounded by C^{-1} and C from below and above, respectively, and $\|\boldsymbol{\Theta}_{*j}\|_0$ is the cardinality of the nonzero entries of the jth column of $\boldsymbol{\Theta}$ (see

Section 1.3 for precise notation). The set $\mathcal{M}(s)$ restricts our attention to well-conditioned symmetric matrices $\mathbf{\Theta}$, whose induced graphs $G(\mathbf{\Theta})$ have maximum degree of at most s. Given this setup, we aim to characterize necessary conditions on the pair $\mathcal{S}_0, \mathcal{S}_1$ under which the combinatorial inference problem in (1.1) is *testable*. Specifically, recall that a test is any measurable function $\psi : \{\boldsymbol{X}_i\}_{i=1}^n \mapsto \{0, 1\}$. Define the minimax risk of testing $\mathcal{S}_0$ against $\mathcal{S}_1$ as

$$\gamma(\mathcal{S}_0, \mathcal{S}_1) = \inf_{\psi} \Big[\max_{\mathbf{\Theta} \in \mathcal{S}_0} \mathbb{P}_{\mathbf{\Theta}}(\psi = 1) + \max_{\mathbf{\Theta} \in \mathcal{S}_1} \mathbb{P}_{\mathbf{\Theta}}(\psi = 0) \Big]. \tag{1.3}$$

If $\liminf_{n \to \infty} \gamma(\mathcal{S}_0, \mathcal{S}_1) = 1$, we say that the problem (1.1) is *untestable* since any test fails to distinguish between $\mathcal{S}_0$ and $\mathcal{S}_1$ in the asymptotic minimax sense. We are specifically interested in an asymptotic setting where the dimension d is a function of the sample size, that is, $d = d(n)$ so that $d \to \infty$ as $n \to \infty$. This setting will be implicitly understood throughout the paper. Due to the close relationship between the sets of precision matrices $(\mathcal{S}_0, \mathcal{S}_1)$ and the sub-decomposition $(\mathcal{G}_0, \mathcal{G}_1)$ [recall that for all $\mathbf{\Theta} \in \mathcal{S}_0, \mathcal{S}_1$ we have $G(\mathbf{\Theta}) \in \mathcal{G}_0, \mathcal{G}_1$ resp.], we anticipate that the sub-decomposition can capture the intrinsic challenge of the test in (1.1). Indeed, in Sections 2 and 4 we develop a framework capable of capturing the impact of the combinatorial structures of $\mathcal{G}_0$ and $\mathcal{G}_1$ to the lower bound $\gamma(\mathcal{S}_0, \mathcal{S}_1)$. Such lower bounds provide necessary conditions for any valid test. We then develop practical procedures that match the obtained lower bounds. To understand how the sub-decomposition $(\mathcal{G}_0, \mathcal{G}_1)$ affects the intrinsic difficulty of the problem in (1.1), we consider the three examples given before. Our lower bound framework distinguishes between two types of sub-decompositions—in the first type, one can find graphs belonging to $\mathcal{G}_0$ and $\mathcal{G}_1$ differing in only one single edge, while in the second type all graphs belonging to $\mathcal{G}_0$ must differ on multiple edge sets from the graphs belonging to $\mathcal{G}_1$.

One can check that in the first two examples (connectivity and cycle presence testing) there exist graphs belonging to $\mathcal{G}_0$ and $\mathcal{G}_1$ differing in only one single edge. For instance, when testing connectivity, consider a tree with a single edge removed (thus it becomes a forest) versus a connected tree. Extending this intuition, for a fixed graph $G_0 = (\overline{V}, E_0) \in \mathcal{G}_0$, we call the edge set $\mathcal{C} = \{e_1, \ldots, e_m\}$ a single-edge *null-alternative divider*, or simply a divider for short, if for all edges $e \in \mathcal{C}$ the graphs $(\overline{V}, E_0 \cup \{e\}) \in \mathcal{G}_1$. Intuitively, the bigger the cardinality of a divider is, the harder it is to tell the null from the alternatives. In Section 2, we detail that $\gamma(\mathcal{S}_0, \mathcal{S}_1)$ is asymptotically 1, when the signal strength of separation between $\mathcal{S}_0$ and $\mathcal{S}_1$ is low [see (2.2) and (2.3) for a formal definition] and there exists a divider with sufficiently large packing number. The packing number, formally defined in Definition 2.3, represents the cardinality of a subset of edges in $\mathcal{C}$ which are "far" apart, where the proximity measure of two edges is a predistance (compared to distance, a predistance does not have to satisfy the triangle inequality) based on the graph G_0. Recall that for a graph G and two vertices u and v, a graph geodesic distance is defined by

$$d_G(u, v) := \text{length of the shortest path between } u \text{ and } v \text{ within } G.$$

Using the notion of geodesic distance, one can define a predistance between two edges, by taking the minimum over the geodesic distances of their corresponding nodes.

If the difference between null and alternative is more than one edge, as in the maximum degree testing $\mathcal{G}_0 = \{G \mid d_{\max}(G) \leq s_0\}$ versus $\mathcal{G}_1 = \{G \mid d_{\max}(G) \geq s_1\}$, for example, the packing number does not always capture the lower bound of the tests. In Section 4, we develop a novel mechanism to handle this more sophisticated case. We introduce a concept called "buffer entropy" which can overcome the disadvantages of the packing number and produce sharper lower bounds.

On the other hand, to match the lower bound, we propose the alternative witness test as a general algorithm for combinatorial testing. Our algorithm identifies a critical structure and proceeds to test whether this structure indeed belongs to the true graph. We prove that alternative witness tests can control both the type I and type II errors asymptotically.

1.1. *Contributions*. There are three major contributions of this paper.

Our first contribution is to develop a novel strategy for obtaining minimax lower bounds on the signal strength required to distinguish combinatorial graph structures which are separable via a single-edge divider. In particular, we relate the information-theoretic lower bounds to the packing number of the divider, which is an intuitive combinatorial quantity. To obtain this connection, we relate the chi-square divergence between two probability measures to the number of "closed walks" on their corresponding Markov graphs. Our analysis hinges on several technical tools including Le Cam's lemma, matrix perturbation inequalities and spectral graph theory. The usefulness of the approach is demonstrated by obtaining generic and interpretable lower bounds in numerous examples such as testing connectivity, connected components, self-avoiding paths and cycles.

Our second contribution is to provide a device for proving lower bounds under the settings where the null and alternative graphs differ in multiple edges. Under such case, the packing number does not always provide tight lower bounds. In order to overcome this issue we formalize a graph quantity called buffer entropy. The buffer entropy is a complexity measure of the structural tests and provides lower bounds. We apply buffer entropy to derive lower bounds for testing the maximum degree and detecting a sparse clique and cycles.

Our third contribution is to propose an alternative witness test (1.1), which matches the lower bounds on the signal strength. Our algorithm works on subdecompositions which are stable with respect to addition of edges, that is, given a graph $G \in \mathcal{G}_1$ adding edges to G yields graphs which belong to $\mathcal{G}_1$. The alternative witness test is a two step procedure—in the first step, it identifies a minimal structure "witnessing" the alternative hypothesis, and in the second step it attempts to certify the presence of this structure in the graph. The alternative witness test utilizes recent advances in high-dimensional inference and provides honest tests for combinatorial inference problems. It has two advantages compared to the sup-

port recovery procedures in Meinshausen and Bühlmann (2006), Ravikumar et al. (2011), Cai, Liu and Luo (2011): First, it allows us to control the type I error at any given level; second, it does not require perfect recovery of the underlying graph to conduct valid inference.

1.2. *Related work.* Graphical model inference is relatively straightforward when $d < n$, but becomes notoriously challenging when $d \gg n$. In high-dimensions, estimation procedures were studied by Yuan and Lin (2006), Friedman, Hastie and Tibshirani (2008), Lam and Fan (2009), Cai, Liu and Luo (2011) among others, while for variable selection procedures see Meinshausen and Bühlmann (2006), Raskutti et al. (2008), Liu, Lafferty and Wasserman (2009), Ravikumar et al. (2011), Cai, Liu and Luo (2011) and references therein. Recently, motivated by Zhang and Zhang (2014), various inferential methods for high-dimensional graphical models were suggested [e.g., Liu (2013), Janková and van de Geer (2015), Chen et al. (2016), Ren et al. (2015), Neykov et al. (2018b), Gu et al. (2015)], most of which focus on testing the presence of a single edge (except Liu (2013) who took the FDR approach [Benjamini and Hochberg (1995)] to conduct multiple tests and Gu et al. (2015) who developed procedures of edge testing in Gaussian copula models). None of the aforementioned works address the problem of combinatorial structure testing.

In addition to estimation and model selection procedures, efforts have been made to understand the fundamental limits of these problems. Lower bounds on estimation were obtained by Ren et al. (2015), where the authors show that the parametric estimation rate $n^{-1/2}$ is unattainable unless $s \log d/\sqrt{n} = o(1)$. Lower bounds on the minimal sample size required for model selection in Ising models were established by Santhanam and Wainwright (2012), where it is shown that support recovery is unattainable when $n \ll s^2 \log d$. In a follow-up work, Wang, Wainwright and Ramchandran (2010) studied model selection limits on the sample size in Gaussian graphical models. The latter two works are remotely related to ours, in that both works exploit graph properties to obtain information-theoretic lower bounds. However, our problem differs significantly from theirs since we focus on developing lower bounds for testing graph structure, which is a fundamentally different problem from estimating the whole graph.

Our problem is most closely related to those in Addario-Berry et al. (2010), Arias-Castro, Bubeck and Lugosi (2012, 2015), Arias-Castro et al. (2015), which are inspired by the large body of research on minimax hypothesis testing [e.g., Ingster (1982), Ingster, Tsybakov and Verzelen (2010), Arias-Castro, Candès and Durand (2011), Arias-Castro, Candès and Plan (2011)] among many others. Addario-Berry et al. (2010) quantify the signal strength as the mean parameter of a standard Gaussian distribution, while Arias-Castro, Bubeck and Lugosi (2012, 2015) impose models on the covariance matrix of a multivariate Gaussian distribution. In our setup, the parameter spaces of interest are designed to reflect the graphical model structure, and hence the signal strength is naturally imposed on the

precision matrix. Arias-Castro, Candès and Plan (2011) provide detection bounds for the linear model. This is related to our work since one can view a linear model with Gaussian design as a Gaussian graphical model. Arias-Castro et al. (2015) address testing on a lattice based Gaussian Markov random field. For specific problems they establish lower bounds on the signal strength required to test the empty graph versus an alternative hypothesis. This is different from the setting of our problems, where the null hypothesis is usually not the empty graph.

1.3. *Notation*. The following notation is used throughout the paper. For a vector $\mathbf{v} = (v_1, \ldots, v_d)^T \in \mathbb{R}^d$, let $\|\mathbf{v}\|_q = (\sum_{i=1}^d |v_i|^q)^{1/q}$, $1 \le q < \infty$, $\|\mathbf{v}\|_0 = |\mathrm{supp}(\mathbf{v})|$, where $\mathrm{supp}(\mathbf{v}) = \{j \mid v_j \ne 0\}$, and $|A|$ denotes the cardinality of a set A. Furthermore, let $\|\mathbf{v}\|_\infty = \max_i |v_i|$ and $\mathbf{v}^{\otimes 2} = \mathbf{v}\mathbf{v}^T$. For a matrix $\mathbf{A}$, we denote $\mathbf{A}_{*j}$ and $\mathbf{A}_{j*}$ to be the jth column and row of $\mathbf{A}$, respectively. For any $n \in \mathbb{N}$, we use the shorthand notation $[n] = \{1, \ldots, n\}$. For two integer sets $S_1, S_2 \subseteq [d]$, we denote $\mathbf{A}_{S_1 S_2}$ to be the sub-matrix of $\mathbf{A}$ with elements $\{A_{jk}\}_{j \in S_1, k \in S_2}$. Moreover, we denote $\|\mathbf{A}\|_{\max} = \max_{jk} |A_{jk}|$, $\|\mathbf{A}\|_p = \max_{\|\mathbf{v}\|_p = 1} \|\mathbf{A}\mathbf{v}\|_p$ for $p \ge 1$. For a symmetric matrix $\mathbf{A} \in \mathbb{R}^{d \times d}$ and a constants c, C, with a slight abuse of notation we write $c \le \mathbf{A} \le C$ to mean that the matrices $\mathbf{A} - c\mathbf{I}_d$ and $C\mathbf{I}_d - \mathbf{A}$ are positive semidefinite, where $\mathbf{I}_d$ denotes the $d \times d$ identity matrix.

For a graph G, we use $V(G)$, $E(G)$, $d_{\max}(G)$ to refer to the vertex set, edge set and maximum degree of G, respectively. We also denote $V(E)$ as the vertex set of the edge set E. We reserve special notation for the complete vertex set $\overline{V} := [d]$, the complete edge set $\overline{E} := \{e \in 2^{[d]} \mid |e| = 2\}$ and the complete graph $\overline{G} := (\overline{V}, \overline{E})$. For two integers $j, k \in [d]$, we use unordered pairs $(j, k) = (k, j)$ to denote undirected edges between vertex k and vertex j. Any symmetric matrix $\mathbf{A} \in \mathbb{R}^{d \times d}$ naturally induces an undirected graph $G(\mathbf{A})$, with vertices in the set $\overline{V}$ and edge set $E(G(\mathbf{A})) = \{(j, k) \mid A_{jk} \ne 0, j \ne k\}$. Additionally, if E is an arbitrary edge set (i.e., $E \subseteq \overline{E}$) for $e := (j, k) \in E$ we use the notation $A_e = A_{jk} = A_{kj}$ interchangeably to denote the element e of the matrix $\mathbf{A}$.

Given two sequences $\{a_n\}, \{b_n\}$, we write $a_n = O(b_n)$ if there exists a constant $C < \infty$ such that $a_n \le C b_n$; $a_n = o(b_n)$ if $a_n / b_n \to 0$, and $a_n \asymp b_n$ if there exists positive constants c and C such that $c < a_n / b_n < C$. Finally, we use the shorthand notation $\wedge$ and $\vee$ for min and max of two numbers, respectively.

1.4. *Organization of the paper.* The paper is structured as follows. A lower bound on single edge dividers along with applications to several examples is presented in Section 2. In Section 3, we outline the alternative witness test, and illustrate how to apply it to the examples considered in Section 2. In Section 4, we generalize the lower bounds strategies from the single edge divider to multiple edge divider stetting. A brief discussion is provided in Section 5. Full proof of the main result of Section 2 is presented in Section 6. Numerical studies, real data analysis and all remaining proofs are deferred to the Supplementary Material [Neykov, Lu and Liu (2018a)].

2. Single-edge null-alternative dividers. In this section, we derive a novel and generic lower bound strategy, applicable to null and alternative hypotheses which differ in one single edge: that is, under the Gaussian model, there exist two matrices $\mathbf{\Theta}_0 \in \mathcal{S}_0$ and $\mathbf{\Theta}_1 \in \mathcal{S}_1$ whose induced graphs $G_0 := G(\mathbf{\Theta}_0)$ and $G_1 := G(\mathbf{\Theta}_1)$ differ in a single edge. We formalize this concept in the definition below.

DEFINITION 2.1 (Single-edge null-alternative divider). For a sub-decomposition $(\mathcal{G}_0, \mathcal{G}_1)$ of $\mathcal{G}$, let $G_0 = (\overline{V}, E_0) \in \mathcal{G}_0$ be a graph under the null. We refer to an edge set $\mathcal{C} = \{e_1, \ldots, e_m\}$, as a (single-edge) *null-alternative divider* with the null base G_0 if for any $e \in \mathcal{C}$ the graphs $G_e := (\overline{V}, E_0 \cup \{e\}) \in \mathcal{G}_1$.

As remarked in the Introduction, if a large divider exists, it is expected that differentiating G_0 from an alternative graph $G_e \in \mathcal{G}_1$ is more challenging. Indeed, our main result of this section confirms this intuition. We proceed to define a predistance for a graph G and two edges e, e' [which need not belong to $E(G)$] which plays a key role in our lower bound result.

DEFINITION 2.2 (Edge geodesic predistance). Let $G = (\overline{V}, E)$ and $\{e, e'\}$ be a pair of edges (e and e' may or may not belong to E). We define

$$d_G(e, e') := \min_{u \in e, v \in e'} d_G(u, v),$$

where $d_G(u, v)$ denotes the geodesic distance between vertices u and v on the augmented graph G. If such a path does not exist, $d_G(e, e') = \infty$.

By definition, $d_G(e, e')$ is a predistance, that is, $d_G(e, e) = 0$ and $d_G(e, e') \geq 0$. Moreover, $d_G(e, e')$ has the same value regardless of whether $e, e' \in E(G)$. See Figure 1 for an illustration of $d_G(e, e')$. Inspired by the classical concept of packing entropy on metric spaces [e.g., Yang and Barron (1999)], we propose the structural packing entropy for graphs in an attempt to characterize information-theoretic lower bounds for combinatorial inference.

DEFINITION 2.3 (Structural packing entropy). Let $\mathcal{C}$ be a nonempty edge set and G be a graph. For any $r \geq 0$, we call the edge set $N_r \subset \mathcal{C}$ an r-packing of $\mathcal{C}$ if for any $e, e' \in N_r$ we have $d_G(e, e') \geq r$. Define the structural r-packing entropy as

$$M(\mathcal{C}, d_G, r) := \log \max\big\{|N_r| \mid N_r \subset \mathcal{C}, N_r \text{ is a } r\text{-packing of } \mathcal{C}\big\}. \tag{2.1}$$

The packing entropy in Definition 2.3 is an analog to the classical packing entropy on metric spaces in the sense that it is defined over an edge set $\mathcal{C}$ equipped with a predistance $d_G(e, e')$ based on the graph G.

To study minimax lower bounds, we only need to focus on the Gaussian graphical model whose structural properties are completely characterized by the precision matrices. We now formally define the sets of precision matrices $\mathcal{S}_0$ and $\mathcal{S}_1$ used in this section. Let

$$\mathcal{S}_0(\theta, s) := \Big\{ \mathbf{\Theta} \in \mathcal{M}(s) \mid G(\mathbf{\Theta}) \in \mathcal{G}_0, \min_{e \in E(G(\mathbf{\Theta}))} |\Theta_e| \geq \theta \Big\} \quad \text{and} \tag{2.2}$$

$$\mathcal{S}_1(\theta, s) := \Big\{ \mathbf{\Theta} \in \mathcal{M}(s) \mid G(\mathbf{\Theta}) \in \mathcal{G}_1, \min_{e \in E(G(\mathbf{\Theta}))} |\Theta_e| \geq \theta \Big\}, \tag{2.3}$$

where $\mathcal{M}(s)$ is defined in (1.2). The parameter θ in the definitions of $\mathcal{S}_0(\theta, s)$ and $\mathcal{S}_1(\theta, s)$ denotes the *signal strength*, and as we show below, its magnitude plays an important role in determining whether one can distinguish between graphical models in $\mathcal{S}_0(\theta, s)$ and $\mathcal{S}_1(\theta, s)$.

THEOREM 2.1 (Necessary signal strength). *Let D be a fixed integer. Suppose that*

$$\theta \leq \max_{\substack{G_0 \in \mathcal{G}_0 : d_{\max}(G_0) \leq D, \\ \mathcal{C} \text{ divider with null base } G_0}} \kappa \sqrt{\frac{M(\mathcal{C}, d_{G_0}, \log|\mathcal{C}|)}{n}} \wedge \frac{(1 - C^{-1}) \wedge e^{-\frac{1}{2}}}{\sqrt{2}(D+2)}, \tag{2.4}$$

where C is defined in (1.2). Then if $M(\mathcal{C}, d_{G_0}, \log|\mathcal{C}|) \to \infty$ as $n \to \infty$, there exists a sufficiently small constant κ in (2.4) (depending on D, C, L) such that $\liminf_{n\to\infty} \gamma(\mathcal{S}_0(\theta, s), \mathcal{S}_1(\theta, s)) = 1$.

Theorem 2.1 allows us to quantify the signal strength necessary for combinatorial inference via combinatorial constructions. The radius $\log|\mathcal{C}|$ of the packing entropy in (2.4) ensures that the pairs of distinct edges are sufficiently far apart. The constant term $\frac{(1-C^{-1})\wedge e^{-\frac{1}{2}}}{\sqrt{2}(D+2)}$ in (2.4) ensures that precision matrices with signal strength θ indeed belong to $\mathcal{M}(s)$.

In Appendix B of the Supplementary Material [Neykov, Lu and Liu (2018a)] we also provide a *deletion-edge* version of Theorem 2.1, which proceeds in the opposite direction, that is, it starts from an alternative graph G_1 and deletes edges from the divider $\mathcal{C}$ to produce graphs under the null hypothesis. This strategy can yield sharper results than Theorem 2.1 in certain situations, and we illustrate this with two examples in Appendix B.

PROOF SKETCH. The proof of Theorem 2.1 can roughly be divided into four steps. Full details of the proof will be provided in Section 6.

Step 1 (*Connect the structural parameters to geometric parameters*). Given the adjacency matrices of the null and alternative graphs G_0 and $\{G_e\}_{e\in\mathcal{C}}$, we construct the corresponding precision matrices and make sure that they belong to $\mathcal{S}_0(\theta, s)$ and $\mathcal{S}_1(\theta, s)$.

Step 2 (*Construct minimax risk lower bound via Le Cam's method*). The second step uses Le Cam's method to lower bound $\gamma(\mathcal{S}_0(\theta, s), \mathcal{S}_1(\theta, s))$. This requires us

to evaluate the chi-square divergence between a normal and a mixture normal distribution. The chi-square divergence can be expressed via ratios of determinants. In particular, we show that the log chi-square divergence can be equivalently re-expressed via an infinite sum of differences among trace operators of adjacency matrix powers.

Step 3 (*Represent the lower bound by the number of shortest closed walks in the graph*). In this step, we control the deviations of the differences of the trace operators. Since the trace of the power of an adjacency matrix equals the number of closed walks within the corresponding graph, we eliminate the trace powers which are smaller than the shortest closed walks. The traces of the higher powers are handled via matrix perturbation bounds.

Step 4 (*Characterize the smallest magnitude of the geometric parameter using the packing entropy*). Lastly, we show that condition (2.4) ensures that the closed walks on the packing of the divider are sufficiently lengthy, which implies that the chi-square divergence vanishes when the signal strength θ is small. □

A typical application of Theorem 2.1 proceeds by constructing a graph G_0 under the null hypothesis, which is one edge apart from the alternative. Next, one builds a divider $\mathcal{C}$ with as large as possible packing number, so that *adding* any edge from $\mathcal{C}$ to G_0 results in an alternative graph. Clearly, choosing the graph G_0 is crucial for this strategy to work. Below we give several examples of explicit constructions of G_0 and divider. At the end of the section, we also provide somewhat general guidance how to select G_0.

2.1. *Some applications*. In this section, we give several examples of combinatorial testing, which readily fall into the framework developed in Section 2. We show one more additional example on self-avoiding paths in Appendix B of the Supplementary Material [Neykov, Lu and Liu (2018a)].

EXAMPLE 2.1 (Connectivity testing). Consider the sub-decomposition $\mathcal{G}_0 = \{G \in \mathcal{G} \mid G \text{ disconnected}\}$ versus $\mathcal{G}_1 = \{G \in \mathcal{G} \mid G \text{ connected}\}$. We construct a base graph $G_0 := (\overline{V}, E_0)$ where

$$E_0 := \big\{(j, j+1)_{j=1}^{\lfloor d/2 \rfloor - 1}, (\lfloor d/2 \rfloor, 1), (j, j+1)_{j=\lfloor d/2 \rfloor + 1}^{d}, (\lfloor d/2 \rfloor + 1, d)\big\},$$

and let $\mathcal{C} := \{(j, \lfloor d/2 \rfloor + j)_{j=1}^{\lfloor d/2 \rfloor}\}$ (see Figure 1). Clearly, adding any edge from $\mathcal{C}$ to G_0 connects the graph, so $\mathcal{C}$ is a single edge divider with a null base G_0. Furthermore, the maximum degree of G_0 equals 2 by construction. To construct a packing set of $\mathcal{C}$, we collect all edges $(j, \lfloor d/2 \rfloor + j)$ satisfying $\lceil \log |\mathcal{C}| \rceil$ divides j except if $j > \lfloor d/2 \rfloor - \lceil \log |\mathcal{C}| \rceil$. This procedure results in a packing set with radius at least $\lceil \log |\mathcal{C}| \rceil$ which has cardinality of at least $\lfloor \frac{|\mathcal{C}|}{\lceil \log |\mathcal{C}| \rceil} \rfloor - 1$. Therefore,

$$M\big(\mathcal{C}, d_{G_0}, \log |\mathcal{C}|\big) \geq \log\left[\left\lfloor \frac{|\mathcal{C}|}{\lceil \log |\mathcal{C}| \rceil} \right\rfloor - 1\right] \asymp \log |\mathcal{C}| \asymp \log d.$$

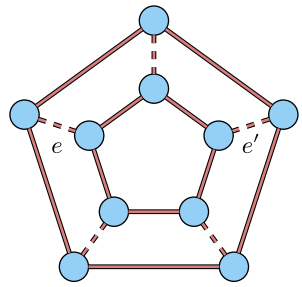


FIG. 1. *The graph G_0 with two edges $e, e' \in \mathcal{C} : d_{G_0}(e, e') = 2$, $d = 10$.*

Theorem 2.1 implies that the asymptotic minimax risk is 1 if $\theta < \kappa\sqrt{\log d/n} \wedge \frac{(1-C^{-1})\wedge e^{-\frac{1}{2}}}{4\sqrt{2}}$.

EXAMPLE 2.2 ($\mathfrak{m}+1$ versus $\mathfrak{m}$ connected components, $\mathfrak{m} \geq \sqrt{d}$). Let $\mathfrak{m} \geq \sqrt{d}$ be an integer. In this example, we are interested in testing whether the graph contains $\mathfrak{m}+1$ connected components versus $\mathfrak{m}$ connected components. The reason to assume $\mathfrak{m} \geq \sqrt{d}$ is to make sure there are sufficiently many edges for constructing a single edge divider in order to obtain sharp bounds. The case when $\mathfrak{m} < \sqrt{d}$ is treated in Example B.2 via a different divider construction. (In fact, the case $\mathfrak{m} < \sqrt{d}$ requires deleting edges from the alternative rather than adding edges to the null base. See Appendix B of the Supplementary Material [Neykov, Lu and Liu (2018a)] for more details.) Formally we have the sub-decomposition $\mathcal{G}_0 = \{G \in \mathcal{G} \mid G \text{ has } \geq \mathfrak{m}+1 \text{ connected components}\}$ versus $\mathcal{G}_1 = \{G \in \mathcal{G} \mid G \text{ has} \leq \mathfrak{m} \text{ connected components}\}$. Construct the null base graph $G_0 = (\overline{V}, E_0)$, where $E_0 := \{(j, j+1)_{j=1}^{d-\mathfrak{m}-1}\}$, and we let $\mathcal{C} := \{(j, j+1)_{j=d-\mathfrak{m}}^{d-1}\}$ (see Figure 2). Adding an edge $e \in \mathcal{C}$ to G_0 converts the base graph G_0 into a graph with $\mathfrak{m}$ connected components and, therefore, $\mathcal{C}$ is a single edge divider with a null base G_0. Additionally, the maximum degree of G_0 is 2 by construction. Note that the distance between any two edges in $\mathcal{C}$ is 0 if and only if they share a common vertex, and ∞ in all other cases. This implies that we can construct a packing set by taking every other edge in the set $\mathcal{C}$. We conclude that $M(\mathcal{C}, d_{G_0}, \log|\mathcal{C}|) \asymp \log(|\mathcal{C}|/2) \asymp \log d$. Hence, by Theorem 2.1 the minimax risk goes to 1 when $\theta < \kappa\sqrt{\log d/n} \wedge \frac{(1-C^{-1})\wedge e^{-\frac{1}{2}}}{4\sqrt{2}}$.

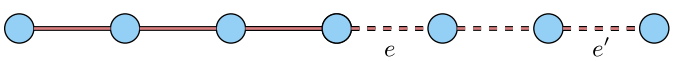


FIG. 2. *Null base graph G_0 with $d-\mathfrak{m}-1$ edges, divider $\mathcal{C}$ (dashed), $d_{G_0}(e, e') = \infty$, $d = 7$.*

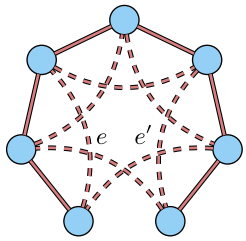


FIG. 3. *The graph G_0 with two (dashed) edges $e, e' \in \mathcal{C}$ such that $d_{G_0}(e, e') = 2$, $d = 7$.*

EXAMPLE 2.3 (Cycle testing). Consider testing whether the graph is a forest vs the graph contains a cycle. Let $\mathcal{G}_0 = \{G \in \mathcal{G} \mid G \text{ is cycle-free}\}$ and $\mathcal{G}_1 = \{G \in \mathcal{G} \mid G \text{ contains a cycle}\}$. Define the null base graph $G_0 = (\overline{V}, E_0)$, where $E_0 := \{(j, j+1)_{j=1}^{d-1}\}$. Let the edge set $\mathcal{C} := \{(j, j+2)_{j=1}^{d}\}$, where the addition is taken modulo d (refer to Figure 3 for a visualization). By construction, we have $G_0 \in \mathcal{G}_0$ and $|\mathcal{C}| = d$. Adding any edge from $\mathcal{C}$ to G_0 results in a graph with a cycle, and hence the edge set $\mathcal{C}$ is a single edge divider with a null base G_0. The maximum degree of G_0 equals 2, and is thus bounded. Moreover, there exists a $(\log|\mathcal{C}|)$-packing set of $\mathcal{C}$ of cardinality at least $\frac{|\mathcal{C}|-2}{\lceil \log|\mathcal{C}|\rceil+2}$ which can be produced by collecting the edges $(j, j+2)$ for $j = k(\log|\mathcal{C}|+2)+1$ for $k = 0, 1, \ldots$ and $j \le d-2$. The last observation implies that $M(\mathcal{C}, d_{G_0}, \log|\mathcal{C}|) \asymp \log \frac{|\mathcal{C}|-2}{\lceil \log|\mathcal{C}|\rceil+2} \asymp \log d$. Hence by Theorem 2.1 we conclude that the minimax risk goes to 1 when $\theta < \kappa\sqrt{\log d/n} \wedge \frac{(1-C^{-1})\wedge e^{-\frac{1}{2}}}{4\sqrt{2}}$.

EXAMPLE 2.4 (Tree versus connected graph with cycles). The construction in Example 2.3 also shows that we have the same signal strength limitation to test for cycles, even if we restrict to the subclass of connected graphs, that is, the class of graphs under the null hypothesis is the class of all trees $\mathcal{G}_0 = \{G \in \mathcal{G} \mid G \text{ is a tree}\}$, and the alternative is the class of all connected graphs which contain a cycle—$\mathcal{G}_1 \in \{G \in \mathcal{G} \mid G \text{ is connected but is not a tree}\}$.

EXAMPLE 2.5 (Triangle-free graph). Consider testing whether the graph contains a triangle (i.e., 3-clique). More formally, let the decomposition of $\mathcal{G}$ be $\mathcal{G}_0 = \{G \in \mathcal{G} \mid G \text{ is triangle-free}\}$ and $\mathcal{G}_1 = \{G \in \mathcal{G} \mid \exists\ 3\text{-clique in } G\}$. It is clear that in this case we can reuse the set $\mathcal{C}$ and its null base $\tilde{G}_0 = (\overline{V}, E_0 \cup \{(1, d)\})$, where E_0 and $\mathcal{C}$ are taken as in Example 2.3.

2.2. *General remarks on choosing a null base G_0.* The following result sheds some light on reasonable choices of G_0.

PROPOSITION 2.1. *Let the graph $G_0 = (\overline{V}, E_0) \in \mathcal{G}_0$ have bounded maximum degree. Suppose there exist constants $0 < c, \gamma \le 1$ so that for each vertex $v \in \overline{V}$,*

one can find a set of vertices W_v satisfying $|W_v| \geq cd^\gamma$ and for all $w \in W_v$, we have $(\overline{V}, E_0 \cup \{(v, w)\}) \in \mathcal{G}_1$. Then there exists a divider with null base G_0 satisfying $M(\mathcal{C}, d_{G_0}, \log|\mathcal{C}|) \asymp \log d$.

Of note, for any edge set $\mathcal{C}$ one has $M(\mathcal{C}, d_{G_0}, \log|\mathcal{C}|) \leq \log\binom{d}{2} \asymp \log d$, which implies that graphs $G_0 \in \mathcal{G}_0$ as in Proposition 2.1 give scalar optimal bounds. The existence of such graphs is dependent on the sub-decomposition $(\mathcal{G}_0, \mathcal{G}_1)$. Notably, all examples in Section 2.1 fall under the framework of Proposition 2.1. Its proof can be found in the Supplementary Material [Neykov, Lu and Liu (2018a)].

REMARK 2.1. When $\mathbf{\Theta} \in \mathcal{M}(s)$, the results in Theorem 2.1 (and Theorem B.1) suggest that a signal strength of order $\sqrt{\log d/n}$ is necessary for controlling the minimax risk (1.3). In fact, Theorem 7 of Cai, Liu and Luo (2011) shows that under such signal strength condition, support recovery of $\mathbf{\Theta}$ is indeed achievable, which further implies that controlling the minimax risk (1.3) is possible. A naive procedure for matching the lower bound is to first perfectly recover the graph structure. Then construct a test based on examining whether the graph has the desired combinatorial structure. Though such an approach is theoretically feasible, it is not practical. First, such an approach is overly conservative and does not allow us to tightly control the type I error at a desired level. Second, such an approach crucially depends on having a suitable thresholding parameter to estimate the graph, which is in general not realistic.

In the next section, we present a family of testing procedures, which do not require perfect support recovery of the full graph. Compared to the naive approach described in Remark 2.1, our tests explicitly exploit the combinatorial structure of the targeted hypotheses and can control the type I error at any desired level.

3. Alternative witness test. We start with a high-level outline of a new combinatorial inference approach for graphical models. For clarity, we mainly present using the case of Gaussian graphical models, and comment on extensions to other graphical models in Appendix D of the Supplementary Material [Neykov, Lu and Liu (2018a)].

Importantly, the algorithms we develop in this section apply to alternative graph classes which are stable under edge addition. Formally, for any $G = (\overline{V}, E) \in \mathcal{G}_1$ and any edge e, we require that the graph $(\overline{V}, E \cup \{e\}) \in \mathcal{G}_1$. Note that due to edge addition stability under the alternative, the full graph $\overline{G} = (\overline{V}, \overline{E})$ belongs to $\mathcal{G}_1$. For a graph $G \in \mathcal{G}_1$, define the following class of edge sets:

$$\mathcal{W}_1(G) = \big\{E \mid \forall E, E \subseteq E(G) \Rightarrow (\overline{V}, E) \in \mathcal{G}_1\big\}.$$

The set $\mathcal{W}_1(G)$ collects all edge sets forming graphs in $\mathcal{G}_1$, which can be obtained by the graph G via iteratively pruning one edge at a time. We use the shorthand notation

$$\overline{\mathcal{W}}_1 = \mathcal{W}_1(\overline{G}).$$

Consider the following parameter sets:

$$\mathcal{S}_0(s) := \big\{\boldsymbol{\Theta} \in \mathcal{M}(s) \mid G(\boldsymbol{\Theta}) \in \mathcal{G}_0\big\} \quad \text{and} \tag{3.1}$$

$$\mathcal{S}_1(\theta, s) := \Big\{\boldsymbol{\Theta} \in \mathcal{M}(s) \;\Big|\; G(\boldsymbol{\Theta}) \in \mathcal{G}_1,\ \max_{E \in \overline{\mathcal{W}}_1} \min_{e \in E} |\Theta_e| \geq \theta\Big\}. \tag{3.2}$$

The parameter set (3.1) does not impose any assumption on the minimum signal strength, thus is broader than the one defined in (2.2). In Definition (3.2), the signal strength is not imposed on all edges of the alternative graphs. In fact, we only need to impose the signal strength assumption on a subset of edges which can be obtained by pruning the complete graph. Such a condition is much weaker than the usual condition needed for perfect graph recovery. We note that for sub-decompositions $(\mathcal{G}_0, \mathcal{G}_1)$ satisfying $\mathcal{W}_1(G) \subseteq \overline{\mathcal{W}}_1$ for all $G \in \mathcal{G}_1$, the parameter set (3.2) is strictly larger than the parameter set (2.3).

Given n independent samples $\boldsymbol{X}_i \sim N(0, (\boldsymbol{\Theta}^*)^{-1})$ and a sub-decomposition $(\mathcal{G}_0, \mathcal{G}_1)$, we formulate a procedure for testing $\mathbf{H}_0 : \boldsymbol{\Theta}^* \in \mathcal{S}_0(s)$ vs $\mathbf{H}_1 : \boldsymbol{\Theta}^* \in \mathcal{S}_1(\theta, s)$. Let $\hat{\boldsymbol{\Sigma}} := n^{-1} \sum_{i=1}^n \boldsymbol{X}_i^{\otimes 2}$ be the empirical covariance matrix. Let $\hat{\boldsymbol{\Theta}}$ be any estimator of the precision matrix $\boldsymbol{\Theta}^*$ satisfying for some fixed constant $K > 0$:

$$\big\|\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*\big\|_{\max} \leq K\sqrt{\log d/n}, \tag{3.3}$$

$$\big\|\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*\big\|_1 \leq Ks\sqrt{\log d/n}, \qquad \|\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\Theta}} - \mathbf{I}_d\|_{\max} \leq K\sqrt{\log d/n}, \tag{3.4}$$

with probability at least $1 - d^{-1}$ uniformly over the parameter space $\mathcal{M}(s)$ [recall definition (1.2)]. An example of an estimator of $\boldsymbol{\Theta}^*$ with this properties is the CLIME procedure introduced by Cai, Liu and Luo (2011) [see also (A.2)]. An overview of the *alternative witness test* is sketched below:

i. In the first step, the alternative witness test identifies a *minimal structure witnessing the alternative*.

ii. In the second step, the alternative witness test attempts to *certify* that the minimal structure identified by the first step is indeed present in the graph.

Split the data $\mathcal{D} = \{\boldsymbol{X}_i\}_{i=1}^n$ in two approximately equal-sized sets $\mathcal{D}_1 = \{\boldsymbol{X}_i\}_{i=1}^{\lfloor n/2 \rfloor}$, $\mathcal{D}_2 = \{\boldsymbol{X}_i\}_{i=\lfloor n/2 \rfloor + 1}^n$ and obtain estimates $\hat{\boldsymbol{\Theta}}^{(1)}$, $\hat{\boldsymbol{\Theta}}^{(2)}$ on $\mathcal{D}_1$ and $\mathcal{D}_2$ correspondingly. For the first step, we exploit $\hat{\boldsymbol{\Theta}}^{(1)}$ to solve the following max-min combinatorial optimization problem:

$$\widehat{E} := \operatorname*{argmax}_{E \in \overline{\mathcal{W}}_1} \min_{e \in E} |\hat{\Theta}_e^{(1)}|, \tag{3.5}$$

where edge sets with the smallest cardinality are preferred, and further ties are broken arbitrarily. Program (3.5) aims at identifying the smallest edge set in $\overline{\mathcal{W}}_1$ whose minimal signal is as large as possible. Given a consistent estimator $\hat{\boldsymbol{\Theta}}^{(1)}$ and sufficiently strong signal strength, the solution of (3.5) identifies a minimal

substructure of $G(\mathbf{\Theta}^*)$ belonging to $\mathcal{G}_1$. This strategy is motivated by the definition of the alternative parameter set $\mathcal{S}_1(\theta, s)$ (3.2). We remark that solving program (3.5) could be computationally challenging for some combinatorial tests. However, for all examples considered in this paper, simple and efficient polynomial time algorithms are available. We refer to the graph $(\overline{V}, \widehat{E})$ as the *minimal structure witnessing the alternative*. Although the minimal structure witness is defined in full generality, for the ease of presentation we justify its validity on a case-by-case basis.

In the second step, the alternative witness test attempts to certify the witness structure using the estimate $\hat{\mathbf{\Theta}}^{(2)}$. Formally, we aim to test the hypothesis

$$\mathbf{H}_0 : \exists e \in \widehat{E} : \Theta_e^* = 0 \quad \text{vs} \quad \mathbf{H}_1 : \forall e \in \widehat{E} : \Theta_e^* \neq 0. \tag{3.6}$$

A rejection of the null hypothesis in (3.6) certifies the presence of the alternative witness structure. If the test fails to reject, the alternative witness test cannot reject the null structure hypothesis. In Section 3.1, we give a detailed description on how the second step of the test works.

3.1. *Minimal structure certification.* In this section, we detail an algorithm for testing (3.6). In fact, we present a general test for the following multiple testing problem:

$$\mathbf{H}_0 : \exists e \in E \text{ s.t. } \Theta_e^* = 0 \quad \text{vs} \quad \mathbf{H}_1 : \forall e \in E, \Theta_e^* \neq 0,$$

using the data $\mathcal{X} = \{\boldsymbol{X}_i\}_{i=1}^n$, $\boldsymbol{X}_i \sim N(0, (\mathbf{\Theta}^*)^{-1})$, where E is a pre-given edge set. Following Neykov et al. (2018b), for any $j, k \in [d]$ we define the bias corrected estimate

$$\tilde{\Theta}_{jk} := \hat{\Theta}_{jk} - \frac{\hat{\mathbf{\Theta}}_{*j}^T(\hat{\mathbf{\Sigma}}\hat{\mathbf{\Theta}}_{*k} - \mathbf{e}_k)}{\hat{\mathbf{\Theta}}_{*j}^T\hat{\mathbf{\Sigma}}_{*j}}, \tag{3.7}$$

where $\mathbf{e}_k$ is a canonical unit vector with 1 at its kth entry. Under mild regularity conditions, we show that if $\hat{\mathbf{\Theta}}$ satisfies (3.3) and (3.4), $\tilde{\Theta}_{jk}$ admits the following Bahadur representation:

$$\sqrt{n}\tilde{\Theta}_{jk} = n^{-1/2}\sum_{i=1}^n(\mathbf{\Theta}_{*j}^{*T}\boldsymbol{X}_i^{\otimes 2}\mathbf{\Theta}_{*k}^* - \Theta_{jk}^*) + o_p(1). \tag{3.8}$$

This motivates a mutiplier bootstrap scheme for approximating the distribution of the statistic $\sqrt{n}\tilde{\Theta}_{jk}$ under the null hypothesis:

$$\hat{W}_{jk} = n^{-1/2}\sum_{i=1}^n(\hat{\mathbf{\Theta}}_{*j}^T\boldsymbol{X}_i^{\otimes 2}\hat{\mathbf{\Theta}}_{*k} - \hat{\Theta}_{jk})\zeta_i,$$

where $\zeta_i \sim N(0, 1)$, $i \in [n]$ are independent and identically distributed. To approximate the null distribution of the statistic $\max_{(j,k)\in S}\sqrt{n}|\tilde{\Theta}_{jk}|$ over a subset $S \subseteq E$,

let $c_{1-\alpha,S}$ denote the $(1-\alpha)$-quantile of the statistic $\max_{(j,k)\in S}|\hat{W}_{jk}|$ (conditioning on the dataset $\mathcal{X}$). Formally, we let

$$c_{1-\alpha,S} := \inf_{t\in\mathbb{R}}\Big\{t \mid \mathbb{P}\Big(\max_{(j,k)\in S}|\hat{W}_{jk}| \le t \mid \mathcal{X}\Big) \ge 1-\alpha\Big\}. \tag{3.9}$$

As defined $c_{1-\alpha,S}$ is a population quantity (conditioning on $\mathcal{X}$). In practice, an arbitrarily accurate estimate $\hat{c}_{1-\alpha,S}$ of $c_{1-\alpha,S}$ can be obtained via Monte Carlo simulations. Below we describe a multiple edge testing procedure returning a subset $\hat{E^{\mathrm{nc}}} \subseteq E$ of rejected edges (hypotheses). Our procedure is based on the step-down construction of Chernozhukov, Chetverikov and Kato (2013), which is inspired by the multiple testing method of Romano and Wolf (2005). Decompose the edge set $E = E^{\mathrm{n}} \cup E^{\mathrm{nc}}$, where $E^{\mathrm{n}} \cap E^{\mathrm{nc}} = \varnothing$, E^{n} is the subset of true null edges and E^{nc} is the set of nonnull edges. Define the parameter set

$$\mathcal{M}_{E^{\mathrm{n}},E^{\mathrm{nc}}}(s,\kappa) = \Big\{\boldsymbol{\Theta}\in\mathcal{M}(s) \;\Big|\; \max_{e\in E^{\mathrm{n}}}|\Theta_e| = 0,\ \min_{e\in E^{\mathrm{nc}}}|\Theta_e| \ge \kappa\sqrt{\frac{\log d}{n}}\Big\}. \tag{3.10}$$

For a fixed edge set E, we say that an edge set $\widehat{E^{\mathrm{r}}}$ has strong control of the familywise error rate if

$$\limsup_{n\to\infty}\ \sup_{E^{\mathrm{n}}\subset E}\ \sup_{\boldsymbol{\Theta}^*\in\mathcal{M}_{E^{\mathrm{n}},E^{\mathrm{nc}}}(s,\kappa)} \mathbb{P}\big(E^{\mathrm{n}}\cap\widehat{E^{\mathrm{r}}}\neq\varnothing\big) \le \alpha \tag{3.11}$$

for some pre-specified size $\alpha > 0$. Our next result shows that Algorithm 1 returns an edge set $\widehat{E^{\mathrm{nc}}}$ with strong control of the familywise error rate.

PROPOSITION 3.1 (Strong familywise error rate test). *Let*

$$\boldsymbol{\Theta}^*\in\mathcal{M}_{E^{\mathrm{n}},E^{\mathrm{nc}}}(s,\kappa),$$

and furthermore

$$s\log(nd)\sqrt{\log d\log(nd)}/\sqrt{n} = o(1), \qquad \big(\log(dn)\big)^6/n = o(1). \tag{3.12}$$

Then for any fixed edge set E, the output $\widehat{E^{\mathrm{nc}}}$ of Algorithm 1 satisfies (3.11). In addition, if the constant κ in (3.10) satisfies $\kappa \ge C'C^4$ for a fixed absolute constant $C' > 0$, we have $\liminf_{n\to\infty}\inf_{E^{\mathrm{n}}\subseteq E}\inf_{\boldsymbol{\Theta}^*\in\mathcal{M}_{E^{\mathrm{n}},E^{\mathrm{nc}}}(s,\kappa)}\mathbb{P}(E^{\mathrm{nc}} = \widehat{E^{\mathrm{nc}}}) = 1$.

Algorithm 1 Multiple edge testing

Initialize $\widehat{E^{\mathrm{n}}} \leftarrow E$.
repeat
  Reject $R \leftarrow \{e\in\widehat{E^{\mathrm{n}}} : \sqrt{n}|\tilde{\Theta}_e| \ge c_{1-\alpha,\widehat{E^{\mathrm{n}}}}\}$.
  Update $\widehat{E^{\mathrm{n}}} \leftarrow \widehat{E^{\mathrm{n}}}\setminus R$.
until $R = \varnothing$ or $\widehat{E^{\mathrm{n}}} = \varnothing$
return $\hat{E^{\mathrm{nc}}} \leftarrow E\setminus\widehat{E^{\mathrm{n}}}$

The first condition in (3.12) ensures the validity of the Bahadur representation in (3.8). The second condition in (3.12) is to guarantee validity of the high-dimensional bootstrap, and a similar condition is required by Chernozhukov, Chetverikov and Kato (2013). While the first condition of (3.12) is not necessarily sharp, it is nearly optimal by ignoring logarithmic terms of dimension and sample size comparing to the minimax rate established in Ren et al. (2015).

Of note, when κ is sufficiently large, Algorithm 1 achieves exact control of the familywise error rate, that is, we have equality in (3.11). This happens since all edges in E^{nc} will be rejected with overwhelming probability, while the bootstrap comparison is asymptotically exact for the remaining edges E^{n}. As a consequence of this result, if the null hypothesis set $\mathcal{S}_0(s)$ considered in (3.1) exhibits signal strength as in definition (2.2), the alternative witness tests are exact.

DEFINITION 3.1. For an edge set E, let $\hat{E^{\mathrm{nc}}}$ be the output of Algorithm 1 on data $\mathcal{X}$ with level $1-\alpha$. Define the following test function:

$$\psi^{B}_{\alpha,E}(\mathcal{X}) := \mathbb{1}\big(E = \widehat{E^{\mathrm{nc}}}\big).$$

$\psi^{B}_{\alpha,E}(\mathcal{X})$ tests whether the set E is comprised only of nonnull edges.

3.2. *Examples*. In this section, we describe practical algorithms based on the alternative witness test for testing problems outlined in Section 2.1. Our tests can distinguish the null from the alternative hypotheses when the minimum signal strength is sufficiently large. As we shall see, the magnitude of the required signal strength is precisely of order $\sqrt{\log d/n}$ and, therefore, in view of Section 2 these tests are minimax optimal. Recall that we observe n i.i.d. samples $\{\boldsymbol{X}_i\}_{i=1}^n$ from $\boldsymbol{X}_i \sim N_d(0, (\boldsymbol{\Theta}^*)^{-1})$. We split the data into $\mathcal{D}_1 = \{\boldsymbol{X}_i\}_{i=1}^{\lfloor n/2 \rfloor}$, $\mathcal{D}_2 = \{\boldsymbol{X}_i\}_{i=\lfloor n/2 \rfloor+1}^{n}$ and obtain estimates $\hat{\boldsymbol{\Theta}}^{(1)}$, $\hat{\boldsymbol{\Theta}}^{(2)}$ on $\mathcal{D}_1$ and $\mathcal{D}_2$ correspondingly, for which (3.3) and (3.4) hold. For space considerations, we present only the connectivity test in full details, and we elaborate on the minimal structures for the remaining tests. Full details can be found in Appendix C of the Supplementary Material [Neykov, Lu and Liu (2018a)].

3.2.1. *Connectivity testing*. This example proposes a new procedure for honestly testing whether $G(\boldsymbol{\Theta}^*)$ is a connected graph. Accordingly, the subdecomposition is $\mathcal{G}_0 := \{G \in \mathcal{G} \mid G \text{ disconnected}\}$ and $\mathcal{G}_1 := \{G \in \mathcal{G} \mid G \text{ connected}\}$. The pair $(\mathcal{G}_0, \mathcal{G}_1)$ determines the parameter sets definitions (3.1) and (3.2).

Finding the minimal structure witness (3.5) reduces to finding a maximum spanning tree (MST) $\hat{T}$ on the full graph with edge weights $|\hat{\Theta}_e^{(1)}|$. The complexity of finding a MST is $O(d^2 \log d)$, where d is the number of vertices. We summarize the procedure in Algorithm 2.

The results on connectivity testing are summarized in the following.

Algorithm 2 Connectivity test

Input: $\mathcal{D} = \{\boldsymbol{X}_i\}_{i=1}^n$, level $0 < \alpha < 1$.
Split the data $\mathcal{D}_1 = \{\boldsymbol{X}_i\}_{i=1}^{\lfloor n/2 \rfloor}$, $\mathcal{D}_2 = \{\boldsymbol{X}_i\}_{i=\lfloor n/2 \rfloor+1}^n$
Using $\mathcal{D}_1$ obtain estimate $\hat{\boldsymbol{\Theta}}^{(1)}$ satisfying (3.3)
Find MST $\hat{T}$ on the full graph with weights $|\hat{\Theta}_e^{(1)}|$.
Output: $\psi_{\alpha,\hat{T}}^B(\mathcal{D}_2)$.

COROLLARY 3.1. *Let* $\theta = (2K + C'C^4)\sqrt{\log d/\lfloor n/2 \rfloor}$ *for an absolute constant* $C' > 0$ *and assume that* (3.12) *holds. Then for any fixed level* α, *the test* $\psi_{\alpha,\hat{T}}^B(\mathcal{D}_2)$ *from the output of Algorithm* 2 *satisfies*

$$\limsup_{n\to\infty} \sup_{\boldsymbol{\Theta}^* \in \mathcal{S}_0(s)} \mathbb{P}(\textit{reject } \mathbf{H}_0) \le \alpha, \qquad \liminf_{n\to\infty} \inf_{\boldsymbol{\Theta}^* \in \mathcal{S}_1(\theta,s)} \mathbb{P}(\textit{reject } \mathbf{H}_0) = 1.$$

3.2.2. *Connected components testing.* Connected component testing is more general compared to connectivity testing. For $\mathfrak{m} \in [d-1]$, let $\mathbf{H}_0$: # connected components $\ge \mathfrak{m}+1$ vs $\mathbf{H}_1$: # connected components $\le \mathfrak{m}$. Testing connectivity is a special case when $\mathfrak{m} = 1$.

For a fixed $\mathfrak{m} \in [d-1]$, define the sub-decomposition $\mathcal{G}_0 = \mathcal{F}_{\mathfrak{m}+1}$ and $\mathcal{G}_1 = \bigcup_{j \le \mathfrak{m}} \mathcal{F}_j$ where

$$\mathcal{F}_j := \{G \in \mathcal{G} \mid G \text{ has exactly } j \text{ connected components}\} \qquad \text{for all } j \in [d].$$

The sub-decomposition $(\mathcal{G}_0, \mathcal{G}_1)$ also defines the parameter sets (3.1) and (3.2). Recall that a sub-graph F of G, that is, $E(F) \subset E(G)$ and $V(F) = V(G) = \overline{V}$, is called a *spanning forest* of G, if F contains no cycles, adding any edge $e \in E(G) \setminus E(F)$ to $E(F)$ creates a cycle, and $|E(F)|$ is maximal. This definition extends naturally to graphs with positive weights on their edges. It is easy to check that the minimal structure witness (3.5) is the maximal spanning forrest with $\mathfrak{m}$ connected components, and can be found efficiently via a greedy algorithm. For more details, see Appendix C of the Supplementary Material [Neykov, Lu and Liu (2018a)].

3.2.3. *Cycle testing.* In this example, we sketch how to test whether the graph is a forest. Recall the sub-decomposition $\mathcal{G}_0 = \{G \in \mathcal{G} \mid G \text{ is a forest}\}$ and $\mathcal{G}_1 = \{G \in \mathcal{G} \mid G \text{ contains a cycle}\}$. The pair $(\mathcal{G}_0, \mathcal{G}_1)$ also defines the parameter sets (3.1) and (3.2). The minimal structure witness (3.5) is a cycle, and can be found via greedily adding edges until a cycle is formed. For more details, see Appendix C of the Supplementary Material [Neykov, Lu and Liu (2018a)].

3.2.4. *Triangle-free graph testing.* In this example, we discuss an algorithm for testing whether the graph is triangle-free. The corresponding sub-decomposition is $\mathcal{G}_0 = \{G \in \mathcal{G} \mid G \text{ is triangle-free}\}$ and $\mathcal{G}_1 = \{G \in \mathcal{G} \mid \exists\ 3\text{-clique}$

subgraph of $G\}$. The pair $(\mathcal{G}_0, \mathcal{G}_1)$ also defines the parameter sets (3.1) and (3.2). The minimal structure witness is a triangle, and can be greedily identified.

4. Multi-edge dividers. The results of Section 2 (and Appendix B of the Supplementary Material) have two major limitations. First, the null base G_0 is assumed to be of bounded degree. Second, our results cover only tests for which there exists a singe-edge divider. In this section, we relax both of these conditions. The following motivating example illustrates a relevant testing problem which does not fall into the framework of Section 2.

Maximum degree testing. Consider testing whether the maximum degree of the graph $d_{\max}$ satisfies $d_{\max} \leq s_0$ vs $d_{\max} \geq s_1$, where $s_0 < s_1 \leq s$ are integers which are allowed to scale with n. In this case, it is impossible to simultaneously construct a null base graph G_0 of bounded degree and a single-edge divider $\mathcal{C}$.

To handle multiple edge dividers, we first extend Definitions 2.1 and 2.2 to allow for the above examples.

DEFINITION 4.1 (Null-alternative divider). Let $G_0 = (\overline{V}, E_0) \in \mathcal{G}_0$ be a fixed graph under the null with adjacency matrix $\mathbf{A}_0$. We call a collection of edge sets $\boldsymbol{C}$ a (multi-edge) divider with null base G_0, if for all edge sets $S \in \boldsymbol{C}$ we have $S \cap E_0 = \varnothing$ and $(\overline{V}, E_0 \cup S) \in \mathcal{G}_1$. For any edge set $S \in \boldsymbol{C}$, we denote the adjacency matrix of the graph $(\overline{V}, S)$ with $\mathbf{A}_S$.

DEFINITION 4.2 (Edge set geodesic predistance). For two edge sets S and S' and a given graph G, let $d_G(S, S') = \min_{e \in S, e' \in S'} d_G(e, e')$.

We provide two generic strategies for obtaining combinatorial inference lower bounds on the signal strength. The first strategy, described in Section 4.1, assumes that all $S \in \boldsymbol{C}$ satisfy $|S| \leq U$ for some fixed constant U. The second strategy, presented in Section 4.2, does not require bounded cardinality of the edge sets S, but requires that the null bases and dividers have some special combinatorial properties.

4.1. *Bounded edge sets.* Below we consider an extension of Theorem 2.1 for multi-edge dividers, where the number of edges in each set $S \in \boldsymbol{C}$ satisfy $|S| \leq U$ for some fixed integer $U \in \mathbb{N}$. In contrast to Section 2, here the graph G_0 is allowed to have unbounded degree.

THEOREM 4.1. *Let $G_0 \in \mathcal{G}_0$ be a graph under the null, and let $\boldsymbol{C}$ be a multi-edge divider with null base G_0. Suppose that for some sufficiently small absolute constant $\kappa > 0$*

$$\theta \leq \kappa \sqrt{\frac{M(\boldsymbol{C}, d_{G_0}, \log|\boldsymbol{C}|)}{nU}} \wedge \frac{\kappa}{U(\|\mathbf{A}_0\|_2 + 2U)} \wedge \frac{1 - C^{-1}}{4(\|\mathbf{A}_0\|_1 + 2U)}. \tag{4.1}$$

If $M(\boldsymbol{C}, d_{G_0}, \log|\boldsymbol{C}|) \to \infty$, we have $\liminf_{n\to\infty} \gamma(\mathcal{S}_0(\theta, s), \mathcal{S}_1(\theta, s)) = 1$.

Theorem 4.1 is an extension of Theorem 2.1. Specifically, Theorem 2.1 corresponds the setting where $U = 1$, and $\|\mathbf{A}_0\|_2 \le \|\mathbf{A}_0\|_1 \le D$ (recall that D is an upper bound of the graph degree). Even though by assumption $U = \max_{S\in\boldsymbol{C}} |S|$ is bounded, we explicitly keep the dependency on U in (4.1) to reflect how the bound changes if U is allowed to scale. The first term on the right-hand side of (4.1) is the structural packing entropy, while the remaining two terms ensure the parameter θ is small enough to construct a valid packing set (more details are provided in the proof).

We illustrate the usefulness of Theorem 4.1 by an example similar to the ones in Section 2.1. Consider testing whether the maximum degree of the graph $G(\boldsymbol{\Theta}^*)$ is at most s_0 versus it is at least s_1, where $s_0 < s_1 \le s$ can increase with n but the null-alternative gap $s_1 - s_0$ remains bounded. Therefore, we cannot apply Theorem 2.1 but should use Theorem 4.1 instead. Define the sub-decomposition $\mathcal{G}_0 = \{G \mid d_{\max}(G) \le s_0\}$ and $\mathcal{G}_1 = \{G \mid d_{\max}(G) \ge s_1\}$, respectively.

EXAMPLE 4.1 (Maximum degree test with bounded null-alternative gap). Let $\mathcal{S}_0(\theta, s)$ and $\mathcal{S}_1(\theta, s)$ be defined in (2.2) and (2.3). Assume that $s \log d/n = o(1)$, $s\sqrt{\log d/n} = O(1)$ and $s = O(d^\gamma)$ for some $\gamma < 1$. Then if κ is small enough and $\theta < \kappa\sqrt{\log d/n}$, we have $\liminf_{n\to\infty} \gamma(\mathcal{S}_0(\theta, s), \mathcal{S}_1(\theta, s)) = 1$.

The proof of Example 4.1 is deferred to the Supplementary Material [Neykov, Lu and Liu (2018a)].

4.2. *Scaling edge sets.* Theorem 4.1 requires the cardinalities of the edge sets in the divider $\boldsymbol{C}$ to be bounded. In this section, we consider multi-edge dividers $\boldsymbol{C}$ allowing the sizes of $S \in \boldsymbol{C}$ to increase with n. For this case, the previous notion of packing entropy based on geodesic predistance is no longer effective. Instead, we introduce a new mechanism called buffer entropy to quantify the lower bound under scaling multi-edge dividers.

We first intuitively explain why the structural entropy in Theorem 4.1 may not be sufficient for handling dividers with scaling edge sets. Recall that Theorem 4.1 uses the structural entropy $M(\boldsymbol{C}, d_{G_0}, \log|\boldsymbol{C}|)$ to characterize the lower bound. In turn, the structural entropy is calculated based on the edge set geodesic predistance d_{G_0} in Definition 4.2. One difference between fixed and scaling edge sets sizes is that one can only pack a limited number of edge sets or large size which are sufficiently far apart (and hence do not overlap). A less wasteful strategy would be to allow for the edge sets to overlap. However, in general, different edge sets $S, S' \in \boldsymbol{C}$ may have multiple overlapping vertices and the notion of geodesic predistance is no longer precise enough to reflect the closeness between S and S'.

Below we introduce a concept called vertex buffer, which helps to measure the closeness between edge sets S and S' more precisely than the geodesic predistance.

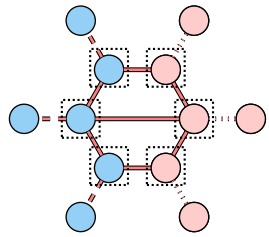


FIG. 4. *Visualization of the vertex buffer in $\mathcal{V}_{S,S'}$. Here, S, S' are plotted with dashed and dotted edges, respectively, and G_0 is in solid edges. The vertices in the buffer are marked in the dashed squares.*

DEFINITION 4.3 (Vertex buffer). Let $G_0 = (\overline{V}, E_0)$ be a given graph and S, S' be two edge sets. The vertex buffer of S, S' under G_0 is defined as

$$\mathcal{V}_{S,S'} := \{V(E_0 \cup S) \cap V(S')\} \cup \{V(E_0 \cup S') \cap V(S)\}.^{3}$$

An important property of the set $\mathcal{V}_{S,S'}$ is that all paths passing through at least one edge in both S and S' must contain at least one vertex in $\mathcal{V}_{S,S'}$. In that sense, a large buffer size $|\mathcal{V}_{S,S'}|$ indicates that the edge sets S and S' are close to each other. We visualize an example of a vertex buffer in Figure 4.

In contrast to the bounded edge sets case, when the edge sets in $\boldsymbol{C}$ are allowed to scale in size, it is not effective to build packing sets based on the predistance, since this strategy limits the number of edge sets we can build. One way to increase the cardinality of $\boldsymbol{C}$ is to consider a larger number of potentially overlapping structures, and use the buffer size as a more precise closeness measure between these structures. Below we formalize the concept of buffer entropy which quantifies this intuition.

DEFINITION 4.4 (Buffer entropy). Let $\boldsymbol{C}$ be a multi-edge divider with a base graph G_0. The buffer entropy is defined as

$$M_{\mathrm{B}}(\boldsymbol{C}, G_0) := \log\Big(\Big[\max_{S \in \boldsymbol{C}} \mathbb{E}_{S'} |\mathcal{V}_{S,S'}|\Big]^{-1}\Big), \tag{4.2}$$

where the expectation $\mathbb{E}_{S'}$ is taken from uniformly sampling S' from $\boldsymbol{C}$.

We want the buffer entropy to be as large as possible to achieve sharp lower bounds. Note the following trivial bound on the size $|\mathcal{V}_{S,S'}|$:

$$|\mathcal{V}_{S,S'}| \le \sum_{v \in V(S)} \mathbb{1}(v \in \mathcal{V}_{S,S'}) + \sum_{v \in V(S')} \mathbb{1}(v \in \mathcal{V}_{S,S'}).$$

[3]We suppress the dependence of $\mathcal{V}_{S,S'}$ on G_0 to ease the notation.

An important condition allowing us to relate the signal strength lower bounds to buffer entropy requires that the divider is such that the variables $\{\mathbb{1}(v \in \mathcal{V}_{S,S'})\}_{v\in V(S)}$ are negatively associated.

DEFINITION 4.5 (Incoherent divider). The collection of edge sets $\boldsymbol{C}$ is called an incoherent divider with a null base G_0, if for any fixed $S \in \boldsymbol{C}$, the random variables $\{\mathbb{1}(v \in \mathcal{V}_{S,S'})\}_{v\in V(S)}$ with respect to a uniformly sampled S' from $\boldsymbol{C}$ are negatively associated. In other words, for any pair of disjoint sets $I, J \subseteq V(S)$ and any pair of coordinatewise nondecreasing functions f, g we have

$$\operatorname{Cov}\big(f\big(\{\mathbb{1}(v \in \mathcal{V}_{S,S'})\}_{v\in I}\big), g\big(\{\mathbb{1}(v \in \mathcal{V}_{S,S'})\}_{v\in J}\big)\big) \le 0.$$

We show concrete constructions of incoherent dividers in Examples 4.2, 4.3 and 4.4. As a remark, negative association is satisfied by a variety of classical discrete distributions such as the multinomial and hypergeometric, and even more generally by the class of permutation distributions [e.g., Joag-Dev and Proschan (1983), Dubhashi and Ranjan (1998)]. It is a standard assumption that has been exploited in other works [e.g., Addario-Berry et al. (2010)] for obtaining lower bounds.

Besides the packing entropy, the lower bound in Theorem 4.1 involves the maximum degree $\|\mathbf{A}_0\|_1$ and the spectral norm $\|\mathbf{A}_0\|_2$. We define similar quantities for the scaling edge sets case. For a divider $\boldsymbol{C}$ with null base G_0 and any two edge sets $S, S' \in \boldsymbol{C}$ define the notation

$$\mathbf{A}_{S,S'} := \mathbf{A}_0 + \mathbf{A}_S + \mathbf{A}_{S'}. \tag{4.3}$$

As the sizes of $S, S' \in \boldsymbol{C}$ are no longer ignorable, we need to consider the matrix $\mathbf{A}_{S,S'}$ (4.3) instead. Denote the uniform maximum degree as $\Gamma := \max_{S,S'\in\boldsymbol{C}} \|\mathbf{A}_{S,S'}\|_1$ and uniform spectral norm as $\Lambda := \max_{S,S'\in\boldsymbol{C}} \|\mathbf{A}_{S,S'}\|_2$. We define

$$\mathcal{R} := \max_{S,S'\in\boldsymbol{C}} \frac{|S \cap S'|}{|\mathcal{V}_{S,S'}|}, \qquad \mathcal{B} := \Lambda^4 \wedge \max_{S,S'\in\boldsymbol{C}} (\Gamma^2 |\mathcal{V}_{S,S'}|).$$

$\mathcal{R}$ is an edge-node ratio measuring how dense the edge set $S \cap S'$ is compared to the vertex buffers. The quantity $\mathcal{B}$ is an auxiliary quantity which assembles maximum degrees, spectral norms and buffer sizes and helps to obtain a compact lower bound formulation.

Below we connect the structural features we defined above to the lower bound. Recall Definitions (2.2) and (2.3) on $\mathcal{S}_0(\theta, s)$ and $\mathcal{S}_1(\theta, s)$. We have the following theorem.

THEOREM 4.2. *Let $\boldsymbol{C}$ be an incoherent divider with a null base G_0. Then if $M_B(\boldsymbol{C}, G_0) \to \infty$ and*

$$\theta \le \sqrt{\frac{M_B(\boldsymbol{C}, G_0)}{4n\mathcal{R}}} \wedge \sqrt{\frac{\mathcal{R}}{\mathcal{B}}} \wedge \frac{1 - C^{-1}}{2\sqrt{2}\Gamma}, \tag{4.4}$$

the minimax risk satisfies $\liminf_{n\to\infty} \gamma(\mathcal{S}_0(\theta, s), \mathcal{S}_1(\theta, s)) = 1$.

When the sample size n is sufficiently large, the buffer entropy term on the right-hand side of (4.4) is the smallest term and drives the bound which bares similarity to Theorem 4.1.

To better illustrate the usage of Theorem 4.2, we consider three examples. First, we focus on the problem of testing whether the maximum degree in the graph is $\leq s_0$ versus $\geq s_1$. When $s_0 = 0$, this problem is related to the problem of detecting a set of s_1 signals in the normal means model [e.g., Ingster (1982), Baraud (2002), Donoho and Jin (2004), Addario-Berry et al. (2010), Verzelen and Villers (2010), Arias-Castro, Candès and Plan (2011)]. However the two problems are distinct, since we are studying structural testing in the graphical model setting. Given $s_0 < s_1 \leq s$, we let the sub-decomposition be $\mathcal{G}_0 = \{G \mid d_{\max}(G) \leq s_0\}$ and $\mathcal{G}_1 = \{G \mid d_{\max}(G) \geq s_1\}$. We summarize our results in the following

EXAMPLE 4.2 (Maximum degree test with scaling divider). Let $\mathcal{S}_0(\theta, s)$ and $\mathcal{S}_1(\theta, s)$ be defined in (2.2) and (2.3). Assume that $s\sqrt{\log d/n} = O(1)$ and $s = O(d^\gamma)$ for some $\gamma < 1/2$. Then for a small enough absolute constant κ if $\theta < \kappa\sqrt{\log d/n}$, we have $\liminf_{n\to\infty} \gamma(\mathcal{S}_0(\theta, s), \mathcal{S}_1(\theta, s)) = 1$.

Due to space limitations, we show how this example follows from Theorem 4.2 in Appendix E of the Supplementary Material [Neykov, Lu and Liu (2018a)]. Here, we simply sketch the construction of the divider in Figure 5. On an important note, the negative association of the random variables $\{\mathbb{1}(v \in \mathcal{V}_{S,S'})\}_{v\in V(S)}$ can be easily deduced by a result of Joag-Dev and Proschan (1983). Our second example further illustrates the usage of Theorem 4.2 with a clique detection problem. Define the null and alternative parameter spaces: $\mathcal{S}_0 := \{\mathbf{I}_d\}$ and

$$\mathcal{S}_1(\theta, s) := \big\{\mathbf{I}_d + \theta\big(\mathbf{v}\mathbf{v}^T - \mathbf{I}_d\big) \mid \theta \in (0, 1), \forall j : v_j \in \{\pm 1, 0\}, \|\mathbf{v}\|_2^2 = s\big\}.$$

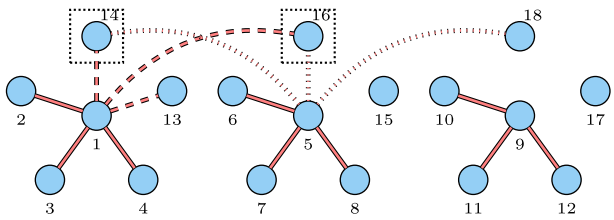


FIG. 5. *Test for maximum degree* $\mathcal{G}_0 = \{G \mid d_{\max}(G) \leq s_0\}$ *and* $\mathcal{G}_1 = \{G \mid d_{\max}(G) \geq s_1\}$ *with* $s_0 = 3$ *and* $s_1 = 6$. *We split the vertices into two parts* $\{1, \ldots, \lfloor\sqrt{d}\rfloor\}$ *and* $\{\lfloor\sqrt{d}\rfloor + 1, \ldots, d\}$. *We use the first part of vertices to construct* s_0*-star graphs as* G_0 (*visualized with solid edges*). *To construct the divider* $\boldsymbol{C}$, *we select any* $s_1 - s_0$ *vertices* (*e.g., vertices* 13, 14, 16) *from the second vertices part* $\{\lfloor\sqrt{d}\rfloor + 1, \ldots, d\}$ *and connect them to any center of the* s_0*-star graphs in* G_0 (*e.g., vertex* 1). *This gives us one of the edge sets* $S \in \boldsymbol{C}$ [*e.g.,* $S = \{(1, 13), (1, 14), (1, 16)\}$ *depicted in dashed edges in the figure*]. $\boldsymbol{C}$ *is comprised by all such edge sets. We depicted the vertex buffer* $\mathcal{V}_{S,S'} = \{14, 16\}$ *for* $S = \{(1, 13), (1, 14), (1, 16)\}$ *and* $S' = \{(5, 14), (5, 16), (5, 18)\}$ *and* $S \cap S' = \varnothing$.

This setup is related to that in Berthet and Rigollet (2013), Johnstone and Lu (2009). Our case is different from previous works because we parametrize the precision matrix rather than the covariance matrix, and the parametrization is distinct. Under our parametrization, the graph in the alternative hypothesis consists of a single s-clique.

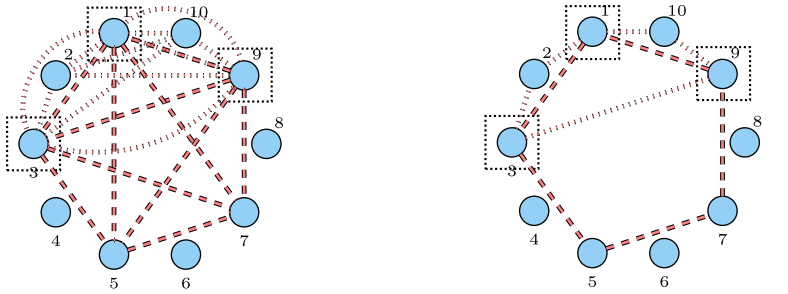
EXAMPLE 4.3 (Sparse clique detection). Suppose $s = O(d^\gamma)$ for a $\gamma < 1/2$. For values of $\theta < \frac{1}{4\sqrt{2}s} \wedge \sqrt{\frac{\log(d/s^2)}{4ns}}$, we have $\liminf_{n\to\infty} \gamma(\mathcal{S}_0, \mathcal{S}_1(\theta, s)) = 1$.

We show how Example 4.3 follows from Theorem 4.2 in Appendix E of the Supplementary Material [Neykov, Lu and Liu (2018a)]. The divider construction we use is simply drawing s vertices and connecting them to form a s-clique. Figure 6(a) illustrates two sets from the divider along with their vertex buffer.

We conclude this section by a final example on cycle detection. In this example, the sub-decomposition is $\mathcal{G}_0 = \{(\overline{V}, \varnothing)\}$, and $\mathcal{G}_1 = \{(\overline{V}, C) | C = \{(v_1, v_2), \ldots, (v_{s-1}, v_s), (v_s, v_1)\}, v_i \in \overline{V}\}$ for an integer $s \in \mathbb{N}$. We have the following example, whose proof can be found in Appendix E of the Supplementary Material [Neykov, Lu and Liu (2018a)]. We show two sets from the divider on Figure 6(b).

EXAMPLE 4.4 (Sparse cycle detection). Suppose $s = O(d^\gamma)$ for a $\gamma < 1/2$. Then for a small enough absolute constant κ if $\theta < \kappa\sqrt{\log d/n}$, we have $\liminf_{n\to\infty} \gamma(\mathcal{S}_0(\theta, s), \mathcal{S}_1(\theta, s)) = 1$.

4.3. *Upper bounds*. Algorithms matching the lower bounds developed in Sections 4.1 and 4.2 are discussed in Appendix E.1 of the Supplementary Material [Neykov, Lu and Liu (2018a)].



(a) Clique Detection Divider (b) Cycle Detection Divider

FIG. 6. *Sparse clique and cycle detection with $s = 5$. In both cases, we set $G_0 = (\overline{V}, \varnothing)$ and visualize two intersecting $S, S' \in \boldsymbol{C}$: in panel* (a), *S is the 5-clique $K_{\{1,3,5,7,9\}}$, S' is the 5-clique $K_{\{1,2,3,9,10\}}$ and $\mathcal{V}_{S,S'} = \{1, 3, 9\}$; in panel* (b), *$S$ is the 5-cycle $C_{\{1,3,5,7,9\}}$, S' is the 5-cycle $C_{\{1,2,3,9,10\}}$ and $\mathcal{V}_{S,S'} = \{1, 3, 9\}$.*

5. Discussion. In this manuscript, we provide general results for upper and lower bounds of testing graph properties. There is still room to improve the proof techniques for lower bounding. Our arguments rely only on "one-sided" alternatives, and it is possible to obtain sharper bounds by additional randomization such as in Baraud (2002). Additionally, we use the Gaussian distribution to quantify the lower bounds. We are further interested in generalizing our results to other important graphical models, such as the Ising model, in our future studies.

6. Proof of Theorem 2.1. In this section, we prove the main result of Section 2. To begin with, we give a high level picture of our proof. The argument consists of four major steps. Our first three steps will show that, there exists a constant R such that if $\theta \leq \frac{1-C^{-1}}{\sqrt{2}(\|\mathbf{A}_0\|_1+2)}$ we have

$$\gamma\big(\mathcal{S}_0(\theta,s),\mathcal{S}_1(\theta,s)\big) \geq 1-\frac{1}{2}\sqrt{\frac{1}{|\mathcal{C}|^2}\sum_{e,e'\in\mathcal{C}}\exp\Big(n\frac{(R\theta)^{2d_{G_0}(e,e')+2}}{d_{G_0}(e,e')+1}\Big)-1}. \tag{6.1}$$

To establish this result, in the first step, we select one precision matrix from the null $\mathcal{S}_0(\theta,s)$ and a set of precision matrices from the alternative $\mathcal{S}_1(\theta,s)$. In the second step, we apply Le Cam's lemma to the precision matrices constructed above to get a lower bound of $\gamma(\mathcal{S}_0(\theta,s),\mathcal{S}_1(\theta,s))$. In the third step, we establish trace perturbation inequalities to further connect the lower bound achieved in the second step to the geometric quantities of the graphs. In the fourth step, we prove the theorem by showing that the right-hand side of (6.1) goes to 1 if (2.4) is satisfied.

Step 1 (*Matrix construction*). In this step, we construct a class of precision matrices based on the null base graph G_0 and the divider set $\mathcal{C}$ and verify that these matrices indeed belong to the sets $\mathcal{S}_0(\theta,s)$ and $\mathcal{S}_1(\theta,s)$. We begin with giving the upper bound of matrix norms of adjacent inequalities. Let $\mathbf{A}_0$ be the adjacency matrix of the graph G_0. Observe that since $\mathbf{A}_0$ is symmetric, by Hölder's inequality $\|\mathbf{A}_0\|_2 \leq \sqrt{\|\mathbf{A}_0\|_1\|\mathbf{A}_0\|_\infty} = \|\mathbf{A}_0\|_1 \leq D$.

Similarly, denote with $\mathbf{A}_e$ the adjacency matrix of the graph $(\overline{V},\{e\})$ for $e\in\mathcal{C}$. Under our assumptions, it follows that $\mathbf{A}_0+\mathbf{A}_e$ is the adjacency matrix of the graph $G_e=(\overline{V},E_0\cup\{e\})$. For brevity, for any two edges $e,e'\in\mathcal{C}$ we define the shorthand notation $\mathbf{A}_{e,e'} := \mathbf{A}_0+\mathbf{A}_e+\mathbf{A}_{e'}$. Take $\mathbf{\Theta}_0=\mathbf{I}+\theta\mathbf{A}_0$, $\mathbf{\Theta}_e=\mathbf{I}+\theta(\mathbf{A}_0+\mathbf{A}_e)$, $\mathbf{\Theta}_{e,e'}=\mathbf{I}+\theta\mathbf{A}_{e,e'}$, for $e,e'\in\mathcal{C}$, $\theta>0$. By the triangle inequality for any $e,e'\in\mathcal{C}$, we have

$$\max\big(\|\mathbf{A}_0\|_2,\|\mathbf{A}_0+\mathbf{A}_e\|_2,\|\mathbf{A}_{e,e'}\|_2\big)\leq\|\mathbf{A}_0\|_2+2,$$
$$\max\big(\|\mathbf{A}_0\|_1,\|\mathbf{A}_0+\mathbf{A}_e\|_1,\|\mathbf{A}_{e,e'}\|_1\big)\leq\|\mathbf{A}_0\|_1+2.$$

Recall Definition (1.2) of the set $\mathcal{M}(s)$. Next, we make sure that the matrices $\mathbf{\Theta}_0$ and $\mathbf{\Theta}_e$ fall into $\mathcal{M}(s)$ and in addition the matrix $\mathbf{\Theta}_{e,e'}>0$. For the upper bounds,

it suffices to choose η satisfying

$$\max\bigl(\|\mathbf{\Theta}_0\|_2, \|\mathbf{\Theta}_e\|_2\bigr) \le 1 + \bigl(\|\mathbf{A}_0\|_2 + 2\bigr)\theta \le C,$$
$$\max\bigl(\|\mathbf{\Theta}_0\|_1, \|\mathbf{\Theta}_e\|_1\bigr) \le 1 + \bigl(\|\mathbf{A}_0\|_1 + 2\bigr)\theta \le L.$$

Recall that $\|\mathbf{A}_0\|_2 \le \|\mathbf{A}_0\|_1$, and $C \le L$ hence both inequalities are implied if $1 + (\|\mathbf{A}_0\|_1 + 2)\theta \le C$. This inequality holds since

$$\theta < \frac{(1 - C^{-1}) \wedge e^{-1/2}}{\sqrt{2}(D+2)} \le \frac{C-1}{(\|\mathbf{A}_0\|_1 + 2)},$$

where the last inequality is true since $D = \|\mathbf{A}_0\|_1$, and $C \ge 1$ and, therefore, $C - 1 \ge 1 - C^{-1}$. Furthermore, by Weyl's inequality,

$$\lambda_d(\mathbf{\Theta}_0), \lambda_d(\mathbf{\Theta}_e), \lambda_d(\mathbf{\Theta}_{e,e'}) \ge 1 - \theta\bigl(\|\mathbf{A}_0\|_2 + 2\bigr) \ge 1 - \theta\bigl(\|\mathbf{A}_0\|_1 + 2\bigr), \tag{6.2}$$

where λ_d denotes the smallest eigenvalue of the corresponding matrix. We want to ensure that the last term is at least C^{-1}. Since by assumption $\theta < \frac{1-C^{-1}}{\sqrt{2}(\|\mathbf{A}_0\|_1+2)}$, the above inequalities are satisfied. Furthermore, we have $G_0 \in \mathcal{G}_0$, $G_e \in \mathcal{G}_1$ for all $e \in \mathcal{C}$, and hence the induced graphs $G(\mathbf{\Theta}_0) \in \mathcal{G}_0$ and $G(\mathbf{\Theta}_e) \in \mathcal{G}_1$ for all $e \in \mathcal{C}$. This shows that $\mathbf{\Theta}_0 \in \mathcal{S}_0(\theta, s)$ and $\mathbf{\Theta}_e \in \mathcal{S}_1(\theta, s)$. We also obtain as a by-product that $\mathbf{\Theta}_{e,e'} \ge 0$.

Step 2 (*Minimax risk lower bound*). In this step, we obtain a lower bound on the minimax risk driven by Le Cam's lemma [Le Cam (1973)]. Using a determinant identity, we control the chi-square divergence by the traces of adjacency matrices' powers. Put the uniform prior on $\mathcal{C}$ and consider the models generated by $N(0, (\mathbf{\Theta}_e)^{-1})$ where $e \in \mathcal{C}$. Define

$$\overline{\mathbb{P}} = \frac{1}{|\mathcal{C}|}\sum_{e\in\mathcal{C}} \mathbb{P}_{\mathbf{\Theta}_e},$$

where $\mathbb{P}_{\mathbf{\Theta}_e}$ we define the probability measure when the data is i.i.d. $\boldsymbol{X}_i \sim N(0, (\mathbf{\Theta}_e)^{-1})$, and let $\mathbb{P}_{\mathbf{\Theta}_0}$ be the probability measure when the data is i.i.d. $\boldsymbol{X}_i \sim N(0, (\mathbf{\Theta}_0)^{-1})$. Next, by Neyman–Pearson's lemma, we have

$$\gamma(\mathcal{S}_0, \mathcal{S}_1) \ge \inf_{\psi}\bigl[\mathbb{P}_{\mathbf{\Theta}_0}(\psi = 1) + \overline{\mathbb{P}}(\psi = 0)\bigr] = 1 - \mathrm{TV}(\overline{\mathbb{P}}, \mathbb{P}_{\mathbf{\Theta}_0}), \tag{6.3}$$

where for two probability measures $P, Q \ll \lambda$ on a measurable space $(\Omega, \mathcal{A})$, TV stands for *total variation distance*, and is defined as

$$\mathrm{TV}(P, Q) = \sup_{A\in\mathcal{A}} \bigl|P(A) - Q(A)\bigr| = \frac{1}{2}\int \left|\frac{dP}{d\lambda}(\omega) - \frac{dQ}{d\lambda}(\omega)\right| d\lambda(\omega).$$

By Cauchy–Schwarz, one has

$$1 - \mathrm{TV}(\overline{\mathbb{P}}_\pi, \mathbb{P}_{\mathbf{\Theta}_0}) \ge 1 - \frac{1}{2}\sqrt{D_{\chi^2}(\overline{\mathbb{P}}_\pi, \mathbb{P}_{\mathbf{\Theta}_0})}, \tag{6.4}$$

where $D_{\chi^2}(P, Q)$ is the chi-square divergence between the measures P, Q and is defined as

$$D_{\chi^2}(P, Q) = \int \left(\frac{dP}{dQ}(\omega) - 1\right)^2 dQ(\omega) = \int \left(\frac{dP}{dQ}(\omega)\right)^2 dQ(\omega) - 1,$$

assuming that $P \ll Q$. Observe that $D_{\chi^2}(\overline{\mathbb{P}}, \mathbb{P}_{\mathbf{\Theta}_0})$ can be equivalently expressed as

$$D_{\chi^2}(\overline{\mathbb{P}}, \mathbb{P}_{\mathbf{\Theta}_0}) = \mathbb{E}_{\mathbf{\Theta}_0} L_{\mathbf{\Theta}_0}^2 - 1, \tag{6.5}$$

where $L_{\mathbf{\Theta}_0} = \frac{1}{|\mathcal{C}|} \sum_{e \in \mathcal{C}} \frac{d\mathbb{P}_{\mathbf{\Theta}_e}}{d\mathbb{P}_{\mathbf{\Theta}_0}}$ is the integrated likelihood ratio, and $\mathbb{E}_{\mathbf{\Theta}_0}$ denotes the expectation under $\boldsymbol{X}_i \sim N(0, (\mathbf{\Theta}_0)^{-1})$. Hence by (6.3) and (6.4), it suffices to obtain upper bounds on the integrated likelihood ratio in order to lower bound the minimax risk (1.3). Writing out the likelihood ratio comparing the normal distribution with precision matrix $\mathbf{\Theta}_0$ to the uniform mixture of normal distribution with precision matrix $\mathbf{\Theta}_e$ for $e \in \mathcal{C}$, we get

$$L_{\mathbf{\Theta}_0} = \frac{1}{|\mathcal{C}|} \sum_{e \in \mathcal{C}} \left(\frac{\det(\mathbf{\Theta}_e)}{\det(\mathbf{\Theta}_0)}\right)^{n/2} \prod_{i=1}^{n} \exp(-\boldsymbol{X}_i^T \theta \mathbf{A}_e \boldsymbol{X}_i / 2).$$

To calculate the chi-square distance in (6.5), next we square this expression and take its expectation under $\mathbb{P}_{\mathbf{\Theta}_0}$ to obtain

$$\begin{aligned} \mathbb{E}_{\mathbf{\Theta}_0} L_{\mathbf{\Theta}_0}^2 &= \frac{1}{|\mathcal{C}|^2} \sum_{e, e' \in \mathcal{C}} \frac{(\det(\mathbf{\Theta}_e) \det(\mathbf{\Theta}_{e'}))^{n/2}}{(\det(\mathbf{\Theta}_0))^n} \\ &\quad \times \mathbb{E}_{\mathbf{\Theta}_0} \exp\left(-\frac{\sum_{i=1}^n \boldsymbol{X}_i^T \theta (\mathbf{A}_e + \mathbf{A}_{e'}) \boldsymbol{X}_i}{2}\right) \\ &= \frac{1}{|\mathcal{C}|^2} \sum_{e, e' \in \mathcal{C}} \left(\frac{\det(\mathbf{\Theta}_e)}{\det(\mathbf{\Theta}_0)}\right)^{n/2} \left(\frac{\det(\mathbf{\Theta}_{e'})}{\det(\mathbf{\Theta}_{e,e'})}\right)^{n/2}. \end{aligned} \tag{6.6}$$

Next, we will expand the determinants above. Recall that we have ensured that $1 - \theta(\|\mathbf{A}_0\|_2 + 2) > 0$ (see 6.2). This implies

$$\theta \max\left(\|\mathbf{A}_0\|_2, \|\mathbf{A}_0 + \mathbf{A}_e\|_2, \|\mathbf{A}_0 + \mathbf{A}_{e'}\|_2, \|\mathbf{A}_{e,e'}\|_2\right) \leq 1.$$

For what follows, for a symmetric matrix $\mathbf{A}_{d \times d}$ we denote its ordered eigenvalues with $\lambda_1(\mathbf{A}) \geq \lambda_2(\mathbf{A}) \geq \cdots \geq \lambda_d(\mathbf{A})$. Let $\mathbf{A} \in \mathbb{R}^{d \times d}$ be a symmetric matrix such that $\|\mathbf{A}\|_2 \leq 1$. Then we have

$$\begin{aligned} \log\det(\mathbf{I} + \mathbf{A}) &= \sum_{j=1}^{d} \log \lambda_j(\mathbf{I} + \mathbf{A}) = \sum_{j=1}^{d} \log(1 + \lambda_j(\mathbf{A})) \\ &= \sum_{j=1}^{d} \sum_{k=1}^{\infty} (-1)^{k+1} \frac{\lambda_j^k(\mathbf{A})}{k} = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{\operatorname{Tr}(\mathbf{A}^k)}{k}. \end{aligned}$$

Using the form $\det(\mathbf{I}+\mathbf{A}) = \exp(\log\det(\mathbf{I}+\mathbf{A}))$ and plugging the above equation into (6.6), we conclude that

$$\mathbb{E}_{\mathbf{\Theta}_0} L^2_{\mathbf{\Theta}_0} = \frac{1}{|\mathcal{C}|^2} \sum_{e,e' \in \mathcal{C}} \exp\left(\frac{n}{2} \sum_{k=1}^{\infty} \frac{(-\theta)^k}{k} (T_1^k + T_2^k) \right),$$

where

$$T_1^k := \operatorname{Tr}\big[\mathbf{A}_0^k - (\mathbf{A}_0 + \mathbf{A}_e)^k\big] T_2^k := \operatorname{Tr}\big[(\mathbf{A}_{e,e'})^k - (\mathbf{A}_0 + \mathbf{A}_{e'})^k\big].$$

Step 3 (*Trace perturbation inequalities*). In this step, we control $T_1^k + T_2^k$ in terms of k and link it with the geometric quantities of the graph. We view T_1^k as the perturbation difference between $\operatorname{Tr}[\mathbf{A}_0^k]$ and $\operatorname{Tr}[(\mathbf{A}_0 + \mathbf{A}_e)^k]$ and we treat T_2^k similarly. In the following step, we aim to develop the perturbation inequalities for the trace of matrix powers.

First, we will argue that $T_1^k + T_2^k \geq 0$ for all $k \in \mathbb{N}$. To see this, recall that the trace operator of an adjacency matrix $\mathbf{M}$ satisfies

$$\operatorname{Tr}(\mathbf{M}^k) = \text{ number of all closed walks of length } k.$$

First, we consider case $e \neq e'$. Notice that all closed walks in $G(\mathbf{A}_0 + \mathbf{A}_e)$ that do not belong to $G(\mathbf{A}_0)$ have to pass through the edge e at least once. Similarly, all closed walks in $G(\mathbf{A}_0 + \mathbf{A}_{e'})$ that do not belong to $G(\mathbf{A}_0)$ have to pass through the edge e' at least once. Furthermore, all closed walks of length k passing through either e or e' belong to $G(\mathbf{A}_{e,e'})$. In addition, $G(\mathbf{A}_{e,e'})$ might contain extra closed walks passing through both e and e'. This shows

$$T_1^k + T_2^k \geq 0$$

for all k. This shows that when k is odd we have $(-\theta)^k(T_1^k + T_2^k) \leq 0$, and thus to control $\mathbb{E}_{\mathbf{\Theta}_0} L^2_{\mathbf{\Theta}_0}$ it suffices to focus only on even k.

Next, we prove that for $k < 2d_{G_0}(e, e') + 2$, we have $T_1^k + T_2^k \equiv 0$. To see this, first consider the case $e \neq e'$. Notice that the graph $G(\mathbf{A}_{e,e'})$ cannot contain paths passing through both e and e' unless $k \geq 2d_{G_0}(e, e') + 2$. To see this, notice that no even length closed walk between e and e' can exist if the length of this walk is smaller than $2d_{G_0}(e, e')$ plus the two edges e and e'. This proves our claim in the case $e \neq e'$. In the special case $e = e'$, the length of the path trivially needs to be at least of length 2 to pass through both e and e'.

We will now argue that for even $k \in \mathbb{N}$ we have $T_1^k + T_2^k \leq 2(\|\mathbf{A}_0\|_2 + 2)^k$. In fact, we will prove that $T_1^k \leq 0$ for all k and $T_2^k \leq 2(\|\mathbf{A}_0\|_2 + 2)^k$ for all even k. To see that $T_1^k \leq 0$, note that $G(\mathbf{A}_0)$ contains less closed walks than $G(\mathbf{A}_0 + \mathbf{A}_e)$.

Recall that for a symmetric matrix $\mathbf{A}_{d\times d}$ we denote its ordered eigenvalues with $\lambda_1(\mathbf{A}) \geq \lambda_2(\mathbf{A}) \geq \cdots \geq \lambda_d(\mathbf{A})$. To this end, we state a helpful result whose proof is deferred to the Supplementary Material [Neykov, Lu and Liu (2018a)].

LEMMA 6.1. *For two symmetric $m \times m$ matrices $\mathbf{A}$ and $\mathbf{B}$, and any constants $c_1 \geq c_2 \geq \cdots \geq c_m$, and a permutation σ on $\{1, \ldots, m\}$ we have*

$$\sum_{j=1}^{m} c_{\sigma(j)} \lambda_j(\mathbf{A}+\mathbf{B}) \leq \sum_{j=1}^{m} c_{\sigma(j)} \lambda_j(\mathbf{A}) + \sum_{j=1}^{m} c_j \lambda_j(\mathbf{B}).$$

Using Lemma 6.1 for the matrices $\mathbf{A} = \mathbf{A}_e, \mathbf{B} = \mathbf{A}_0 + \mathbf{A}_{e'}$ with constants

$$c_{\sigma(j)} = \operatorname{sign}\big(\lambda_j(\mathbf{A}_{e,e'}) - \lambda_j(\mathbf{A}_e)\big)\big|\lambda_j(\mathbf{A}_{e,e'}) - \lambda_j(\mathbf{A}_e)\big|^{k-1},$$

we obtain

$$\begin{aligned}\sum_{j=1}^{d} \big|\lambda_j(\mathbf{A}_{e,e'}) - \lambda_j(\mathbf{A}_e)\big|^k &\leq \sum_{j=1}^{d} c_j \lambda_j(\mathbf{A}_0 + \mathbf{A}_{e'}) \\ &\leq \left[\sum_{j=1}^{d} |c_j|^{\frac{k}{k-1}}\right]^{\frac{k-1}{k}} \left[\sum_{j=1}^{d} \big|\lambda_j(\mathbf{A}_0 + \mathbf{A}_{e'})\big|^k\right]^{\frac{1}{k}},\end{aligned}$$

where the last inequality follows by Hölder's inequality. We conclude that

$$\sum_{j=1}^{d} \big|\lambda_j(\mathbf{A}_{e,e'}) - \lambda_j(\mathbf{A}_e)\big|^k \leq \sum_{j=1}^{d} \big|\lambda_j(\mathbf{A}_0 + \mathbf{A}_{e'})\big|^k. \tag{6.7}$$

Next, observe that the negative adjacency matrix $-\mathbf{A}_e$ of the single edge graph $(\overline{V}, \{e\})$ has very simple eigenvalue structure: $1, -1$ and $d-2$ zeros. Hence we conclude that for even k

$$\begin{aligned}T_2^k &= \operatorname{Tr}\big((\mathbf{A}_{e,e'})^k\big) - \operatorname{Tr}\big((\mathbf{A}_0 + \mathbf{A}_{e'})^k\big) \\ &= \sum_{j=1}^{d} \big|\lambda_j(\mathbf{A}_{e,e'})\big|^k - \sum_{j=1}^{d} \big|\lambda_j(\mathbf{A}_0 + \mathbf{A}_{e'})\big|^k \\ &\leq \big|\lambda_1(\mathbf{A}_{e,e'})\big|^k - \big|\lambda_1(\mathbf{A}_{e,e'}) - 1\big|^k + \big|\lambda_d(\mathbf{A}_{e,e'})\big|^k - \big|\lambda_d(\mathbf{A}_{e,e'}) + 1\big|^k \\ &\leq \big|\lambda_1(\mathbf{A}_{e,e'})\big|^k + \big|\lambda_d(\mathbf{A}_{e,e'})\big|^k \leq 2\|\mathbf{A}_{e,e'}\|_2^k \leq 2\big(\|\mathbf{A}_0\|_2 + 2\big)^k.\end{aligned}$$

The last shows that indeed $T_1^k + T_2^k \leq 2(\|\mathbf{A}_0\|_2 + 2)^k$ as claimed. Putting everything together, we obtain

$$\begin{aligned}\sum_{k=1}^{\infty} \frac{(-\theta)^k}{k}\big[T_1^k + T_2^k\big] &\leq \sum_{2|k, k \geq 2d_{G_0}(e,e')+2}^{\infty} \frac{\theta^k}{k}\big[T_1^k + T_2^k\big] \\ &\leq \sum_{2|k, k \geq 2d_{G_0}(e,e')+2}^{\infty} \frac{2((\|\mathbf{A}_0\|_2 + 2)\theta)^k}{k}\end{aligned}$$

$$\leq \frac{2((\|\mathbf{A}_0\|_2+2)\theta)^{2d_{G_0}(e,e')+2}}{(2d_{G_0}(e,e')+2)(1-(\theta(\|\mathbf{A}_0\|_2+2))^2)}$$
$$\leq \frac{2((\|\mathbf{A}_0\|_2+2)\theta)^{2d_{G_0}(e,e')+2}}{d_{G_0}(e,e')+1},$$

where in the last inequality we used the fact that $\theta \leq \frac{1}{\sqrt{2}(\|\mathbf{A}_0\|_2+2)}$ which follows by the requirements on θ. This completes the proof of (6.1) where $R = \sqrt{2}(\|\mathbf{A}_0\|_2+2)$.

Step 4 (*Rate control*). The goal in this final step is to show that if (2.4) holds, the minimax risk

$$\liminf_{n\to\infty} \gamma\big(\mathcal{S}_0(\theta,s),\mathcal{S}_1(\theta,s)\big) = 1.$$

The proof is technical, but the high-level idea is to clip the first $\log|\mathcal{C}|$ degrees in (6.1) and deal with two separate summations. It turns out that the scaling assumed on θ in (2.4) is precisely enough to control both the summation of all degrees below $\log|\mathcal{C}|$ and the summation of all degrees above $\log|\mathcal{C}|$. Define the following quantities:

$$K_r := \big|\{(e,e') \mid e,e' \in \mathcal{C}, d_{G_0}(e,e') = r\}\big|,$$

where (e,e') are unordered edge pairs, and observe that $\sum_{r\geq 0} K_r = \binom{|\mathcal{C}|}{2} + |\mathcal{C}|$ by definition. We will in fact, first show that if $\theta \leq \kappa\sqrt{\frac{\log|\mathcal{C}|}{n}}$ for some small κ, and

$$\sum_{r=0}^{\lfloor \log|\mathcal{C}|\rfloor} K_r = O\big(|\mathcal{C}|^{2-\gamma}\big) \tag{6.8}$$

for some $1/2 < \gamma \leq 1$, then $\liminf \gamma(\mathcal{S}_0(\theta,s),\mathcal{S}_1(\theta,s)) = 1$ provided that $|\mathcal{C}| \to \infty$. We will then derive the theorem as a corollary to this observation.

By (6.1), it suffices to control

$$\underbrace{\frac{2}{|\mathcal{C}|^2}\sum_{(e,e'):d_{G_0}(e,e')\geq 1} \exp\left(n\frac{\overline{\theta}^{2d_{G_0}(e,e')+2}}{d_{G_0}(e,e')+1}\right)}_{I_1} + \underbrace{\frac{2K_0-|\mathcal{C}|}{|\mathcal{C}|^2}\exp(n\overline{\theta}^2)}_{I_2}, \tag{6.9}$$

where we will write $\overline{\theta}$ for $R\theta$ for brevity.

First, observe that since $\theta < \frac{1-C^{-1}}{\sqrt{2}(D+2)}$ then we have

$$\overline{\theta} < R\frac{(1-C^{-1})\wedge e^{-1/2}}{\sqrt{2}(D+2)} \leq \big(1-C^{-1}\big)\wedge e^{-1/2} < e^{-1/2} < 1.$$

We will show that when $\bar{\theta}$ is small, (6.9) is bounded by 1 asymptotically, which in turn suffices to show that $\liminf \gamma(\mathcal{S}_0(\theta,s),\mathcal{S}_1(\theta,s)) = 1$. Notice that $\overline{\theta}$ and θ

are the same quantity up to the constant R, and hence $\theta < \kappa\sqrt{\frac{\log|\mathcal{C}|}{n}}$ is equivalent to $\overline{\theta} < \bar{\kappa}\sqrt{\frac{\log|\mathcal{C}|}{n}}$ for some sufficiently small $\bar{\kappa}$. We will require

$$\bar{\kappa} < \sqrt{\gamma} \wedge \sqrt{2\gamma/c_0} \wedge (ec_0)^{-1/2}.$$

Observe that since $K_0 = O(|\mathcal{C}|^{2-\gamma})$, $\bar{\kappa}^2 < \gamma$, and $\overline{\theta} < \bar{\kappa}\sqrt{\frac{\log|\mathcal{C}|}{n}}$ we have

$$I_2 \leq \frac{(2K_0 - |\mathcal{C}|)|\mathcal{C}|^{\bar{\kappa}^2}}{|\mathcal{C}|^2} \to 0.$$

Next, we tackle the term I_1 in (6.9). We will show that since $\overline{\theta} < 1$ by assumption, this term goes to 1:

$$\begin{aligned} I_1 &= \frac{2}{|\mathcal{C}|^2}\sum_{r\geq 1} K_r \exp\bigl(n\overline{\theta}^{2r+2}/(r+1)\bigr) = \frac{2}{|\mathcal{C}|^2}\sum_{r=1}^{d-1} K_r|\mathcal{C}|^{\bar{\kappa}^2\overline{\theta}^{2r}/(r+1)} + \frac{2K_\infty}{|\mathcal{C}|^2} \\ &< \frac{2}{|\mathcal{C}|^2}\sum_{r=1}^{d-1} K_r|\mathcal{C}|^{\overline{\theta}^{2r}/(r+1)} + \frac{2K_\infty}{|\mathcal{C}|^2}, \end{aligned}$$

where the last inequality follows by the fact that $\bar{\kappa}^2 < \gamma < 1$. Splitting out the first $\lfloor \log|\mathcal{C}| \rfloor$ terms out of this summation yields

$$I_1 < \underbrace{\frac{2}{|\mathcal{C}|^2}\sum_{r=1}^{\lfloor\log|\mathcal{C}|\rfloor} K_r|\mathcal{C}|^{\overline{\theta}^{2r}/(r+1)}}_{I_{11}} + \underbrace{\frac{2}{|\mathcal{C}|^2}\sum_{r=\lfloor\log|\mathcal{C}|\rfloor+1}^{d-1} K_r|\mathcal{C}|^{\overline{\theta}^{2r}/(r+1)} + \frac{2K_\infty}{|\mathcal{C}|^2}}_{I_{12}}.$$

The first term is bounded by $I_{11} \leq 2(\sum_{r=1}^{\lfloor\log|\mathcal{C}|\rfloor} K_r)\frac{|\mathcal{C}|^{\overline{\theta}^2/2}}{|\mathcal{C}|^2} = o(1)$, where we used $(\sum_{r=2}^{\lfloor\log|\mathcal{C}|\rfloor} K_r) = O(|\mathcal{C}|^{2-\gamma})$ and the fact that $\overline{\theta}^2/2 < 1/2 < \gamma$. Next, we will argue that $|\mathcal{C}|^{\overline{\theta}^{2r}/(r+1)} \leq 1 + 3\overline{\theta}^{2r}(|\mathcal{C}|-1)/(r+1)$. This follows by

$$\exp\bigl(\log(|\mathcal{C}|)\overline{\theta}^{2r}/(r+1)\bigr) \leq 1 + 3\log(|\mathcal{C}|)\overline{\theta}^{2r}/(r+1) \leq 1 + 3\bigl(|\mathcal{C}|-1\bigr)\overline{\theta}^{2r}/(r+1),$$

with the first inequality holding when $\log(|\mathcal{C}|)\overline{\theta}^{2r}/(r+1) < 1$, which is true since $\overline{\theta} < 1$, and $r \geq \lfloor \log|\mathcal{C}| \rfloor + 1$. Hence we have

$$\begin{aligned} I_{12} &\leq \frac{2}{|\mathcal{C}|^2}\sum_{r=\lfloor\log|\mathcal{C}|\rfloor+1}^{d-1} K_r\bigl(1 + 3\bigl(|\mathcal{C}|-1\bigr)\overline{\theta}^{2r}/(r+1)\bigr) + \frac{2K_\infty}{|\mathcal{C}|^2} \\ &\leq \left(1 - \frac{O(|\mathcal{C}|^{2-\gamma})}{|\mathcal{C}|^2}\right) + \frac{6(|\mathcal{C}|-1)}{|\mathcal{C}|^2}\sum_{r=\lfloor\log|\mathcal{C}|\rfloor+1}^{d-1}\frac{K_r}{r}\overline{\theta}^{2r} \\ &\leq \left(1 - \frac{O(|\mathcal{C}|^{2-\gamma})}{|\mathcal{C}|^2}\right) + \frac{6\overline{\theta}^{2\lfloor\log|\mathcal{C}|\rfloor+2}}{1-\overline{\theta}^2}|\mathcal{C}|^{-1}\max_{\lfloor\log|\mathcal{C}|\rfloor+1\leq r}\frac{K_r}{r}. \end{aligned}$$

Paying closer attention to the second term, we have

$$\frac{6\overline{\theta}^{2\lfloor \log|\mathcal{C}|\rfloor+2}}{1-\overline{\theta}^2}|\mathcal{C}|^{-1}\max_{\lfloor \log|\mathcal{C}|\rfloor+1\le r}\frac{K_r}{r}\le\frac{6}{1-e^{-1}}\overline{\theta}^{2\lfloor \log|\mathcal{C}|\rfloor+2}\frac{\binom{|\mathcal{C}|}{2}+|\mathcal{C}|}{|\mathcal{C}|}$$
$$\le\frac{6}{1-e^{-1}}\overline{\theta}^{2\lfloor \log|\mathcal{C}|\rfloor+2}|\mathcal{C}|=o(1),$$

with the last equality holds since $\overline{\theta}<\exp(-1/2)$, as we required. This combined with (6.1) concludes the proof of $\liminf_n \gamma(\mathcal{S}_0(\theta,s),\mathcal{S}_1(\theta,s))=1$, when $\theta\le\kappa\sqrt{\frac{\log|\mathcal{C}|}{n}}$.

Finally, notice that any subset of a divider $\mathcal{C}$ is trivially a divider. Hence we can apply what we just showed to the set $N_{\log|\mathcal{C}|}\subset\mathcal{C}$ —the maximal $\log|\mathcal{C}|$-packing of $\mathcal{C}$. Evaluating the constants K_r on $N_{\log|\mathcal{C}|}$ gives

$$K_0=|N_{\log|\mathcal{C}|}|,\ K_r=0\qquad\text{for all } r\le\lfloor\log|\mathcal{C}|\rfloor,$$

and since $\lfloor N_{\log|\mathcal{C}|}\rfloor\le\lfloor\log|\mathcal{C}|\rfloor$ we conclude that $\sum_{r=0}^{\lfloor N_{\log|\mathcal{C}|}\rfloor}K_r=|N_{\log|\mathcal{C}|}|=O(|N_{\log|\mathcal{C}|}|^{2-\gamma})$ for any $0<\gamma\le1$.

Acknowledgments. The authors would like to thank the Editor, Associate Editor and two referees for their suggestions, comments and remarks which led to significant improvements in the presentation of this manuscript.

SUPPLEMENTARY MATERIAL

Supplement to "Combinatorial inference for graphical models" (DOI: 10.1214/17-AOS1650SUPP; .pdf). The Supplementary Material contains proofs and derivations of some of the main results of the paper, as well as simulation results and real data analysis.

REFERENCES

ADDARIO-BERRY, L., BROUTIN, N., DEVROYE, L. and LUGOSI, G. (2010). On combinatorial testing problems. *Ann. Statist.* **38** 3063–3092. MR2722464

ARIAS-CASTRO, E., BUBECK, S. and LUGOSI, G. (2012). Detection of correlations. *Ann. Statist.* **40** 412–435. MR3014312

ARIAS-CASTRO, E., BUBECK, S. and LUGOSI, G. (2015). Detecting positive correlations in a multivariate sample. *Bernoulli* **21** 209–241. MR3322317

ARIAS-CASTRO, E., CANDÈS, E. J. and DURAND, A. (2011). Detection of an anomalous cluster in a network. *Ann. Statist.* **39** 278–304. MR2797847

ARIAS-CASTRO, E., CANDÈS, E. J. and PLAN, Y. (2011). Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *Ann. Statist.* **39** 2533–2556. MR2906877

ARIAS-CASTRO, E., BUBECK, S., LUGOSI, G. and VERZELEN, N. (2015). Detecting Markov random fields hidden in white noise. Preprint. Available at arXiv:1504.06984.

BARAUD, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli* **8** 577–606. MR1935648

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

BERTHET, Q. and RIGOLLET, P. (2013). Optimal detection of sparse principal components in high dimension. *Ann. Statist.* **41** 1780–1815. MR3127849

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

CHEN, M., REN, Z., ZHAO, H. and ZHOU, H. (2016). Asymptotically normal and efficient estimation of covariate-adjusted Gaussian graphical model. *J. Amer. Statist. Assoc.* **111** 394–406. MR3494667

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

DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. MR2065195

DUBHASHI, D. and RANJAN, D. (1998). Balls and bins: A study in negative dependence. *Random Structures Algorithms* **13** 99–124. MR1642566

FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441.

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

INGSTER, Y. I. (1982). Minimax nonparametric detection of signals in white Gaussian noise. *Problemy Peredachi Informatsii* **18** 61–73. MR0689340

INGSTER, Y. I., TSYBAKOV, A. B. and VERZELEN, N. (2010). Detection boundary in sparse regression. *Electron. J. Stat.* **4** 1476–1526. MR2747131

JANKOVÁ, J. and VAN DE GEER, S. (2015). Confidence intervals for high-dimensional inverse covariance estimation. *Electron. J. Stat.* **9** 1205–1229. MR3354336

JOAG-DEV, K. and PROSCHAN, F. (1983). Negative association of random variables, with applications. *Ann. Statist.* **11** 286–295. MR0684886

JOHNSTONE, I. M. and LU, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.* **104** 682–693. MR2751448

LAM, C. and FAN, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. Statist.* **37** 4254–4278. MR2572459

LE CAM, L. (1973). Convergence of estimates under dimensionality restrictions. *Ann. Statist.* **1** 38–53. MR0334381

LIU, W. (2013). Gaussian graphical model estimation with false discovery rate control. *Ann. Statist.* **41** 2948–2978. MR3161453

LIU, H., LAFFERTY, J. and WASSERMAN, L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *J. Mach. Learn. Res.* **10** 2295–2328. MR2563983

MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. MR2278363

NEYKOV, M., LU, J. and LIU, H. (2018a). Supplement to "Combinatorial inference for graphical models." DOI:10.1214/17-AOS1650SUPP.

NEYKOV, M., NING, Y., LIU, J. S. and LIU, H. (2018b). A unified theory of confidence regions and testing for high dimensional estimating equations. *Statist. Sci* **33** 427–443. MR3843384

RASKUTTI, G., YU, B., WAINWRIGHT, M. J. and RAVIKUMAR, P. K. (2008). Model selection in Gaussian graphical models: High-dimensional consistency of ℓ_1-regularized MLE. In *Advances in Neural Information Processing Systems* 1329–1336.

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

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

ROMANO, J. P. and WOLF, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *J. Amer. Statist. Assoc.* **100** 94–108.

SANTHANAM, N. P. and WAINWRIGHT, M. J. (2012). Information-theoretic limits of selecting binary graphical models in high dimensions. *IEEE Trans. Inform. Theory* **58** 4117–4134.

VERZELEN, N. and VILLERS, F. (2010). Goodness-of-fit tests for high-dimensional Gaussian linear models. *Ann. Statist.* **38** 704–752. MR2604699

WANG, W., WAINWRIGHT, M. J. and RAMCHANDRAN, K. (2010). Information-theoretic bounds on model selection for Gaussian Markov random fields. In *ISIT* 1373–1377.

YANG, Y. and BARRON, A. (1999). Information-theoretic determination of minimax rates of convergence. *Ann. Statist.* **27** 1564–1599. MR1742500

YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 49–67. MR2212574

ZHANG, C.-H. and ZHANG, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 217–242. MR3153940

M. NEYKOV
DEPARTMENT OF STATISTICS
AND DATA SCIENCE
CARNEGIE MELLON UNIVERSITY
PITTSBURGH, PENNSYLVANIA 15213
USA
E-MAIL: mneykov@stat.cmu.edu

J. LU
DEPARTMENT OF OPERATIONS RESEARCH
AND FINANCIAL ENGINEERING
PRINCETON UNIVERSITY
PRINCETON, NEW JERSEY 08540
USA
E-MAIL: junweil@princeton.edu

H. LIU
DEPARTMENT OF ELECTRICAL ENGINEERING
AND COMPUTER SCIENCE
AND
DEPARTMENT OF STATISTICS
NORTHWESTERN UNIVERSITY
EVANSTON, ILLINOIS 60208
USA
E-MAIL: hanliu@northwestern.edu