Deviance Tests for Graph Estimation From Multi-Attribute Gaussian Data

Jitendra K. Tugnait, *Life Fellow, IEEE*

***Abstract*—We consider the problem of inferring the conditional independence graph (CIG) of Gaussian vectors from multi-attribute data. Most existing methods for graph estimation are based on single-attribute models where one associates a scalar random variable with each node. In multi-attribute graphical models, each node represents a random vector. For single-attribute graphical models, considerable body of work exists where one first tests for exclusion of each edge from the saturated model, and then infers the CIG. In this paper, we propose and analyze a deviance test based on generalized likelihood ratio, for edge exclusion in multi-attribute Gaussian graphical models. The null distribution of the test statistic is derived to allow analytical calculation of the test threshold. An alternative computationally fast version of the deviance test statistic is also derived. Numerical results based on synthetic as well as real data are presented.**

***Index Terms*—Multi-attribute data, graphical model selection, undirected graph, generalized likelihood ratio test, deviance test.**

I. Introduction

GRAPHICAL interaction models (graphical models, in short) are an important and useful tool for analyzing multivariate data [1]–[4]. Graphical modeling is a form of multivariate analysis where one uses graphs to represent models. A central concept is that of conditional independence. Given a collection of random variables, one wishes to assess the relationship between two variables, conditioned on the remaining variables. In graphical models, graphs are used to display the conditional independence structure of the variables.

Consider a graph $\mathcal{G} = (V, \mathcal{E})$ with a set of p vertices (nodes) $V = \{1, 2, \ldots, p\} = [1, p]$, and a corresponding set of (undirected) edges $\mathcal{E} \subseteq V \times V$. Given a (real-valued) random vector $\boldsymbol{x} = [x_1\ x_2\ \cdots\ x_p]^\top$, in the corresponding graph $\mathcal{G}$, each variable x_i is represented by a node (i in V), and associations between variables x_i and x_j are represented by edges between nodes i and j of $\mathcal{G}$. In a conditional independence graph (CIG), there is no edge between nodes i and j if and only if (iff) x_i and x_j are conditionally independent given the remaining $p-2$ variables x_ℓs.

Manuscript received December 26, 2019; revised July 22, 2020; accepted September 8, 2020. Date of publication September 11, 2020; date of current version October 7, 2020. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Mark A. Davenport. This work was supported by National Science Foundation under Grants CCF-1617610 and ECCS-2040536.

The author is with the Department of Electrical & Computer Engineering, 200 Broun Hall, Auburn University, Auburn, AL 36849 USA (e-mail: tugnajk@eng.auburn.edu).

Digital Object Identifier 10.1109/TSP.2020.3023575

Suppose real-valued $\boldsymbol{x}$ is multivariate Gaussian with positive-definite covariance matrix $\boldsymbol{\Sigma}$ and inverse covariance matrix (also known as precision matrix or concentration matrix) $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$. Then Ω_{ij}, the (i,j)-th element of $\boldsymbol{\Omega}$, is zero iff x_i and x_j are conditionally independent [1, Proposition 5.2]. A CIG $\mathcal{G}$ and real multivariate Gaussian $\boldsymbol{x}$ specify a real-valued Gaussian graphical model for $\boldsymbol{x}$ where the distribution of $\boldsymbol{x}$ obeys the conditional independence restrictions of CIG $\mathcal{G}$ [1, Sec. 5.2], [2, p. 165]. Such models for real-valued $\boldsymbol{x}$ have been extensively studied [4]–[8].

A vast majority of work on estimating undirected graphical models is focused on cases where each node represents a scalar random variable. In many applications, there may be more than one random variable associated with a node. This class of graphical models have been called multi-attribute graphical models in [9]–[11]. In [11] real data examples involving studies of biological regulatory networks and Alzheimer's disease have been presented. Similarly, [9] considers genomics examples involving gene and protein profiles. As noted in [11], "Multi-attribute data appear naturally in social media and scientific data analysis. For example, in a study of social networks, one may use personal information, including demographics, interests, and many other features, as nodal attributes." In [12], improper complex Gaussian graphical models have been modeled as graphical models with two variables per node: the real and imaginary parts of an improper complex random variable. In this paper, we generalize the approach of [12] to the case where there are arbitrary number of random variables per node. In particular, we consider p random vectors $\boldsymbol{z}_i \in \mathbb{R}^m$, $i = 1, 2, \ldots, p$, $m \geq 2$. We associate $\boldsymbol{z}_i$ with the ith node of an undirected graph $\mathcal{G} = (V, \mathcal{E})$ where $V = [1, p]$, and $\mathcal{E} \subseteq V \times V$ is the set of edges that describe the conditional dependencies among vectors $\{\boldsymbol{z}_i,\ i \in V\}$. As in the scalar case ($m = 1$), there is no edge between node i and node j in $\mathcal{G}$ iff random vectors $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$ are conditionally independent given the remaining $p-2$ vectors.

We now discuss, in some detail, some specific cases where use of multi-attribute models is relevant.

- *Gene Regulatory Networks [9]–[11], [13], [14]:* Antibodies are proteins, and proteins are encoded by genes. Expression of a gene encoding for a protein can be measured either at the transcriptome level, in terms of its quantity of mRNA (messenger ribonucleic acid molecule), or at the protein (proteomic) level, in terms of the concentration of the associated protein. The US National Cancer Institute (NCI) website provides the NCI-60 database for 60 human tumor cell lines. It consists of different molecular profiles on a

1053-587X © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Jitendra Tugnait. Downloaded on October 08,2020 at 21:51:17 UTC from IEEE Xplore. Restrictions apply.

panel of 60 diverse human cancer cell lines. In particular, protein profiles (i.e., normalized reverse-phase lysate arrays (RPLA) for 92 antibodies) and gene profiles (i.e., normalized RNA microarray intensities from Human Genome U95 Affymetrix chip-set for >9000 genes) have been used in [9]–[11], [13], [14] for graph modeling to investigate links between various proteins/genes based on their two profiles. In [9] a network with 91 nodes (p $=91$ genes), each with two (m=2) attributes comprised of protein and gene profiles, is considered based on $n=60$ samples (cell lines). Since these molecular profiles are on the same set of biological samples, it is argued in [9] (also by others) that the multi-attribute graphical model "proposes a consensus version of the interactions at hand in the cell, and one which is hopefully more robust to noise," compared with single-attribute models based on protein profile alone or gene profile alone. Similar conclusions are reached in [10], [11], [13], [14]. Note that while [9]–[11], [13] consider associations based on conditional independence (Markov networks), [14] deals with association networks (which confound the direct interactions with indirect ones since they only represent marginal associations).

- *Color Image Graphs:* In [15], [16] image graphs for grayscale images are inferred for modeling dependence of a pixel on neighboring pixels; here one has one variable per pixel node. These approaches do not apply to color images where one has three variables (RGB color components) per pixel node. Image graphs for color images is an example of multi-attribute graphical models with a pixel represented by a graph node and three attributes (RGB components) per pixel. Graph-based transform coding and its potential application to signal/image compression is discussed in [17] which is a potential application of image graphs; see also [18], [19]. These contributions are restricted to single attribute models (grayscale images). For color images, multi-attribute graphical modeling is more appropriate to explore dependence of a pixel on neighboring pixels.
- *GDP Network Analysis [13]:* Connections among different industries in the US are explored in [13] to see if the GDP (gross domestic product) of one industry has some effect on that of other industries. Regional GDP data, from the first quarter of 2005 to the second quarter of 2016, available from U.S. Department of Commerce website for 8 regions (New England, Mideast, Great Lakes, etc.) and 20 industries (utilities, construction, manufacturing, etc.) are used resulting in a multi-attribute graph with $p=20$ nodes (industries) and $m=8$ attributes (regions). Exploration in [13] takes regions into consideration, since significant differences in relations may exist because of regional characteristics, which are not possible to capture using only national data.
- *Air Pollutants Interaction Graphs [20]:* In [20] daily time series data from Hong Kong to analyze air pollution of Hong Kong via single attribute graphical models is considered, following the earlier work of [21] based on dependent time series. The time series data of the daily average for four pollutants, sulfer dioxide (SO_2), nitrous dioxide (NO_2), ozone (O_3) and respirable suspended particulates (RSP), from September 2010 to September 2014 over three monitoring stations located at Tsuen Wan (TW), Tap Mun (TM), and Tung Chung (TC) in Hong Kong, are used in [20] to construct a time series graph of $p=12$ nodes with $m=1$ attribute. If the objective is to study conditional dependencies between various pollutants, then a more appropriate model would be a multi-attribute model with $p=4$ pollutant nodes, each with $m=3$ attributes, reflecting measurements at the three monitoring stations. Note that the approaches of [20], [21] (and others referenced therein) do not assume i.i.d. time series.

Suppose we have N i.i.d. observations $\boldsymbol{x}(t)$, $t=0,1,\ldots,N-1$, of $\boldsymbol{x}$. The main objective in graph estimation is to determine if $\{i,j\}\in\mathcal{E}$, given data $\{\boldsymbol{x}(t)\}_{t=0}^{N-1}$. One may classify prior work on graph estimation into two broad categories: low-dimensional models [1], [2], [6], [22]–[24] where $p\ll N$, and high-dimensional models [3]–[5], [7], [8], [25]–[31] where p is of the order of N (or larger). Since in high-dimensional models, p is of the order of N (or larger), the sample covariance matrix may not be invertible or may be quite ill-conditioned. In order to regularize the problem, it is typically assumed in works on high-dimensional models that the non-zero elements of the inverse covariance matrix are sparse. On the other hand, approaches designed for low-dimensional models fail when dealing with $p>N$. In particular, in our approach to multi-attribute graphical modeling, we require $mp\leq N$ so that the sample covariance matrix is invertible with probability one for zero-mean random vectors. This paper is limited to the low-dimensional case. Of the four specific applications discussed earlier, color image graphs and air pollutants interaction graphs fit the definition of low-dimensional graphical models. In Section V-B of the paper, we present real data results dealing with color image graphs.

Early works in high-dimensional settings, such as [3]–[5], [7], [8], [25], [26], are concerned with consistent estimation of the inverse covariance matrix and/or edges in the graphical model. They lack classical measures of uncertainty and statistical significance, such as confidence intervals and p-values, needed for classical hypothesis testing and estimator quality quantification. More recent work (see [27]–[31] and references therein; [29] is a recent overview) has dealt with confidence intervals for estimators of elements of inverse covariance [28]–[31], and hypothesis testing for edge connections [27], [30], [31]. The results are asymptotic in nature with both number of nodes p and sample size N becoming large. In case of hypothesis testing for edge connections for the entire graph, which requires multiple testing, the approaches of [27], [30], [31] yield asymptotic control of false discovery rates. While [27], [30] investigate multiple testing of individual entries of a precision matrix, [31] investigates multiple testing for conditional dependence between subgroups of nodes, i.e., submatrices of the precision matrix. The approach of [31] is relevant to multi-attribute graphical models, and this is discussed in more detail later in Section III-B (see Remark 1 therein). Since we deal with low-dimensional models, our problem is simpler, allowing us to derive exact probability distribution of the test statistic for testing of an edge connection

Authorized licensed use limited to: Jitendra Tugnait. Downloaded on October 08,2020 at 21:51:17 UTC from IEEE Xplore. Restrictions apply.

for finite samples ($N \geq mp$), and then to use this result to control false discovery rate for finite samples, following the Benjamini-Hochberg approach [38].

The prior work reported on multi-attribute graphical models [9]–[11], [13] all address high-dimensional models. They do not consider hypothesis testing to quantify edge connections in the graphical model. Low-dimensional models with two variables per node, are addressed in [12]. In this paper, we extend the approach of [12] to the general case of arbitrary number of attributes per node. Since we deal with low-dimensional models, we do not need to assume sparsity of connected edges and some regularity conditions as in the high-dimensional case [9]–[11], [13], [31], but we do require $N > mp$ so that sample covariance is invertible (in fact, we need $N \gg mp$). Our Lemma 1 generalizes [12, Lemma 2] from $m = 2$ to an arbitrary $m \geq 2$. Lemma 3 of this paper is a key result that helps generalize [12, Theorem 1] from $m = 2$ to an arbitrary $m \geq 2$ in our Theorem 1. There is no counterpart of our Theorem 2 in [12]. Also, Lemma 3 is not in [1], and neither is our Theorem 2. Finally, our Lemma 5 generalizes [12, Lemma 7] from $m = 2$ to an arbitrary $m \geq 2$.

Our approach consists of constructing an associated single-attribute Gaussian graphical model, where we test for joint exclusion of m^2 edges in this single-attribute model corresponding to a single edge in the multi-attribute model. Our proposed deviance test is a generalized likelihood ratio test that yields a specified significance level for finite samples. One could use existing single edge exclusion tests [1, Sec. 5.3.3] to test these m^2 edges, but only in a multiple testing set-up where one tests each edge separately, one-by-one. To achieve a specified overall significance level for testing m^2 edges (equivalent to one-edge in multi-attribute model), one has to use methods such as the Bonferroni method ([24] and [37, Sec. 10.7]) of threshold selection. This results in a significant loss in power of the test. Alternatively, one could estimate m separate models for each attribute separately using existing single edge exclusion tests, and then combine them into a single model; how to combine is not clear [9]–[11], [13]. We illustrate these approaches via a simulation example in Section V-A1 which shows that, at least for the presented example, our proposed approach significantly outperforms these two alternatives.

The rest of the paper is organized as follows. The system model is presented in Section II where we describe the multi-attribute graphical model with m random variables per node, in more detail in Section II-A, and then a binary hypothesis testing problem is formulated in Section II-B using the associated single-attribute Gaussian graphical model, where we test for exclusion/inclusion of m^2 edges in the single-attribute model corresponding to a single edge in the multi-attribute graphical model. A deviance test (a generalized likelihood ratio test (GLRT)) is derived in Section III-B based on [1, Proposition 5.14], which is discussed in Section III-A. The null distribution of the test statistic is specified explicitly in Section III-C, which allows for analytical calculation of the test threshold. A significant simplification of the GLRT statistic of Section III-A is presented in Section IV, and it requires $\mathcal{O}((mp)^3)$ flops compared to $\mathcal{O}((mp)^5)$ flops needed in Section III-A. Numerical results based on synthetic and real data are presented in Section V. In Appendix A we review some basic definitions and useful results from [1], [32] concerning simple undirected graphs, that we use extensively in the paper. In Appendix B we recall some existing results on edge exclusion tests for selection of single-attribute Gaussian graphical models, whose formulation offers a road map for the formulation of the multiple-edge binary hypothesis testing problem formulated in Section II, and solved in Section III. Lemmas 3 and 4 and Theorem 2, stated in Sections III-B, III-C and IV, respectively, are proved in Appendix C.

A. Notation and Terminology

We use $\boldsymbol{S} \succeq 0$ and $\boldsymbol{S} \succ 0$ to denote that symmetric $\boldsymbol{S}$ is positive semi-definite and positive definite, respectively. The set of real numbers is denoted by $\mathbb{R}$. For a square matrix $\boldsymbol{A}$, $|\boldsymbol{A}|$ and $\mathrm{etr}(\boldsymbol{A})$ denote the determinant and the exponential of the trace of $\boldsymbol{A}$, respectively, i.e., $\mathrm{etr}(\boldsymbol{A}) = \exp(\mathrm{tr}(\boldsymbol{A}))$, $[\boldsymbol{B}_k]_{i:l,j:m}$ denotes the submatrix of the matrix $\boldsymbol{B}_k$ comprising its rows i through l and columns j through m, $[\boldsymbol{B}_k]_{ij} = [\boldsymbol{B}_k]_{i,j}$ is its (i,j)-th element, $B_{ij} = B_{i,j}$ is the (i,j)-th element of $\boldsymbol{B}$, x_i is the i-th component of $\boldsymbol{x} \in \mathbb{R}^p$, and $\boldsymbol{I}$ is the identity matrix. The superscript $\top$ denotes the transpose operation, and $\mathbb{E}$ denotes the expectation operation.

For a set V, $|V|$ denotes its cardinality, i.e., the number of elements in V. The set of integers 1 through n is denoted by $[1, n]$. Given $V = [1, n]$, let $\boldsymbol{x}_V \in \mathbb{R}^n$ denote a column vector with its elements indexed by the elements of V. Then for any $A \subseteq V$, we define $\boldsymbol{x}_A = [x_i]_{i \in A}$ = a column vector of dimension $|A|$, consisting of variables associated with the elements in A. The set $\mathbb{R}_+^{p \times p}$ denotes the set of $p \times p$ positive definite symmetric matrices, i.e., $\mathbb{R}_+^{p \times p} = \{\boldsymbol{S} \in \mathbb{R}^{p \times p} \,\big|\, \boldsymbol{S}^\top = \boldsymbol{S} \succ 0\}$. The notation $y = \mathcal{O}(g(x))$ means that there exists some finite real number $b > 0$ such that $\lim_{x \to \infty} |y/g(x)| \leq b$.

The notation $\boldsymbol{x} \sim \mathcal{N}_r(\boldsymbol{m}, \boldsymbol{\Sigma})$ denotes a real random vector $\boldsymbol{x}$ that is Gaussian with mean $\boldsymbol{m}$ and covariance $\boldsymbol{\Sigma}$. For scalar x, $x \sim \mathcal{B}(\alpha, \beta)$ denotes a beta random variable with probability density function (pdf)

$$f_x(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha - 1}(1 - x)^{\beta - 1},\ 0 < x < 1, \quad (1)$$

where $\Gamma(\alpha)$ denotes the (complete) Gamma function $\Gamma(z) := \int_0^\infty t^{z-1} e^{-t}\, dt$, and we do not distinguish between a random variable and the value taken by it. The abbreviations w.p.1., w.r.t. and i.i.d. stand for with probability one, with respect to, and independent and identically distributed, respectively.

II. System Model

We first define a multi-attribute Gaussian graphical model with p nodes and a random vector $\boldsymbol{z}_i \in \mathbb{R}^m$ associated with node i of the graph, in terms of an associated single-attribute Gaussian graphical model with mp nodes. A binary hypothesis testing problem is formulated in Section II-B exploiting the concept of "correct graph" [33], [34], and using the associated single-attribute Gaussian graphical model, where we test for exclusion/inclusion of m^2 edges in the single-attribute Gaussian graphical model corresponding to a single edge in the associated multi-attribute Gaussian graphical model.

Authorized licensed use limited to: Jitendra Tugnait. Downloaded on October 08,2020 at 21:51:17 UTC from IEEE Xplore. Restrictions apply.

A. Multi- and Single-Attribute Graphical Models

The conditional dependency structure among p random variables $x_1, x_1, \ldots, x_p$, components of $\boldsymbol{x} \in \mathbb{R}^p$, is represented using an undirected graph $\mathcal{G} = (V, \mathcal{E})$, where $V = [1, p]$ is the set of p nodes corresponding to the p random variables x_is, and $\mathcal{E} \subseteq V \times V$ is the set of undirected edges describing conditional dependencies among x_is. There is no edge between nodes i and j iff x_i and x_j are conditionally independent given the remaining p-2 variables x_ℓ, $\ell \in [1, p]$, $\ell \neq i$, $\ell \neq j$ [2, p. 60]. We will call $\mathcal{G}$ a single-attribute graphical model for $\boldsymbol{x}$. A real Gaussian graphical model associated with a simple undirected graph $\mathcal{G} = (V, \mathcal{E})$ is defined as the family of p-dimensional real-valued Gaussian random vectors $\boldsymbol{x} \in \mathbb{R}^p$, $p = |V|$, that obey the conditional independence restrictions implied by the edge set $\mathcal{E}$.

Now consider p jointly Gaussian random vectors $\boldsymbol{z}_i \in \mathbb{R}^m$, $i = 1, 2, \ldots, p$. We associate $\boldsymbol{z}_i$ with the ith node of an undirected graph $\mathcal{G} = (V, \mathcal{E})$ where $V = [1, p]$ is the set of p nodes, and $\mathcal{E} \subseteq V \times V$ is the set of edges that describe the conditional dependencies among vectors $\{\boldsymbol{z}_i,\ i \in V\}$. As in the scalar case ($m = 1$), there is no edge between node i and node j in $\mathcal{G}$ (i.e., $\{i, j\} \notin \mathcal{E}$) iff random vectors $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$ are conditionally independent given the remaining p-2 vectors $\boldsymbol{z}_\ell, \ell \in [1, p], \ell \neq i$, $\ell \neq j$ [10], [11]. This is the multi-attribute Gaussian graphical model of interest in this paper. The term multi-attribute Gaussian graphical model has been used in [9] for such models. Thus, the multi-attribute Gaussian graphical model associated with graph $\mathcal{G} = (V, \mathcal{E})$, $p = |V|$, is the family of p m-dimensional jointly Gaussian random vectors $\boldsymbol{z}_i \in \mathbb{R}^m$, that obey the conditional independence restrictions implied by the edge set $\mathcal{E}$.

Define the mp-vector

$$\boldsymbol{x} = [\boldsymbol{z}_1^\top\ \boldsymbol{z}_2^\top\ \cdots\ \boldsymbol{z}_p^\top]^\top \in \mathbb{R}^{mp}. \tag{2}$$

Suppose we have available N i.i.d. observations $\boldsymbol{x}(t)$, $t = 0, 1, \ldots, N-1$, of $\boldsymbol{x}$. Our main objective is to determine if $\{i, j\} \in \mathcal{E}$, given data $\{\boldsymbol{x}(t)\}_{t=0}^{N-1}$.

In order to exploit some results in [1] pertaining to single-attribute Gaussian graphical models, we will associate $\boldsymbol{x}$ with an "enlarged" graph $\bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}})$, where $\bar{V} = [1, mp]$ and $\bar{\mathcal{E}} \subseteq \bar{V} \times \bar{V}$. Now $[\boldsymbol{z}_j]_\ell$, the ℓth component of $\boldsymbol{z}_j$ associated with node j of $\mathcal{G} = (V, \mathcal{E})$, is the random variable $x_q = [\boldsymbol{x}]_q$, where $q = (j-1)m + \ell$, $j = 1, 2, \ldots, p$ and $\ell = 1, 2, \ldots, m$. The random variable x_q is associated with node q of $\bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}})$. Corresponding to the edge $\{j, k\} \in \mathcal{E}$ in the multi-attribute $\mathcal{G} = (V, \mathcal{E})$, there are m^2 edges $\{q, r\} \in \bar{\mathcal{E}}$ specified by $q = (j-1)m + s$ and $r = (k-1)m + t$, where $s = 1, 2, \ldots, m$ and $t = 1, 2, \ldots, m$. The graph $\bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}})$ is a single-attribute graph. In order for $\bar{\mathcal{G}}$ to reflect the conditional independencies encoded in $\mathcal{G}$, we must have the equivalence

$$\{j, k\} \notin \mathcal{E} \Leftrightarrow \bar{\mathcal{E}}^{(jk)} \cap \bar{\mathcal{E}} = \emptyset \tag{3}$$

where

$$\bar{\mathcal{E}}^{(jk)} = \{\{q, \} \,|\, q = (j-1)m + s,\ r = (k-1)m + t,\ s, t = 1, 2, \ldots, m\}. \tag{4}$$

Let $\boldsymbol{R}_{xx} = \mathbb{E}\{\boldsymbol{x}\boldsymbol{x}^\top\} \succ \boldsymbol{0}$ and $\bar{\boldsymbol{\Omega}} = \boldsymbol{R}_{xx}^{-1}$. Define the (j, k)th $m \times m$ subblock $\bar{\boldsymbol{\Omega}}^{(jk)}$ of $\bar{\boldsymbol{\Omega}}$ as

$$[\bar{\boldsymbol{\Omega}}^{(jk)}]_{st} = [\bar{\boldsymbol{\Omega}}]_{(j-1)m+s,(k-1)m+t},\ s, t = 1, 2, \ldots, m. \tag{5}$$

It is established in [11, Sec. 2.1] that

$$\bar{\boldsymbol{\Omega}}^{(jk)} = \boldsymbol{0} \Leftrightarrow \boldsymbol{z}_j \text{ and } \boldsymbol{z}_k \text{ are conditionally independent} \Leftrightarrow \{j, k\} \notin \mathcal{E}. \tag{6}$$

Since $\bar{\boldsymbol{\Omega}}^{(jk)} = \boldsymbol{0}$ is equivalent to $[\bar{\boldsymbol{\Omega}}]_{qr} = 0$ for every $\{q, r\} \in \bar{\mathcal{E}}^{(jk)}$, and since, by [1, Proposition 5.2], $[\bar{\boldsymbol{\Omega}}]_{qr} = 0$ iff x_q and x_r are conditionally independent, hence, iff $\{q, r\} \notin \bar{\mathcal{E}}$, it follows that equivalence (3) holds true. We state this result as Lemma 1.

Lemma 1: Consider the multi-attribute Gaussian graphical model $\mathcal{G} = (V, \mathcal{E})$, $|V| = p$, for p m-dimensional jointly Gaussian random vectors $\boldsymbol{z}_i \in \mathbb{R}^m$, and the associated single-attribute Gaussian graphical model $\bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}})$, $|V| = mp$ for $\boldsymbol{x} \in \mathbb{R}^{mp}$ as defined in (2). Assume that $\boldsymbol{R}_{xx} = \mathbb{E}\{\boldsymbol{x}\boldsymbol{x}^\top\} \succ \boldsymbol{0}$. Then $\forall j, k \in V$, $j \neq k$, $\{j, k\} \notin \mathcal{E}$, i.e., $\boldsymbol{z}_j$ and $\boldsymbol{z}_k$ are conditionally independent given $\boldsymbol{z}_{V\setminus\{j,k\}}$, iff $\bar{\mathcal{E}}^{(jk)} \cap \bar{\mathcal{E}} = \emptyset$ which is equivalent to $[\bar{\boldsymbol{\Omega}}]_{qr} = 0\ \forall \{q, r\} \in \bar{\mathcal{E}}^{(jk)}$, where $\bar{\boldsymbol{\Omega}} = \boldsymbol{R}_{xx}^{-1}$ and $\bar{\mathcal{E}}^{(jk)}$ is defined in (4). •

B. Binary Hypothesis Testing

We assume that the random vectors $\boldsymbol{z}_i$s are zero-mean. We are given N i.i.d. observations $\boldsymbol{x}(0), \boldsymbol{x}(1), \ldots, \boldsymbol{x}(N-1)$ of zero-mean $\boldsymbol{x} \in \mathbb{R}^{mp}$ defined in (2). Assume that $\boldsymbol{x}$ obeys the conditional independence restriction implied by the associated multi-attribute Gaussian graphical model $\mathcal{G} = (V, \mathcal{E})$, $V = [1, p]$. We need to decide if for any given $j, k \in V$, $j \neq k$, edge $\{j, k\} \in \mathcal{E}$ or $\{j, k\} \notin \mathcal{E}$. As in Lemma 1, consider the associated single-attribute Gaussian graphical model $\bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}})$. Then by (3), we need to test if $\bar{\mathcal{E}}^{(jk)} \cap \bar{\mathcal{E}} = \emptyset$. In a binary hypothesis testing framework, we compare two competing models: $\bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}})$ and $\bar{\mathcal{G}}' = (\bar{V}, \bar{\mathcal{E}}')$ with $\bar{\mathcal{E}}' = \bar{\mathcal{E}} \backslash \bar{\mathcal{E}}^{(jk)}$. What should $\bar{\mathcal{E}}$ be in the single-attribute Gaussian graphical model, i.e., what should $\mathcal{E}$ be in the original multi-attribute Gaussian graphical model? Recall that we do not know the true edge set $\mathcal{E}$. Based on the concept of a *correct* graph from [33], [34] (discussed in Appendix B-B), we pick $\bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}})$ to be the saturated (complete) graph $\bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}}_s)$, where $\bar{\mathcal{E}}_s$ denotes the set of all possible edges with vertices in $\bar{V}$.

Therefore, we consider the following hypothesis testing problem for testing the exclusion of single edge $\{j, k\}$ from $\mathcal{E}$:

$$\begin{aligned} &\mathcal{H}_0 :\ \bar{\boldsymbol{\Omega}} = \bar{\boldsymbol{\Omega}}(\bar{\mathcal{G}}) \in \bar{\mathbb{R}}_+(\bar{\mathcal{G}}'),\ \bar{\mathcal{G}}' = (\bar{V}, \bar{\mathcal{E}}'),\ \bar{\mathcal{E}}' = \bar{\mathcal{E}}_s \backslash \bar{\mathcal{E}}^{(jk)} \\ &\mathcal{H}_1 :\ \bar{\boldsymbol{\Omega}} = \bar{\boldsymbol{\Omega}}(\bar{\mathcal{G}}) \in \bar{\mathbb{R}}_+(\bar{\mathcal{G}}),\ \bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}}_s),\ \bar{\mathcal{E}}_s \text{ is saturated}, \end{aligned} \tag{7}$$

where

$$\bar{\mathbb{R}}_+(\bar{\mathcal{G}}) = \Big\{\bar{\boldsymbol{\Omega}} \in \mathbb{R}_+^{(mp)\times(mp)} \,\Big|\, \forall j \neq k \in V :\ \bar{\mathcal{E}}^{(jk)} \cap \bar{\mathcal{E}}_s = \emptyset \Rightarrow \bar{\Omega}_{qr} = 0\ \forall \{q, r\} \in \bar{\mathcal{E}}^{(jk)}\Big\} \tag{8}$$

and since the associated graph determines if $\bar{\Omega}_{qr}$ is zero or nonzero, we will denote $\bar{\boldsymbol{\Omega}}$ as $\bar{\boldsymbol{\Omega}}(\bar{\mathcal{G}})$ to explicitly indicate this dependence.

Authorized licensed use limited to: Jitendra Tugnait. Downloaded on October 08,2020 at 21:51:17 UTC from IEEE Xplore. Restrictions apply.

It follows from Lemma 1 that if $\mathcal{H}_0$ in (7) is true then $\{j,k\} \notin \mathcal{E}$, else $\{j,k\} \in \mathcal{E}$. In the next section we derive a GLRT for testing problem (7). The GLRT for testing a nested, restricted model against a "full" (unconstrained, saturated) model is called a *deviance test* [2, Sec. 6.8], and this term applies to problem (7).

III. Deviance Tests for Multi-Attribute Graphs

Here a GLRT is derived in Section III-B for the hypothesis testing problem (7), based on [1, Proposition 5.14] discussed in Section III-A. The null distribution of the test statistic is specified explicitly in Section III-C, which allows for analytical calculation of the test threshold.

Our objective is to derive the GLRT for the binary hypothesis testing problem specified in (7), given N i.i.d. observations $\boldsymbol{x}(0), \boldsymbol{x}(1), \ldots, \boldsymbol{x}(N-1)$ of zero-mean $\boldsymbol{x} \in \mathbb{R}^{mp}$ which obeys the conditional independence restriction implied by the associated graph $\bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}})$, $\bar{V} = [1, mp]$. Given $\boldsymbol{x}(t)$, let

$$\boldsymbol{X} = [\boldsymbol{x}(0)\ \boldsymbol{x}(1)\ \cdots\ \boldsymbol{x}(N-1)]^\top \in \mathbb{R}^{N\times(mp)}\,. \tag{9}$$

The joint probability density function (pdf) of $\boldsymbol{X}$ is

$$\begin{aligned} f_{\boldsymbol{X}}(\boldsymbol{X}) &= \prod_{t=0}^{N-1} \frac{1}{(2\pi)^{mp/2}\,|\boldsymbol{R}_{xx}|^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{x}^\top(t)\boldsymbol{R}_{xx}^{-1}\boldsymbol{x}(t)\right) \\ &= \frac{|\bar{\boldsymbol{\Omega}}|^{N/2}}{(2\pi)^{Nmp/2}} \mathrm{etr}(-\frac{1}{2}\bar{\boldsymbol{\Omega}}\boldsymbol{X}^\top\boldsymbol{X})\,. \end{aligned} \tag{10}$$

The unknown in (10) is $\bar{\boldsymbol{\Omega}}$ which satisfies the respective restrictions in (7) under the two hypotheses. The GLRT for problem (7) is given by

$$\mathcal{L}(\boldsymbol{X}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau \tag{11}$$

where

$$\mathcal{L}(\boldsymbol{X}) = \frac{\sup_{\bar{\boldsymbol{\Omega}}(\bar{\mathcal{G}})\in\bar{\mathbb{R}}_+(\bar{\mathcal{G}})}\ f_{\boldsymbol{X}|\mathcal{H}_1}(\boldsymbol{X}|\mathcal{H}_1)}{\sup_{\bar{\boldsymbol{\Omega}}(\bar{\mathcal{G}})\in\bar{\mathbb{R}}_+(\bar{\mathcal{G}}')}\ f_{\boldsymbol{X}|\mathcal{H}_0}(\boldsymbol{X}|\mathcal{H}_0)}\,. \tag{12}$$

The solution to (11)–(12) will be obtained by exploiting [1, Proposition 5.14], which we recall next.

A. GLRT for Decomposable Gaussian Graphical Models [1, Sec. 5.3.3]

Consider $\mathbf{y} \in \mathbb{R}^p$ and the associated single-attribute undirected graph $\mathcal{G} = (V, \mathcal{E})$ which is assumed to be decomposable. Let $\mathcal{G}_0 = (V, \mathcal{E}_0)$ denote a decomposable submodel of $\mathcal{G}$ with k (≥ 1) edges less than $\mathcal{G}$. By [1, Lemma 2.21], there is a sequence $\mathcal{G}_0 \subset \cdots \subset \mathcal{G}_k = \mathcal{G}$ of graphs that are decomposable and differ by one edge only. Let e_i denote the edge that is in $\mathcal{G}_i$ but not in $\mathcal{G}_{i-1}$. Consider the hypothesis testing problem

$$\begin{aligned} &\mathcal{H}_0 \,:\, \boldsymbol{\Omega} = \boldsymbol{\Omega}(\mathcal{G}) \in \mathbb{R}_+(\mathcal{G}_0),\ \mathcal{G}_0 = (V, \mathcal{E}_0) \\ &\mathcal{H}_1 \,:\, \boldsymbol{\Omega} = \boldsymbol{\Omega}(\mathcal{G}) \in \mathbb{R}_+(\mathcal{G}),\ \mathcal{G} = (V, \mathcal{E})\,, \end{aligned} \tag{13}$$

given N i.i.d. observations $\boldsymbol{y}(t)$, $t = 0, 1, \ldots, N-1$, of zero-mean $\boldsymbol{y}$, where

$$\mathbb{R}_+(\mathcal{G}) = \left\{\boldsymbol{\Omega} \in \mathbb{R}_+^{p\times p} \,\middle|\, \forall i \neq j \in V \,:\, \{i,j\} \notin \mathcal{E} \right. \\ \left. \Rightarrow\ \Omega_{ij} = 0\right\}\,. \tag{14}$$

Define

$$\boldsymbol{Y} = [\boldsymbol{y}(0)\ \boldsymbol{y}(1)\ \cdots\ \boldsymbol{y}(N-1)]^\top \in \mathbb{R}^{N\times p}\,. \tag{15}$$

The GLRT for problem (13) is given by

$$\mathcal{L}(\boldsymbol{Y}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau \tag{16}$$

where

$$\mathcal{L}(\boldsymbol{Y}) = \frac{\sup_{\boldsymbol{\Omega}(\mathcal{G})\in\mathbb{R}_+(\mathcal{G})}\ f_{\boldsymbol{Y}|\mathcal{H}_1}(\boldsymbol{Y}|\mathcal{H}_1)}{\sup_{\boldsymbol{\Omega}(\mathcal{G})\in\mathbb{R}_+(\mathcal{G}_0)}\ f_{\boldsymbol{Y}|\mathcal{H}_0}(\boldsymbol{Y}|\mathcal{H}_0)}\,. \tag{17}$$

Under the above set-up, the solution (16)–(17) to problem (13) is specified in [1, Proposition 5.14], which we now state as Lemma 2 in the notation of this paper.

Lemma 2: The GLRT for the test (13) is given by (16) where

$$\mathcal{L}(\boldsymbol{Y}) = \left(\frac{|\hat{\boldsymbol{\Omega}}_0|}{|\hat{\boldsymbol{\Omega}}|}\right)^{-N/2}\,, \tag{18}$$

$\hat{\boldsymbol{\Omega}}_0$ maximizes $f_{\boldsymbol{Y}|\mathcal{H}_0}(\boldsymbol{Y}|\mathcal{H}_0)$ under the constraint $\boldsymbol{\Omega}(\mathcal{G}) \in \mathbb{R}_+(\mathcal{G}_0)$, and $\hat{\boldsymbol{\Omega}}$ maximizes $f_{\boldsymbol{Y}|\mathcal{H}_1}(\boldsymbol{Y}|\mathcal{H}_1)$ under the constraint $\boldsymbol{\Omega}(\mathcal{G}) \in \mathbb{R}_+(\mathcal{G})$. Under $\mathcal{H}_0$, $\mathcal{L}^{-2/N}(\boldsymbol{Y}) \sim Z = \prod_{i=1}^k Z_i$ where Z_is are mutually independent beta random variables $\mathcal{B}((N - |C_i^*| + 1)/2, 1/2)$ and C_i^* is the unique clique of $\mathcal{G}_i$ that contains the edge e_i. •

In [1, Proposition 5.14], both mean and inverse covariance are unknown, whereas in our case, we have zero mean $\boldsymbol{y}$. This results in an extra degree of freedom in estimating the covariance matrix, hence $Z_i \sim \mathcal{B}((N - |C_i^*| + 1)/2, 1/2)$ in Lemma 2, instead of $Z_i \sim \mathcal{B}((N - |C_i^*|)/2, 1/2)$ as in [1, Proposition 5.14].

B. Proposed GLRT/ Deviance Test

The solution to (11)–(12) will be obtained by exploiting Lemma 2. To this end, we need to establish that all the requirements of Lemma 2, as applied to (11)–(12), are fulfilled.

Consider the graphs $\bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}}_s)$ and $\bar{\mathcal{G}}' = (\bar{V}, \bar{\mathcal{E}}')$, $\bar{\mathcal{E}}' = \bar{\mathcal{E}}_s \backslash \bar{\mathcal{E}}^{(jk)}$, under $\mathcal{H}_1$ and $\mathcal{H}_0$, respectively. In the notation of Section III-A, $\mathcal{G}_0 = \bar{\mathcal{G}}'$, with $\mathcal{G}_0$ and $\mathcal{G}_{m^2} = \bar{\mathcal{G}}$ differing by m^2 edges. In order for Lemma 2 to be applicable, both $\mathcal{G}_0$ and $\mathcal{G}_{m^2}$ should be decomposable. Since $\mathcal{G}_{m^2} = \bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}}_s)$ is complete (saturated), it is decomposable (by Definition 3 in Appendix A). We need to show that $\mathcal{G}_0 = \bar{\mathcal{G}}'$ is also decomposable; to this end, we will use Lemma 6 (Appendix A).

For $j = 1, 2, \ldots, p$ and $s = 1, 2, \cdots m$, define the set

$$\begin{aligned} A_{j,s} = \{&(j-1)m + s, (j-1)m + s + 1, \ldots, \\ &(j-1)m + m\} \end{aligned} \tag{19}$$

where j indexes the graph nodes and s indexes the attributes at each node. We will use the notation $A_{j,s}$ extensively later in Lemma 3. For the given edge $\{j,k\}$, define the sets

$$B_1 = \bar{V} \backslash A_{j,1}\,, \quad B_2 = \bar{V} \backslash A_{k,1}\,. \tag{20}$$

Authorized licensed use limited to: Jitendra Tugnait. Downloaded on October 08,2020 at 21:51:17 UTC from IEEE Xplore. Restrictions apply.

With $n = 2$ in Definition 4 (Appendix A), we have

$$H_1 = B_1 = \bar{V}\backslash A_{j,1}$$
$$H_2 = B_1 \cup B_2 = \bar{V}$$
$$S_2 = H_1 \cap B_2 = \bar{V}\backslash\{A_{j,1} \cup A_{k,1}\}\,. \tag{21}$$

Since the separator set S_2 is complete (as it induces the complete subgraph $(S_2, \bar{\mathcal{E}}_{S_2})$, $\bar{\mathcal{E}}_{S_2} = (\bar{\mathcal{E}}_s\backslash\bar{\mathcal{E}}^{(jk)}) \cap (S_2 \times S_2)$), and $S_2 \subset B_1$, the sequence B_1, B_2 is perfect (Definition 4, Appenidx A). Moreover, the subsets B_1 and B_2 of $\bar{V}$ are the cliques of graph $\mathcal{G}_0$. Hence, by Lemma 6, $\mathcal{G}_0$ is decomposable. Therefore, Lemma 2 applies.

To calculate (12) similar to (18), we need the determinants of the ML estimates of $\bar{\boldsymbol{\Omega}}$ under the two hypotheses. Define the sample covariance matrix (we now use $\bar{\boldsymbol{\Sigma}} = \boldsymbol{R}_{xx}$)

$$\hat{\bar{\boldsymbol{\Sigma}}} = \frac{1}{N}\boldsymbol{X}^\top\boldsymbol{X} = \frac{1}{N}\sum_{t=0}^{N-1}\boldsymbol{x}(t)\boldsymbol{x}^\top(t)\,. \tag{22}$$

For $\boldsymbol{x}(t)$ associated with $\bar{\mathcal{G}}$ or $\bar{\mathcal{G}}'$, for $\tilde{V} \subseteq \bar{V}$, we define $\boldsymbol{x}_{\tilde{V}}(t) = [x_i(t)]_{i\in\tilde{V}}$ = a column vector of dimension $|\tilde{V}|$. Thus, $\boldsymbol{x}(t) = \boldsymbol{x}_{\bar{V}}(t)$. Under $\mathcal{H}_1$, the graph $\bar{\mathcal{G}}$ is saturated, hence by [1, Theorem 5.1], the ML estimate $\hat{\bar{\boldsymbol{\Omega}}}$ of $\bar{\boldsymbol{\Omega}} = \bar{\boldsymbol{\Sigma}}^{-1}$ is given by (noting that the mean of $\boldsymbol{x}(t)$ is zero)

$$\hat{\bar{\boldsymbol{\Omega}}} = \hat{\bar{\boldsymbol{\Sigma}}}^{-1} \Rightarrow |\hat{\bar{\boldsymbol{\Omega}}}| = 1/|\hat{\bar{\boldsymbol{\Sigma}}}|\,. \tag{23}$$

Under $\mathcal{H}_0$, the graph $\bar{\mathcal{G}}' = \mathcal{G}_0$ is decomposable with cliques B_1 and B_2 specified in (20), and separator S_2 specified in (21). By [1, Proposition 5.9], the determinant of the ML estimate $\hat{\bar{\boldsymbol{\Omega}}}_0$ of $\bar{\boldsymbol{\Omega}} = \bar{\boldsymbol{\Sigma}}^{-1}$ under $\mathcal{H}_0$ is determined by the cliques and separators of the graph $\bar{\mathcal{G}}' = (\bar{V}, \bar{\mathcal{E}}') = \mathcal{G}_0$. By [1, Proposition 5.9], we have

$$|\hat{\bar{\boldsymbol{\Omega}}}_0| = \frac{|\hat{\bar{\boldsymbol{\Sigma}}}_{\bar{V}\backslash\{A_{j,1}\cup A_{k,1}\}}|}{|\hat{\bar{\boldsymbol{\Sigma}}}_{\bar{V}\backslash A_{j,1}}||\hat{\bar{\boldsymbol{\Sigma}}}_{\bar{V}\backslash A_{k,1}}|} \tag{24}$$

where

$$\hat{\bar{\boldsymbol{\Sigma}}}_{\bar{V}\backslash\{A_{j,1}\cup A_{k,1}\}} = \frac{1}{N}\sum_{t=0}^{N-1}\boldsymbol{x}_{\bar{V}\backslash\{A_{j,1}\cup A_{k,1}\}}(t) \times \boldsymbol{x}^\top_{\bar{V}\backslash\{A_{j,1}\cup A_{k,1}\}}(t) \tag{25}$$

$$\hat{\bar{\boldsymbol{\Sigma}}}_{\bar{V}\backslash A_{j,1}} = \frac{1}{N}\sum_{t=0}^{N-1}\boldsymbol{x}_{\bar{V}\backslash A_{j,1}}(t)\boldsymbol{x}^\top_{\bar{V}\backslash A_{j,1}}(t) \tag{26}$$

$$\hat{\bar{\boldsymbol{\Sigma}}}_{\bar{V}\backslash A_{k,1}} = \frac{1}{N}\sum_{t=0}^{N-1}\boldsymbol{x}_{\bar{V}\backslash A_{k,1}}(t)\boldsymbol{x}^\top_{\bar{V}\backslash A_{k,1}}(t)\,. \tag{27}$$

Thus, $\mathcal{L}(\boldsymbol{X})$ in (12) is given by

$$\mathcal{L}(\boldsymbol{X}) = \left[\frac{|\hat{\bar{\boldsymbol{\Sigma}}}_{\bar{V}\backslash\{A_{j,1}\cup A_{k,1}\}}|\,|\hat{\bar{\boldsymbol{\Sigma}}}|}{|\hat{\bar{\boldsymbol{\Sigma}}}_{\bar{V}\backslash A_{j,1}}|\,|\hat{\bar{\boldsymbol{\Sigma}}}_{\bar{V}\backslash A_{k,1}}|}\right]^{-N/2}. \tag{28}$$

We now turn to specifying the distribution of $\mathcal{L}^{-2/N}(\boldsymbol{X})$ under $\mathcal{H}_0$. By [1, Lemma 2.21], there is a sequence $\bar{\mathcal{G}}' = \mathcal{G}_0 \subset \mathcal{G}_1 \subset \cdots \subset \mathcal{G}_{m^2} = \bar{\mathcal{G}}$ of graphs that are decomposable and differ by one edge only. In order to apply Lemma 2 to (11)–(12), we need to determine these graphs $\mathcal{G}_i$s, and edge e_i that is in $\mathcal{G}_i$ but not in $\mathcal{G}_{i-1}$. Starting with $\mathcal{G}_0 = \bar{\mathcal{G}}' = (\bar{V}, \bar{\mathcal{E}}')$, $\bar{\mathcal{E}}' = \bar{\mathcal{E}}_s\backslash\bar{\mathcal{E}}^{(jk)}$, we add edges

$$\{(j-1)m + s, (k-1)m + t\} \tag{29}$$

in the following order (to be read left-to-right and top-to-bottom)

$$\begin{aligned} &s = 1, \quad t = 1, 2, \ldots, m \\ &s = 2, \quad t = 1, 2, \ldots, m \\ &\qquad\vdots \quad\ \vdots \\ &s = m, \quad t = 1, 2, \ldots, m\,. \end{aligned} \tag{30}$$

That is, we first add edge $e_1 = \{(j-1)m + 1, (k-1)m + 1\}$ to $\mathcal{G}_0$ to construct $\mathcal{G}_1$, followed by adding edge $e_2 = \{(j-1)m + 1, (k-1)m + 2\}$ to $\mathcal{G}_1$ to construct $\mathcal{G}_2$, and so on.

Lemma 3 stated below and proved in Appendix C, is central to proving Theorem 1 specifying the exact distribution of $\mathcal{L}^{-2/N}(\boldsymbol{X})$ under $\mathcal{H}_0$.

Lemma 3: The edge $e_i = \{(j-1)m + s, (k-1)m + t\}$ is in the unique clique C_i^* of the graph $\mathcal{G}_i$ where $i = (s-1)m + t$, $s, t = 1, 2, \ldots, m$,

$$C_i^* = \begin{cases} \bar{V}\backslash(A_{j,s+1} \cup A_{k,t+1})\,, & s < m, t < m \\ \bar{V}\backslash A_{k,t+1}\,, & s = m, t < m \\ \bar{V}\backslash A_{j,s+1}\,, & s < m, t = m \\ \bar{V}\,, & s = m, t = m \end{cases} \tag{31}$$

and $|C_i^*| = (p-2)m + s + t$. •

Invoking (28) and Lemma 3, we now state the result of application of Lemma 2 to the hypothesis testing problem (7) as Theorem 1, where we index C_i^* as C_{st}^*.

Theorem 1: The GLRT for the test (7) is given by (11) where $\mathcal{L}(\boldsymbol{X})$ is specified by (28). Under $\mathcal{H}_0$, $\mathcal{L}^{-2/N}(\boldsymbol{X}) \sim Z = \prod_{s,t=1}^m Z_{st}$ where Z_{st}s are mutually independent beta random variables, with $Z_{st} \sim \mathcal{B}((N - |C_{st}^*| + 1)/2, 1/2)$ and $|C_{st}^*| = (p-2)m + s + t$. •

Remark 1: As noted earlier in Section I, hypothesis testing for edge connections in high-dimensional Gaussian graphical models has been addressed in [27], [30], [31]. The results are asymptotic in nature with both number of nodes p and sample size N becoming large. In case of hypothesis testing for edge connections for the entire graph, which requires multiple testing, the approaches of [27], [30], [31] yield asymptotic control of false discovery rates. While [27], [30] investigate multiple testing of individual entries of a precision matrix, [31] investigates multiple testing for conditional dependence between subgroups of nodes. The approach of [31] therefore applies to multi-attribute graphical models. In [31] one has a graphical model with p nodes, N samples and M non-overlapping subgroups of nodes whose conditional independence need to be tested. Translating this to our multi-attribute setting, [31] considers a model with mp nodes, N samples, and p subgroups of m nodes per subgroup. Multiple testing is based upon $m \times m$ submatrices of the precision matrix. Unlike [9]–[11], [13] where group lasso is used to estimate the precision matrix with group penalty on $m \times m$ submatrices of the precision matrix to reflect the

Authorized licensed use limited to: Jitendra Tugnait. Downloaded on October 08,2020 at 21:51:17 UTC from IEEE Xplore. Restrictions apply.

multi-attribute character, [31] uses single attribute penalty such as lasso. Under some regularity and technical conditions, [31] provides a method for multiple testing of all edges in a multi-attribute graph to control false discovery rates asymptotically as p and N become large. Since we deal with low-dimensional models, our problem is simpler, allowing us to derive exact probability distribution of the test statistic for testing of an edge connection for finite samples ($N \geq mp$), and then to use this result to control false discovery rates for finite samples, following the Benjamini-Hochberg approach [38]. Theorem 1 of [31] "corresponds" to our Theorem 1 where while the distribution of the test statistic in our Theorem 1 is exact for finite samples, that in [31] holds asymptotically. □

C. Threshold Selection

We now turn to calculating the pdf of a test statistic, equivalent to that in Theorem 1, under $\mathcal{H}_0$. We follow [35]. Our main result is in Theorem 2, which allows us to calculate the test threshold analytically. Theorem 2 is stated below, and proved in Appendix C.

In the following, χ^2_n denotes a random variable with central chi-square distribution with n degrees of freedom as well as the distribution itself. Also, $B_r(y)$ denotes the Bernoulli polynomial of degree r and order unity. The first five Bernoulli polynomials are ($B_0(y) = 1$):

$$B_1(y) = y - (1/2), \quad B_2(y) = y^2 - y + (1/6),$$
$$B_3(y) = y^3 - (3/2)y^2 + (1/2)y,$$
$$B_4(y) = y^4 - 2y^3 + y^2 - (1/30),$$
$$B_5(y) = y^5 - (5/2)y^4 + (5/3)y^3 - (1/6)y.$$

Theorem 2: The GLRT/deviance test for (7) can be expressed as

$$2\rho \, \ln(\mathcal{L}(\boldsymbol{X})) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau$$

where

$$\rho = 1 - \frac{m(p-1) + 0.5}{N}. \tag{32}$$

The threshold τ is picked to achieve a pre-specified $P_{fa} = 1 - P\{2\rho \, \ln(\mathcal{L}(\boldsymbol{X})) \leq \tau \,|\, \mathcal{H}_0\}$ where

$$\begin{aligned} P\{2\rho \, \ln(\mathcal{L}(\boldsymbol{X})) \leq z \,|\, \mathcal{H}_0\} &= P\{\chi^2_\nu \leq z\} \\ &+ \omega_2 \left[P\{\chi^2_{\nu+4} \leq z\} - P\{\chi^2_\nu \leq z\}\right] \\ &+ \omega_3 \left[P\{\chi^2_{\nu+6} \leq z\} - P\{\chi^2_\nu \leq z\}\right] \\ &+ \Bigg\{\omega_4 \left[P\{\chi^2_{\nu+8} \leq z\} - P\{\chi^2_\nu \leq z\}\right] + \frac{1}{2}\omega_2^2 \left[P\{\chi^2_{\nu+8} \leq z\}\right. \\ &\left. -2P\{\chi^2_{\nu+4} \leq z\} + P\{\chi^2_\nu \leq z\}\right] \Bigg\} + \mathcal{O}\left(\frac{m^2}{N^5}\right), \end{aligned} \tag{33}$$

$$\nu = m^2, \tag{34}$$

$$\omega_r = \frac{(-1)^{r+1}}{r(r+1)(\rho N/2)^r} \left\{ \left(\sum_{s,t=1}^{m} B_{r+1}(a_{st} + 0.5) \right) - \left(\sum_{s,t=1}^{m} B_{r+1}(a_{st} + 1) \right) \right\}. \tag{35}$$

and $a_{st} = (1-\rho)\frac{N}{2} - \frac{m(p-2)+s+t}{2}$. •

Remark 2: By Theorem 1, $\mathcal{L}^{-2/N}(\boldsymbol{X}) \in [0, 1]$ as it is a product of beta random variables, hence, $0 \leq 2\rho \, \ln(\mathcal{L}(\boldsymbol{X})) < \infty$. Therefore, $P\{2\rho \, \ln(\mathcal{L}(\boldsymbol{X})) \leq 0 \,|\, \mathcal{H}_0\} = 0$ and $P\{2\rho \, \ln(\mathcal{L}(\boldsymbol{X})) \leq z_u \,|\, \mathcal{H}_0\} \approx 1$ for some "large" z_u. For given values of parameters p, m and N, the values of ρ, ν and a_{st} are fixed, and so are the values of ω_r, $r = 2, 3, 4$. Then (33) is a function of only z. Using software to numerically calculate the CDF (cumulative distribution function) of a central χ^2 random variable (e.g., MATLAB function cdf('chi2',.,.)), one can use the bisection method to numerically solve $g(\tau) := P\{2\rho \, \ln(\mathcal{L}(\boldsymbol{X})) \leq \tau \,|\, \mathcal{H}_0\} - (1 - P_{fa}) = 0$ for τ starting with endpoints 0 and z_u. We have $g(0) < 0$ and $g(z_u) > 0$, therefore, the bisection method will find the unique solution for τ in the interval $[0, z_u]$ since $g(\tau)$ is monotone (as $P\{2\rho \, \ln(\mathcal{L}(\boldsymbol{X})) \leq \tau \,|\, \mathcal{H}_0\}$ is a CDF) and continuous in τ. □

IV. Deviance Test Simplification

The statistic $\mathcal{L}(\boldsymbol{X})$ of Theorem 1 can be simplified to yield a computationally fast statistic, just as [36, Theorem 1] simplifies [32, Theorem 7.6]. Here we generalize the approach followed in [12, Sec. V]. Assume that $N \geq mp$, so that $\hat{\bar{\boldsymbol{\Sigma}}}^{-1}$ exists w.p.1.

Define the expectations

$$\bar{\boldsymbol{\Sigma}}_{\bar{V}} = \mathbb{E}\{\boldsymbol{x}_{\bar{V}}(t)\boldsymbol{x}^\top_{\bar{V}}(t)\} = \bar{\boldsymbol{\Sigma}} = \bar{\boldsymbol{\Omega}}^{-1} \tag{36}$$

$$\bar{\boldsymbol{\Sigma}}_{\bar{V}\setminus\{A_{j,1}\cup A_{k,1}\}} = \mathbb{E}\{\boldsymbol{x}_{\bar{V}\setminus\{A_{j,1}\cup A_{k,1}\}}(t) \times \boldsymbol{x}^\top_{\bar{V}\setminus\{A_{j,1}\cup A_{k,1}\}}(t)\} \tag{37}$$

$$\bar{\boldsymbol{\Sigma}}_{\bar{V}\setminus A_{j,1}} = \mathbb{E}\{\boldsymbol{x}_{\bar{V}\setminus A_{j,1}}(t)\boldsymbol{x}^\top_{\bar{V}\setminus A_{j,1}}(t)\} \tag{38}$$

$$\bar{\boldsymbol{\Sigma}}_{\bar{V}\setminus A_{k,1}} = \mathbb{E}\{\boldsymbol{x}_{\bar{V}\setminus A_{k,1}}(t)\boldsymbol{x}^\top_{\bar{V}\setminus A_{k,1}}(t)\}. \tag{39}$$

In general, for subsets A and B of $\bar{V}$, we use the notation

$$\bar{\boldsymbol{\Sigma}}_{A,B} = \mathbb{E}\{\boldsymbol{x}_A(t)\boldsymbol{x}^\top_B(t)\}, \; \bar{\boldsymbol{\Sigma}}_A = \mathbb{E}\{\boldsymbol{x}_A(t)\boldsymbol{x}^\top_A(t)\}. \tag{40}$$

Lemma 4 is proved in Appendix C.

Lemma 4: For graph $\bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}})$ and associated $\boldsymbol{y} \in \mathbb{R}^{2p}$ considered in Lemma 2, we have

$$\frac{|\bar{\boldsymbol{\Sigma}}_{\bar{V}\setminus\{A_{j,1}\cup A_{k,1}\}}| \, |\bar{\boldsymbol{\Sigma}}_{\bar{V}}|}{|\bar{\boldsymbol{\Sigma}}_{\bar{V}\setminus A_{j,1}}| \, |\bar{\boldsymbol{\Sigma}}_{\bar{V}\setminus A_{k,1}}|} = \frac{|\bar{\boldsymbol{\Omega}}^{-1}_{(jj)}|}{|[\bar{\boldsymbol{\Omega}}^{-1}_{(jkjk)}]_{1:m,1:m}|} \tag{41}$$

where, with $\bar{\Omega}_{qr}$ denoting the (q, r)-th entry of $\bar{\boldsymbol{\Omega}}$, $\bar{\boldsymbol{\Omega}}_{(jk)}$ is $m \times m$, $\bar{\boldsymbol{\Omega}}_{(jkjk)}$ is $(2m) \times (2m)$,

$$\bar{\boldsymbol{\Omega}}_{(jkjk)} := \begin{bmatrix} \bar{\boldsymbol{\Omega}}_{(jj)} & \bar{\boldsymbol{\Omega}}_{(jk)} \\ \bar{\boldsymbol{\Omega}}_{(kj)} & \bar{\boldsymbol{\Omega}}_{(kk)} \end{bmatrix}, \tag{42}$$

Authorized licensed use limited to: Jitendra Tugnait. Downloaded on October 08,2020 at 21:51:17 UTC from IEEE Xplore. Restrictions apply.

the (q,r)-th element of $\bar{\mathbf{\Omega}}_{(jk)}$ is

$$[\bar{\mathbf{\Omega}}_{(jk)}]_{qr} := \bar{\Omega}_{(j-1)m+q,(k-1)m+r}\,, \quad q,r = 1,2,\ldots,m, \tag{43}$$

and $[\bar{\mathbf{\Omega}}^{-1}_{(jkjk)}]_{1:m,1:m}$ is the $m \times m$ submatrix of $\bar{\mathbf{\Omega}}^{-1}_{(jkjk)}$ comprising its first m rows and first m columns. •

Assume that $N \geq mp$, so that $\hat{\bar{\mathbf{\Sigma}}}^{-1}$ exists w.p.1. Lemma 5 then follows just as Lemma 4, by manipulations of $\hat{\bar{\mathbf{\Sigma}}}^{-1}$ and submatrices of $\hat{\bar{\mathbf{\Sigma}}} = \hat{\bar{\mathbf{\Sigma}}}_{\bar{V}}$.

Lemma 5: An alternative expression for the test statistic of Theorem 1 is given by

$$\frac{|\hat{\bar{\mathbf{\Sigma}}}_{\bar{V}\setminus\{A_{j,1}\cup A_{k,1}\}}|\,|\hat{\bar{\mathbf{\Sigma}}}|}{|\hat{\bar{\mathbf{\Sigma}}}_{\bar{V}\setminus A_{j,1}}|\,|\hat{\bar{\mathbf{\Sigma}}}_{\bar{V}\setminus A_{k,1}}|} = \frac{|\hat{\bar{\mathbf{\Omega}}}^{-1}_{(jj)}|}{|[\hat{\bar{\mathbf{\Omega}}}^{-1}_{(jkjk)}]_{1:m,1:m}|} \tag{44}$$

where, with $\hat{\bar{\Omega}}_{qr}$ denoting the (q,r)-th entry of $\hat{\bar{\mathbf{\Omega}}} = \hat{\bar{\mathbf{\Sigma}}}^{-1}$, $\hat{\bar{\mathbf{\Omega}}}_{(jk)}$ is $m \times m$, $\hat{\bar{\mathbf{\Omega}}}_{(jkjk)}$ is $(2m) \times (2m)$,

$$\hat{\bar{\mathbf{\Omega}}}_{(jkjk)} := \begin{bmatrix} \hat{\bar{\Omega}}_{(jj)} & \hat{\bar{\Omega}}_{(jk)} \\ \hat{\bar{\Omega}}_{(kj)} & \hat{\bar{\Omega}}_{(kk)} \end{bmatrix}, \tag{45}$$

the (q,r)-th element of $\hat{\bar{\mathbf{\Omega}}}_{(jk)}$ is

$$[\hat{\bar{\mathbf{\Omega}}}_{(jk)}]_{qr} := \hat{\bar{\Omega}}_{(j-1)m+q,(k-1)m+r}\,, \quad q,r = 1,2,\ldots,m, \tag{46}$$

and $[\hat{\bar{\mathbf{\Omega}}}^{-1}_{(jkjk)}]_{1:m,1:m}$ is the $m \times m$ submatrix of $\hat{\bar{\mathbf{\Omega}}}^{-1}_{(jkjk)}$ in its top left corner. •

Using log-likelihood, and Theorems 1 and 2, we have the following deviance test

$$2\rho\,\ln(\mathcal{L}(\boldsymbol{X})) = \rho N\left[\ln\left(|[\hat{\bar{\mathbf{\Omega}}}^{-1}_{(jkjk)}]_{1:m,1:m}|\right) - \ln\left(|\hat{\bar{\mathbf{\Omega}}}^{-1}_{(jj)}|\right)\right] \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau \tag{47}$$

where τ is picked to achieve a specified probability of false alarm (significance level) P_{fa}. We summarize the above results in Theorem 3.

Theorem 3: The GLRT for the test (7) is given by (47), where $\hat{\bar{\mathbf{\Omega}}}_{(jkjk)}$ and $\hat{\bar{\mathbf{\Omega}}}_{(jj)}$ are given by (45) and (46), respectively. Under $\mathcal{H}_0$, the test statistic $N\rho\,\ln(\mathcal{L}(\boldsymbol{X}))$ has the distribution specified in Theorem 2. •

A. Flop Count Comparison Between Test Statistics of Theorems 1 and 3

We will compare computation of the left-side of (44) with that of its right-side. Suppose that we need to carry out edge exclusion tests for all $p(p-1)/2$ edges in $\mathcal{E}$. The determinant $|\hat{\bar{\mathbf{\Sigma}}}|$ is computed just once and needs $\mathcal{O}((mp)^3)$ flops, whereas, while $|\hat{\bar{\mathbf{\Sigma}}}_{\bar{V}\setminus A_{j,1}}|$ and $|\hat{\bar{\mathbf{\Sigma}}}_{\bar{V}\setminus A_{k,1}}|$ both need $\mathcal{O}((mp)^3)$ flops, they have to be computed anew for each of the $m^2p(p-1)/2$ distinct edges of $\bar{\mathcal{E}}$, leading to a total of $\mathcal{O}((mp)^5)$ flops. Thus, the left-side of (44) needs $\mathcal{O}((mp)^5)$ flops. To compute the right-side of (44), we first compute $\hat{\bar{\mathbf{\Omega}}} = \hat{\bar{\mathbf{\Sigma}}}^{-1}$ just once with $\mathcal{O}((mp)^3)$ flops. Given $\hat{\bar{\mathbf{\Omega}}}$, we compute $|\hat{\bar{\mathbf{\Omega}}}^{-1}_{(jj)}|$ and $|[\hat{\bar{\mathbf{\Omega}}}^{-1}_{(jkjk)}]_{1:m,1:m}|$ with $\mathcal{O}(m^3)$ flops for each of the $m^2p(p-1)/2$ edges in $\bar{\mathcal{E}}$. Thus, the right-side of (44) needs $\mathcal{O}((mp)^3 + m^5p^2) = \mathcal{O}((mp)^3)$ flops, for $m \ll p$. Hence, the computational complexity of our proposed alternative statistic is $\mathcal{O}((mp)^3)$ versus $\mathcal{O}((mp)^5)$ for the original expression of Theorem 1.

V. Numerical Experiments

We now present numerical results based on synthetic as well as real data to illustrate the proposed test. All synthetic data results are based on 1000 Monte Carlo runs.

A. Synthetic Data

Given realizations of p random vectors $\boldsymbol{z}_i \in \mathbb{R}^m$, there are $p(p-1)/2$ unordered pair of vertices in the associated CIG that may or may not be connected. So we have to perform $p(p-1)/2$ binary hypothesis tests. Thus, we have a multiple testing problem where the main issue is how to control the overall significance level. Instead, as in [12], we will use the ROC (receiver operating characteristic) averaged over all edges, as a performance measure. The ROC curve is the trade-off between the false-alarm rate P_{fa} (type I error) and the detection probability P_d.

We consider a chain graph (an example in [11]) where p nodes are connected in succession. In the upper triangular $\bar{\mathbf{\Omega}}$, using the notation of (5), we set $[\bar{\mathbf{\Omega}}^{(jk)}]_{st} = 0.5^{|s-t|}$ for $j = k = 1,\ldots,p$, $s,t = 1,\ldots,m$. For $j \neq k$, if the two nodes are not connected, we have $\bar{\mathbf{\Omega}}^{(jk)} = \mathbf{0}$, and if nodes j and k are connected in the chain graph, then $[\bar{\mathbf{\Omega}}^{(jk)}]_{st}$ is uniformly distributed over $[-0.4,-0.1] \cup [0.1,0.4]$ if $s \neq t$, and $[\bar{\mathbf{\Omega}}^{(jk)}]_{st} = 0$ if $s = t$. Now add $\rho\boldsymbol{I}$ to $\bar{\mathbf{\Omega}}$ with ρ picked to make minimum eigenvalue of $\bar{\mathbf{\Omega}} + \rho\boldsymbol{I}$ equal to 0.5. This is similar to simulation example 3 in [11, Sec. 5.1]. With $\mathbf{\Phi}\mathbf{\Phi}^\top = (\bar{\mathbf{\Omega}} + \rho\boldsymbol{I})^{-1}$, we generate $\boldsymbol{x} = \mathbf{\Phi}\boldsymbol{w}$ with $\boldsymbol{w} \in \mathbb{R}^{mp}$ as Gaussian $\boldsymbol{w} \sim \mathcal{N}_r(\mathbf{0},\boldsymbol{I})$. We generate N i.i.d. observations from $\boldsymbol{x}$, with $m = 3$, $p = 6$ or 10, and $N = 100$ or 200.

We apply the GLRT (47) coupled with Theorem 2, to compute the test threshold for a given significance level α, to test each of $p(p-1)/2$ edges. The simulation results are shown in Figs. 1 and 2 based on 1000 runs and averaged over all edges. It is seen from Fig. 1 that empirical α tracks the design α quite well. The ROC curves (Fig. 2) show that the performance improves with increasing sample size N and decreasing number of nodes p, for a given value of P_{fa}.

1) Applying Existing Methods: For a given edge in the multi-attribute graphical model, we test for joint exclusion of m^2 edges in an augmented single-attribute graphical model. The single edge exclusion test (53) of [1, Sec. 5.3.3] can be used to test these m^2 (=9) edges, but only in a multiple testing set-up where one tests each edge separately, one-by-one. For zero-mean $\boldsymbol{x}$, under the null hypothesis, $\mathcal{L}_{SAG}$ in (53) is distributed as $\mathcal{B}((N-mp+1)/2, 1/2)$. We applied multiple testing using (53) to synthetic data, based on Bonferroni method ([24] and [37, Sec. 10.7]) of threshold selection. In this approach, to achieve significance level α as in our proposed approach, we pick significance level $\alpha' = \alpha/9$ for each of nine applications of the

Authorized licensed use limited to: Jitendra Tugnait. Downloaded on October 08,2020 at 21:51:17 UTC from IEEE Xplore. Restrictions apply.

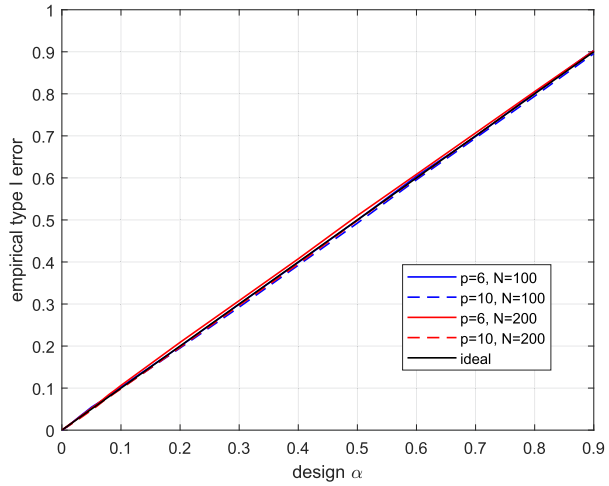


Fig. 1. Empirical significance level vs. design significance level: $m = 3$, $p = 6$ or 10, N =100 or 200.

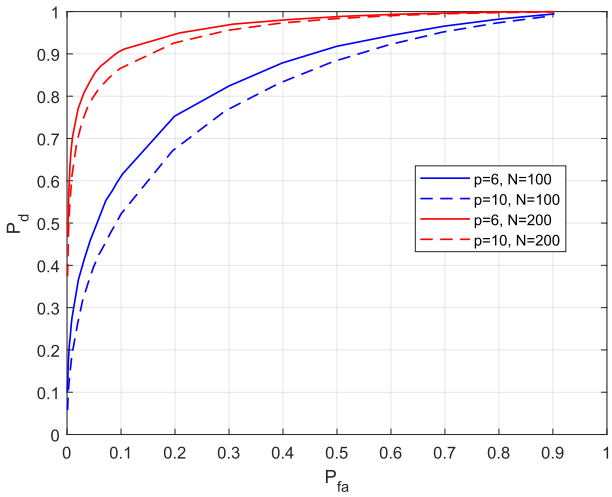


Fig. 2. ROC curves: $m = 3$, $p = 6$ or 10, N =100 or 200.

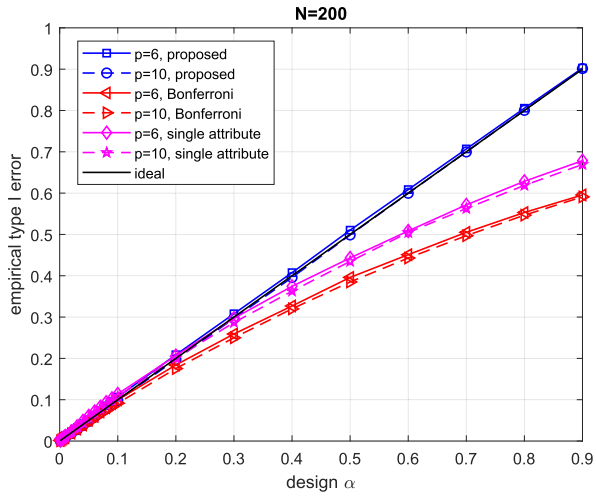


Fig. 3. Empirical significance level vs. design significance level: $m = 3$, $p = 6$ or 10, N =200. Comparison between the proposed, Bonferroni and single-attribute model-based methods.

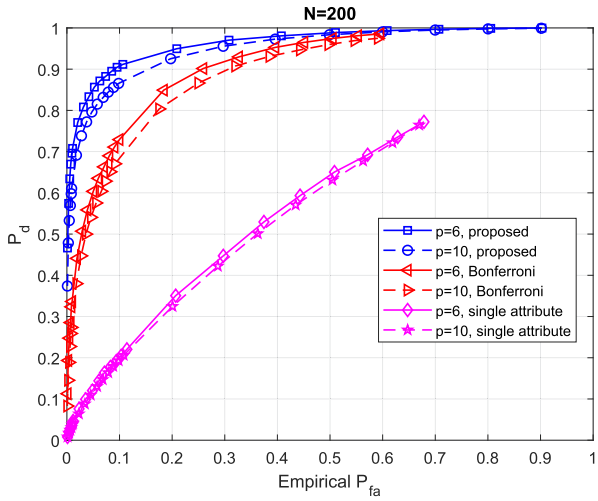


Fig. 4. ROC curves: $m = 3$, $p = 6$ or 10, $N = 200$. Comparison between the proposed, Bonferroni and single-attribute model-based methods.

test (53). The overall significance level then has an upper bound $9\alpha' = \alpha$. Alternatively, one can estimate the graph based on single attributes only, one graph per attribute, and combine these graphs in some way. In this example, we choose to place an edge between two nodes if at least one of the graphs has an edge. To achieve significance level α as in our proposed approach, we pick significance level $\alpha'' = \alpha/3$ for each of the three applications of the test (53) to estimate single attribute graphs. The simulation results based on 1000 Monte Carlo runs, averaged over all edges, are shown in Figs. 3 and 4 for $N = 200$, where the label "Bonferroni" refers to multiple testing over 9 edges, and the label "single attribute" refers to estimating one graph per attribute. It is seen from Fig. 3 that only the proposed approach delivers the design significance level, whereas, while the other two methods keep the empirical α below the upper bound of design value, as expected, there is a wide gap between the design and empirical significance levels. In Fig. 4 we show the ROC curves, P_d versus empirical P_{fa}. It is seen that performances of test (53), with Bonferroni method based threshold selection or single attribute based threshold selection, are significantly worse than that of the proposed test (47).

2) Non-Gaussian Distributions: We now investigate the influence of non-Gaussian distributions on graph estimation using the proposed deviance test. Our theory is based on the assumption of Gaussian data. How does the estimated graph change when this assumption does not hold? We generate data via $\boldsymbol{x} = \boldsymbol{\Phi}\boldsymbol{w}$ where $\boldsymbol{\Phi}$ is exactly as for Figs. 1–4 corresponding to a chain graph, but $\boldsymbol{w} \in \mathbb{R}^{mp}$ now is not necessarily Gaussian. We generate N i.i.d. observations from $\boldsymbol{x}$, with $m = 3$, $p = 10$ and $N = 200$ using $\boldsymbol{w}$ with zero-mean i.i.d. components distributed as (a) Gaussian (as before), (b) Laplace, (c) uniform, or (d) exponential random variables. The resulting performance based on 1000 Monte Carlo runs is shown in Fig. 5 where we show ROC curves as well as empirical significance level vs. design significance level curves. We see little change in performance when the underlying distributions are changed.

Authorized licensed use limited to: Jitendra Tugnait. Downloaded on October 08,2020 at 21:51:17 UTC from IEEE Xplore. Restrictions apply.

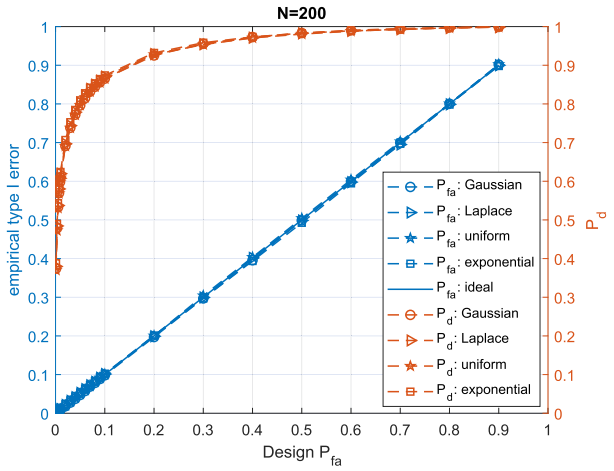


Fig. 5. Performance under different distributions: $m = 3, p = 10, N = 200$.

TABLE I
NUMBER OF ESTIMATED EDGES IN SINGLE ATTRIBUTE (SA) AND MULTI-ATTRIBUTE (MA) IMAGE GRAPHS. POSSIBLE EDGES 2016, $N = 1250$, $p = 64$, $m = 3$, FDR ≤ 0.01.

Approach	Image 1	Image 2
MA	203	216
SA: Red	156	155
SA: Green	154	145
SA: Blue	141	131
Common to MA & Red SA	148	140
Common to MA & Green SA	147	138
Common to MA & Blue SA	138	127
Common to Red & Green SAs	147	143
Common to Green & Blue SAs	134	124
Common to Red & Blue SAs	135	126

Notice that the test statistic (47) is based only on sample covariance which is a sufficient statistic in our case because we assumed Gaussian distribution to begin with. Then the central limit theorem ensures that the sample covariance will be Gaussian for large enough sample size, regardless of the distribution of $\boldsymbol{w}$.

B. Real Data Example: Graphs of Color Texture Images

Following [15], [16] where grayscale texture images from a University of Southern California (USC-SIPI) database are considered, we now consider color textures from the Amsterdam Library of Textures (ALOT) [40]. We use two versions of the image 108, images c111.png and c111r60.png (400×400 patches shown in Figs. 6(a) and 6(b)), photographed from different angles. These two images are 3072×1536 color pixels (RGB) from which we extract 400×400 patches used for inferring image graphs. The 400×400 patches were partitioned into non-overlapping 8×8 blocks, vectorized into 64-pixel columns, 3 colors associated with each pixel. Thus, we have $m = 3$, $p = 64$ and initially 2500 samples. We randomly sample without replacement $N = 1250$ samples from the initial 2500 samples. The data were centered and mean-square value normalized to one before processing. Histograms and autocorrelation functions (with $\pm 2\sigma$ error bounds) of red, green and blue components of a typical pixel in the 8×8 blocks of images 1 and 2 are shown in Figs. 6(c) and 6(d), respectively. (For the autocorrelation function we used function autocorr.m of MATLAB Econometrics Toolbox, which also provides confidence bounds at each lag.) It is seen that the data sequences for each color are approximately decorrelated. The histograms show that probability distribution is not too far from Gaussian, except for the red component of image 1. The simulation results of Section V-A2 suggest that the estimated graphs do not change much with changes in distributions.

The results of multi-attribute graph estimation are shown in Figs. 6(e) and 6(f), based on test designed for a false discovery rate (FDR) of $\alpha = 0.01$. Some statistics about the detected edges are in Table I. We used [38] (a modification of [39] to handle dependent multiple tests, see also [37, Sec. 10.7]) to control FDR where each edge is tested using the proposed test (47) and the resulting p-values are used to control FDR. We also show the results of single-attribute graph estimation in Figs. 6(e) and 6(f) and Table I, based on the test designed for a false discovery rate (FDR) of $\alpha = 0.01$, using the single edge exclusion test (53) of [1, Sec. 5.3.3] ($\mathcal{L}_{SAG}$ in (53)).

The image graphs of Figs. 6(e) and 6(f) are arranged as 8×8 nodes mimicking the 8×8 blocks of original images that were vectorized into 64-pixel columns, and the edges (links between nodes) are shown with colored edge weights where link thickness also reflects edge weight. The following rationale was used to select the edge weights. The single attribute test (53) for edge $\{j, k\}$ can be rewritten as

$$\frac{[\hat{\boldsymbol{\Omega}}]_{jk}^2}{[\hat{\boldsymbol{\Omega}}]_{jj}[\hat{\boldsymbol{\Omega}}]_{kk}} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \bar{\tau} \tag{48}$$

with higher $[\hat{\boldsymbol{\Omega}}]_{jk}^2/([\hat{\boldsymbol{\Omega}}]_{jj}[\hat{\boldsymbol{\Omega}}]_{kk})$ signifying higher probability of an edge between nodes j and k. We take the statistic $[\hat{\boldsymbol{\Omega}}]_{jk}^2/([\hat{\boldsymbol{\Omega}}]_{jj}[\hat{\boldsymbol{\Omega}}]_{kk})$ as a measure of edge weight. The multi-attribute test (47) can be rewritten as

$$1 - \frac{|\hat{\hat{\boldsymbol{\Omega}}}_{(jj)}^{-1}|}{|[\hat{\hat{\boldsymbol{\Omega}}}_{(jkjk)}^{-1}]_{1:m,1:m}|} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tilde{\tau} \tag{49}$$

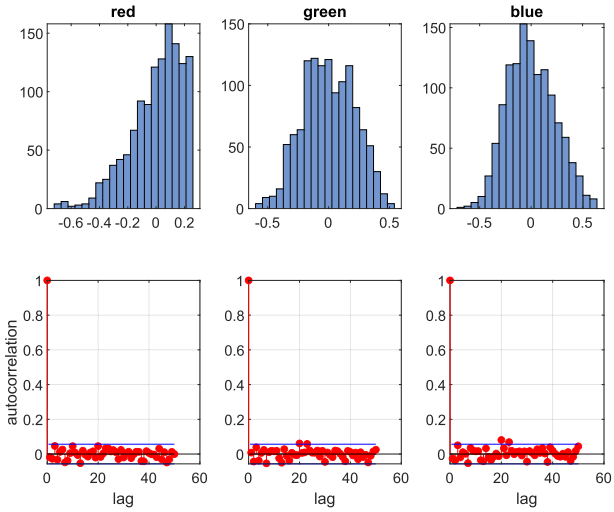
(see also Lemma 5), whose test statistic (left-side) reduces to the statistic $[\hat{\boldsymbol{\Omega}}]_{jk}^2/([\hat{\boldsymbol{\Omega}}]_{jj}[\hat{\boldsymbol{\Omega}}]_{kk})$ when $m = 1$. In conformity with the single attribute case, we take the statistic $1 - |\hat{\hat{\boldsymbol{\Omega}}}_{(jj)}^{-1}|/|[\hat{\hat{\boldsymbol{\Omega}}}_{(jkjk)}^{-1}]_{1:m,1:m}|$ as a measure of the edge weight in the multi-attribute case. For all image graphs shown in Figs. 6(e) and 6(f), we first normalize the maximum edge weight to one (for each graph separately). The colorbar for the colored normalized edge weights is shown in Figs. 6(e) and 6(f), and we quantize the interval [0, 1] to 4 link thicknesses with thicker lines denoting higher statistic value.

Compare Figs. 6(a) and 6(e), and Figs. 6(b) and 6(f), respectively, to note that the strong link weights follow the texture

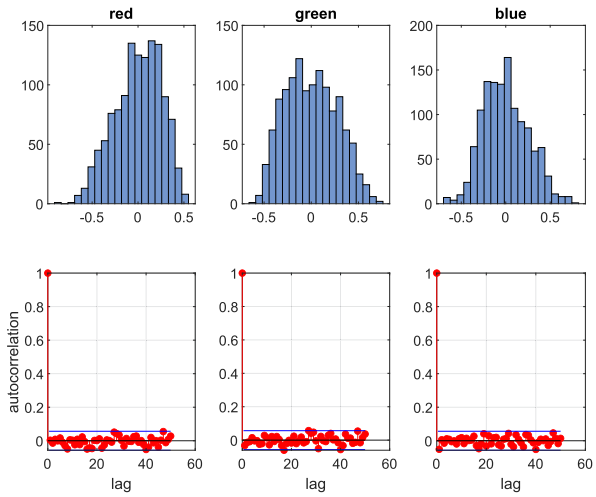
Authorized licensed use limited to: Jitendra Tugnait. Downloaded on October 08,2020 at 21:51:17 UTC from IEEE Xplore. Restrictions apply.

(a) Color texture image 1.

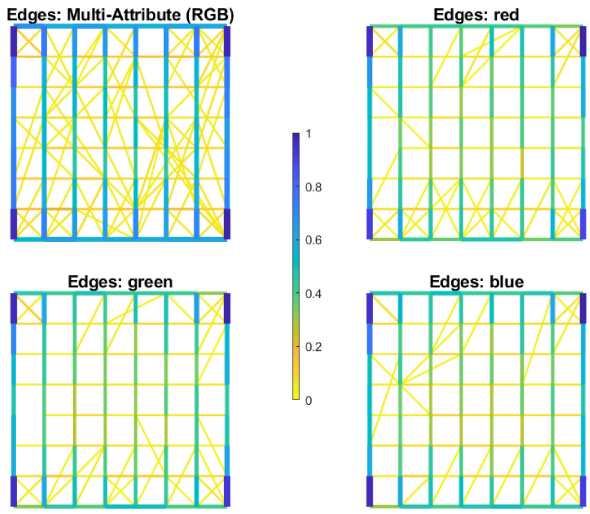
(b) Color texture image 2.



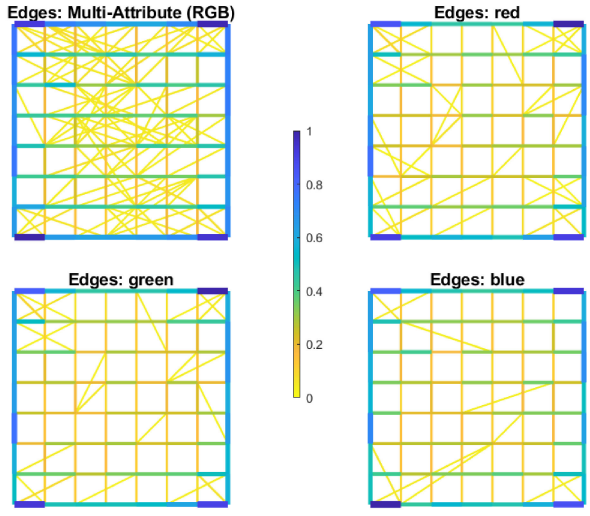
(c) Histograms and autocorrelation functions (with $\pm 2\sigma$ error bounds) of red, green and blue components of a typical pixel in 8×8 block of image 1.



(d) Histograms and autocorrelation functions (with $\pm 2\sigma$ error bounds) of red, green and blue components of a typical pixel in 8×8 block of image 2.



(e) Estimated multi-attribute and single-attribute red, green and blue image graphs for image 1.



(f) Estimated multi-attribute and single-attribute red, green and blue image graphs for image 2.

Fig. 6. Color image graph estimation results based on 400×400 RGB images: $p = 64$, $m = 3$, $N = 1250$.

Authorized licensed use limited to: Jitendra Tugnait. Downloaded on October 08,2020 at 21:51:17 UTC from IEEE Xplore. Restrictions apply.

orientation: vertical in Figs. 6(a) and 6(e), and horizontal in Figs. 6(b) and 6(f). Weaker edge weights connect pixels in "other" directions. These observations provide "visual" support for fitted graphs, confirming our intuition. Note that we do not know the groundtruth. Comparing multi-attribute graphs with single attribute graphs in Figs. 6(e) and 6(f) as well as via statistics in Table I, we see that multi-attribute captures edges not found in single attribute graphs while rejecting some edges in single attribute graphs that do not find support across all three single attribute graphs.

VI. Conclusion

We investigated the problem of inferring the conditional independence graph (CIG) of Gaussian vectors from multi-attribute data. Most existing methods for graph estimation are based on single-attribute models where each node of the graph represents a scalar random variable. In multi-attribute graphical models, each node represents a random vector. We proposed and analyzed a deviance test based on generalized likelihood ratio, for edge exclusion in multi-attribute Gaussian graphical models. The null distribution of the test statistic was derived to allow analytical calculation of the test threshold. The deviance test statistic was also expressed in an alternative form, where for m-dimensional random vectors per node in a p-node graph, the alternative statistic reduces the computational complexity from $\mathcal{O}((mp)^5)$ to $\mathcal{O}((mp)^3)$. Numerical results based on synthetic as well as real data were presented to illustrate the proposed tests.

Appendix A
Simple Undirected Graphs

Here we recall some useful definitions and background material from [1, Chapter 2], [32, Chapter 5] concerning simple undirected graphs, that we use extensively in this paper. The material in this appendix also occurs in [12, Sec. II.A].

A simple undirected graph $\mathcal{G}$ is a pair $(V, \mathcal{E})$, where V is a finite set of elements called vertices (or nodes), and the set $\mathcal{E} \subseteq V \times V$ of unordered pairs (called edges) of distinct vertices. The graph is simple in that there are no loops or multiple edges. We denote the edge between vertices α and β of V as $\{\alpha, \beta\}$. The set $V \backslash A$ denotes the set of all elements in V that are not in A.

A graph is said to be complete, or saturated, if $\mathcal{E} = V \times V$, i.e., if any given vertex is connected to every other vertex. A subset $A \subseteq V$ for $\mathcal{G} = (V, \mathcal{E})$ induces the subgraph $\mathcal{G}_A = (A, \mathcal{E}_A)$ where $\mathcal{E}_A = \mathcal{E} \cap (A \times A)$, i.e., $\mathcal{E}_A$ contains exactly those edges in $\mathcal{E}$ which connect vertices from A. A subset is complete if it induces a complete subgraph.

A path of length n in $\mathcal{G}$ is a sequence of distinct vertices $\alpha_0, \alpha_1, \ldots, \alpha_n$, where $\alpha_j \in V$, such that $\{\alpha_{j-1}, \alpha_j\} \in \mathcal{E}$ for all $j = 1, 2, \ldots, n$. Two subsets $A, B \subseteq V$ are said to be separated by $S \subseteq V$ if all paths from A to B go via S, i.e., the paths from a vertex in A to a vertex in B intersect S at some vertex.

Definition 1. Decomposition: A pair (A, B) of subsets of V of a simple undirected graph $\mathcal{G} = (V, \mathcal{E})$ is said to form a decomposition of $\mathcal{G}$ if $V = A \cup B$, $A \cap B$ is a complete subset of V, and $A \backslash B$ and $B \backslash A$ are separated by $A \cap B$. •

Definition 2. Clique: Let $\mathcal{G} = (V, \mathcal{E})$ be a simple undirected graph. A complete subset of V which is maximal w.r.t. set inclusion is called a clique, i.e.,

$$(C \text{ is complete and } C \subset C' \Rightarrow C' \text{ is not complete}) \Leftrightarrow C \text{ is a clique},$$

where $C, C' \subseteq V$. •

Definition 3. Decomposability: A simple undirected graph $\mathcal{G} = (V, \mathcal{E})$ is said to be decomposable if it is complete, or if there exists a decomposition formed by proper subsets A and B of V into decomposable subgraphs $\mathcal{G}_A$ and $\mathcal{G}_B$. •

Definition 4 is from [1, pp. 14-15], as specialized to this paper.

Definition 4. Perfect sequence and running intersection property: Let $B_1, \ldots, B_n$ be a sequence of n subsets of V of a simple undirected graph $\mathcal{G} = (V, \mathcal{E})$. Let

$$H_j = B_1 \cup \cdots \cup B_j\,,\ j \geq 1$$
$$S_j = H_{j-1} \cap B_j\,,\ j \geq 2\,.$$

The sequence $B_1, \ldots, B_n$ is said to be perfect if the following conditions are fulfilled:

i) for all $i > 1$ there is a $j < i$ such that $S_i \subseteq B_j$;
ii) the sets S_i are complete for all i.

The condition (i) is known as the running intersection property (RIP). •

We conclude this section with Lemma 6, which is [1, Proposition 2.17] and is also in [32, Chapter 5].

Lemma 6: A simple undirected graph $\mathcal{G} = (V, \mathcal{E})$ is decomposable if and only if the cliques of $\mathcal{G}$ can be ordered (or, numbered) to form a perfect sequence. •

Appendix B
Gaussian Graphical Models

Here we recall some existing results on edge exclusion tests for selection of single-attribute Gaussian graphical models. The material in this appendix also occurs in [12, Sec. III.C].

If $\boldsymbol{x} \sim \mathcal{N}_r(\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} \succ 0$ and $\boldsymbol{x} \in \mathbb{R}^p$, the pdf of $\boldsymbol{x}$ is

$$f_{\boldsymbol{x}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{p/2}\,|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{x}\right) = \frac{|\boldsymbol{\Omega}|^{1/2}}{(2\pi)^{p/2}} \exp\left(-\frac{1}{2}\boldsymbol{x}^\top \boldsymbol{\Omega}\boldsymbol{x}\right). \quad (50)$$

Consider $\boldsymbol{x}$ associated with a single-attribute $\mathcal{G} = (V, \mathcal{E})$, $|V| = p$, where $\boldsymbol{x}$ obeys the conditional independence restrictions implied by the edge set $\mathcal{E}$. Denote the inverse covariance matrix as $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$. Then Ω_{ij}, the (i, j)-th element of $\boldsymbol{\Omega}$, is zero iff x_i and x_j are conditionally independent [1, Proposition 5.2].

A. Edge Exclusion Tests

Suppose we are given N i.i.d. observations $\boldsymbol{x}(0), \boldsymbol{x}(1), \ldots, \boldsymbol{x}(N-1)$ of $\boldsymbol{x}$. In graphical model selection (graph estimation), one needs to decide if a given edge $\{j, k\}$, for $1 \leq j < k \leq p$, in the associated graph $\mathcal{G} = (V, \mathcal{E})$, is in $\mathcal{E}$ or not in $\mathcal{E}$. Since $\Omega_{jk} = 0 \Leftrightarrow \{j, k\} \notin \mathcal{E}$, its estimate based on the random vector sample can be used to test if $\{j, k\} \notin \mathcal{E}$;

Authorized licensed use limited to: Jitendra Tugnait. Downloaded on October 08,2020 at 21:51:17 UTC from IEEE Xplore. Restrictions apply.

hence the term edge exclusion test [24]. There are $p(p-1)/2$ distinct edges in an undirected graph with p nodes. So for graphical model selection, one has to determine which set of edges out of total $p(p-1)/2$ edges belong to $\mathcal{E}$. A formulation in the form a generalized likelihood ratio test (GLRT) is in [1, Sec. 5.3.3], which we discuss next.

Consider two undirected graphs $\mathcal{G} = (V, \mathcal{E})$ and $\mathcal{G}' = (V, \mathcal{E}')$ where $|V| = p$, and $\mathcal{G}'$ is a subgraph of $\mathcal{G}$, i.e., $\mathcal{E}'$ is a subset of $\mathcal{E}$. Let $\mathcal{E}$ be saturated, i.e., complete. Now suppose that we remove one edge $\{j,k\}$ from $\mathcal{E}$ to obtain $\mathcal{E}' = \mathcal{E}\backslash\{j,k\}$. Since the associated graph determines if Ω_{jk} is zero or nonzero, we will denote $\mathbf{\Omega}$ as $\mathbf{\Omega}(\mathcal{G})$ to explicitly indicate this dependence. In [1, Sec. 5.3.3], the following hypothesis testing problem is considered

$$\begin{aligned} &\mathcal{H}_0 \,:\, \mathbf{\Omega} = \mathbf{\Omega}(\mathcal{G}) \in \mathbb{R}_+(\mathcal{G}'),\ \mathcal{G}' = (V, \mathcal{E}'),\ \mathcal{E}' = \mathcal{E}\backslash\{j,k\} \\ &\mathcal{H}_1 \,:\, \mathbf{\Omega} = \mathbf{\Omega}(\mathcal{G}) \in \mathbb{R}_+(\mathcal{G}),\ \mathcal{G} = (V, \mathcal{E}),\ \mathcal{E} \text{ is saturated}\,, \end{aligned} \tag{51}$$

where $\mathbb{R}_+(\mathcal{G})$ is defined in (14). The sample covariance matrix for the given data is $\hat{\mathbf{\Sigma}} = \frac{1}{N}\sum_{t=0}^{N-1} \boldsymbol{x}(t)\boldsymbol{x}^\top(t)$, and an estimate of its inverse is $\hat{\mathbf{\Omega}} = \hat{\mathbf{\Sigma}}^{-1}$. The GLRT statistic in [1, Sec. 5.3.3] for problem (51) is (SAG stands for single-attribute Gaussian)

$$\mathcal{L}_{SAG} = 1 - \frac{[\hat{\mathbf{\Omega}}]_{jk}^2}{[\hat{\mathbf{\Omega}}]_{jj}[\hat{\mathbf{\Omega}}]_{kk}}\,, \tag{52}$$

where the null is rejected if (52) is small:

$$\mathcal{L}_{SAG} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \tau. \tag{53}$$

The GLRT for testing a nested, restricted model against a "full" (unconstrained, saturated) model has been called a *deviance test* [1, Secs. 5.2.2, 5.3.3], [2, Sec. 6.8]. The deviance of the restricted model is defined as twice the difference between the unconstrained maximum of the log-likelihood and the maximum taken over the restricted model [2, p. 186]. The GLRT (53) is therefore also a deviance test.

B. Correct Graph

Why was $\mathcal{E}$ picked to be saturated be in (51)? To justify this choice, we first recall the concept of a *correct* graph from [33], [34], where it is defined for time series graphical models, but it applies here as well.

Definition 5. Correct Graph [33, Definition 2], [34, Definition 1]: Let $\mathcal{G} = (V, \mathcal{E})$, $V = [1, p]$, denote the true graphical model for $\boldsymbol{x}$. Then $\tilde{\mathcal{G}} = (V, \tilde{\mathcal{E}})$ is correct for $\mathcal{G} = (V, \mathcal{E})$ if $\mathcal{E} \subseteq \tilde{\mathcal{E}}$. •

Thus, if $\{j,k\} \in \mathcal{E}$, then we must have $\{j,k\} \in \tilde{\mathcal{E}}$, and if $\{j,k\} \notin \tilde{\mathcal{E}}$, then $\{j,k\} \notin \mathcal{E}$. Note that a saturated (complete) graph $\mathcal{G} = (V, \mathcal{E}_s)$ is a correct graph for any graphical model, where $\mathcal{E}_s$ denotes the set of all possible edges with vertices in V.

This definition of a correct graph is central to the approach of [34], patterned after that of [33], to graphical modeling of real time series, and it applies in our case as well. To test if edge $\{j,k\} \notin \mathcal{E}$, we consider if edge $\{j,k\} \notin \mathcal{E}_s$ (null: correct model), or $\{j,k\} \in \mathcal{E}_s$ (alternative: incorrect model), where $(V, \mathcal{E}_s)$ is correct for any $(V, \mathcal{E})$. There are $L = p(p-2)/2$ edges in $(V, \mathcal{E}_s)$. Number edge $\{j,k\}$ as $i = (j-1)p + k$, $1 \le j < k \le p$. Let $(V, \mathcal{E}^i)$ denote the graph where $\mathcal{E}^i = \mathcal{E}_s\backslash\{j,k\}$, with $i = (j-1)p + k$, i.e., only one edge is missing from the saturated graph. Lemma 7 is a restatement of [34, Proposition 2].

Lemma 7: If the graph $(V, \mathcal{E}^i)$ is correct for edges corresponding to $i = i_1, i_2, \ldots, i_s$, and incorrect for all others, then the graphical model $\mathcal{G} = (V, \mathcal{E})$ for $\boldsymbol{x}$ is the graph with only edges $\{i_1, i_2, \ldots, i_s\}$ missing. •

Appendix C
Proofs of Lemmas 3 and 4, and Theorem 2

A. Proof of Lemma 3

Consider the edge $e_i = \{(j-1)m + s, (k-1)m + t\}$ being added to $\mathcal{G}_{i-1}$ in the order specified in (29)–(30). We will alternatively index C_i^* as C_{st}^* also. We consider following four cases:

Case 1: Assume $s < m$ and $t < m$. Define

$$\begin{aligned} &B_1 = \bar{V}\backslash A_{j,s},\ B_2 = \bar{V}\backslash(A_{j,s+1} \cup A_{k,t+1}), \\ &B_3 = \bar{V}\backslash A_{k,1}\,. \end{aligned}$$

With $n = 3$ in Definition 4 (Appendix A), we have

$$\begin{aligned} H_1 &= B_1,\ H_2 = B_1 \cup B_2 = \bar{V}\backslash A_{j,s+1} \\ H_3 &= B_1 \cup B_2 \cup B_3 = \bar{V} \\ S_2 &= H_1 \cap B_2 = \bar{V}\backslash(A_{j,s} \cup A_{k,t+1}) \\ S_3 &= H_2 \cap B_3 = \bar{V}\backslash(A_{j,s+1} \cup A_{k,1})\,. \end{aligned} \tag{54}$$

The separator sets S_2 and S_3 are complete, $S_2 \subset B_1$, and $S_3 \subset B_1$, hence the sequence of subsets B_1, B_2, B_3 of $\bar{V}$ is a perfect sequence. The sets B_1, B_2 and B_3 are all possible cliques of $\bar{\mathcal{G}}_i$, which is, therefore, decomposable by Lemma 6. Note that $B_2 = C_i^* = C_{st}^*$ where $|C_{st}^*| = (p-2)m + s + t$.

Case 2: Assume $s = m$ and $t < m$. Define

$$B_1 = \bar{V}\backslash A_{k,t+1},\ B_2 = \bar{V}\backslash\{(j-1)m + m\}\,.$$

With $m = 2$ in Definition 4, we have

$$\begin{aligned} &H_1 = B_1 = \bar{V}\backslash A_{k,t+1}, \quad H_2 = B_1 \cup B_2 = \bar{V} \\ &S_2 = H_1 \cap B_2 = \bar{V}\backslash(A_{k,t+1} \cup \{(j-1)m + m\})\,. \end{aligned} \tag{55}$$

The sequence of subsets B_1, B_2 of $\bar{V}$ is perfect since S_2 is complete and $S_2 \subset B_1$. The sets B_1 and B_2 are the cliques of $\bar{\mathcal{G}}_1$, which is decomposable by Lemma 6. We have $B_1 = C_i^* = C_{st}^*$ where $|C_{mt}^*| = (p-2)m + m + t$.

Case 3: Assume $s < m$ and $t = m$. Define

$$B_1 = \bar{V}\backslash A_{j,s+1},\ B_2 = \bar{V}\backslash\{(k-1)m + m\}\,.$$

With $m = 2$ in Definition 4, we have

$$\begin{aligned} &H_1 = B_1 = \bar{V}\backslash A_{j,s+1}, \quad H_2 = B_1 \cup B_2 = \bar{V} \\ &S_2 = H_1 \cap B_2 = \bar{V}\backslash(A_{j,s+1} \cup \{(k-1)m + m\})\,. \end{aligned} \tag{56}$$

The sequence of subsets B_1, B_2 of $\bar{V}$ is perfect since S_2 is complete and $S_2 \subset B_1$. The sets B_1 and B_2 are the cliques of $\bar{\mathcal{G}}_1$, which is decomposable by Lemma 6. We have $B_1 = C_i^* = C_{sm}^*$ where $|C_{sm}^*| = (p-2)m + s + m$.

Authorized licensed use limited to: Jitendra Tugnait. Downloaded on October 08,2020 at 21:51:17 UTC from IEEE Xplore. Restrictions apply.

Case 4: Assume $s = m$ and $t = m$. Now we have a complete, hence decomposable, graph $\bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}}_s)$, with a single clique $\bar{V}$ which contains the edge $\{jm, km\}$. Thus, $\bar{V} = C_i^* = C_{mm}^*$ where $|C_{mm}^*| = (p-2)m + m + m = pm$.

This completes the proof. ■

B. Proof of Theorem 2

First we need to recall/develop some auxiliary results. If $Y \sim \mathcal{B}(\alpha, \beta)$, then [1, p. 262]

$$\mathbb{E}\{Y^h\} = \frac{\Gamma(\alpha+h)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha+\beta+h)}. \tag{57}$$

Lemma 8. Under $\mathcal{H}_0$, for any h with $\mathrm{Re}(h) \geq 0$,

$$\mathbb{E}\left\{\frac{1}{\mathcal{L}^h(\boldsymbol{X})} \,|\, \mathcal{H}_0\right\} = \prod_{s,t=1}^{m} \frac{\Gamma(b_{st}+1)\Gamma(b_{st}+0.5+(Nh/2))}{\Gamma(b_{st}+0.5)\Gamma(b_{st}+1+(Nh/2))} \tag{58}$$

where $b_{st} = \frac{N-m(p-2)-s-t}{2}$. •

Proof: By Theorem 1,

$$\mathbb{E}\left\{\frac{1}{\mathcal{L}^h(\boldsymbol{X})} \,|\, \mathcal{H}_0\right\} = \mathbb{E}\left\{\prod_{s,t=1}^{m} Z_{st}^{Nh/2}\right\} = \prod_{s,t=1}^{m} \mathbb{E}\left\{Z_{st}^{Nh/2}\right\} \tag{59}$$

where we have used the fact that Z_{st}s are mutually independent random variables. Since $Z_{st} \sim \mathcal{B}((N - |C_{st}^*| + 1)/2, 1/2)$ and $|C_{st}^*| = (p-2)m + s + t$, we have $(N - |C_{st}^*|)/2 = b_{st}$, so that $Z_{st} \sim \mathcal{B}(b_{st} + 0.5, 0.5)$. Now (58) follows from (57) and (59). □

We also need Lemma 9 which follows from [41, Sec. 8.2.4], [42, Sec. 8.5.1], but is proved in [43, Lemma 9] in the stated form. In order to exploit Lemma 9 (stated next), we need to establish that $0 \leq \mathcal{L}^{-1}(\boldsymbol{X}) \leq 1$. Since Z_{st}s are all beta random variables, their values are in the interval [0,1], hence, $0 \leq \mathcal{L}^{-1}(\boldsymbol{X}) = \prod_{s,t=1}^{m} Z_{st}^{N/2} \leq 1$.

Lemma 9: Consider a random variable W ($0 \leq W \leq 1$) with hth moment

$$\mathbb{E}\{W^h\} = C\left(\frac{\prod_{j=1}^{b} y_j^{y_j}}{\prod_{k=1}^{a} x_k^{x_k}}\right)^h \frac{\prod_{k=1}^{a}\Gamma(x_k(1+h)+\xi_k)}{\prod_{j=1}^{b}\Gamma(y_j(1+h)+\eta_j)} \tag{60}$$

where a and b are integers, C is a constant such that $\mathbb{E}\{W^0\} = 1$ and $\sum_{k=1}^{a} x_k = \sum_{j=1}^{b} y_j$. Define

$$\nu = -2\left[\sum_{k=1}^{a}\xi_k - \sum_{j=1}^{b}\eta_j - \frac{1}{2}(a-b)\right], \tag{61}$$

$$\rho = 1 - \frac{1}{\nu}\left[\sum_{k=1}^{a} x_k^{-1}\left(\xi_k^2 - \xi_k + \frac{1}{6}\right) - \sum_{j=1}^{b} y_j^{-1}\left(\eta_j^2 - \eta_j + \frac{1}{6}\right)\right], \tag{62}$$

$$\beta_k = (1-\rho)x_k, \quad \epsilon_j = (1-\rho)y_j, \tag{63}$$

and

$$\omega_r = \frac{(-1)^{r+1}}{r(r+1)}\left\{\sum_{k=1}^{a}\frac{B_{r+1}(\beta_k+\xi_k)}{(\rho x_k)^r} - \sum_{j=1}^{b}\frac{B_{r+1}(\epsilon_j+\eta_j)}{(\rho y_j)^r}\right\}. \tag{64}$$

Then

$$\begin{aligned} P\{-2\rho\ln(W) \leq z\} &= P\{\chi_\nu^2 \leq z\} + \omega_2\left[P\{\chi_{\nu+4}^2 \leq z\} - P\{\chi_\nu^2 \leq z\}\right] \\ &\quad + \omega_3\left[P\{\chi_{\nu+6}^2 \leq z\} - P\{\chi_\nu^2 \leq z\}\right] \\ &\quad + \Big\{\omega_4\left[P\{\chi_{\nu+8}^2 \leq z\} - P\{\chi_\nu^2 \leq z\}\right] + \frac{1}{2}\omega_2^2 \\ &\quad \times \left[P\{\chi_{\nu+8}^2 \leq z\} - 2P\{\chi_{\nu+4}^2 \leq z\} + P\{\chi_\nu^2 \leq z\}\right]\Big\} \\ &\quad + \sum_{k=1}^{a}\mathcal{O}(x_k^{-5}) + \sum_{j=1}^{b}\mathcal{O}(y_j^{-5}), \end{aligned} \tag{65}$$

where ρ is has been chosen to yield $\omega_1 = 0$. •

Compare (60) with (58) and note that instead of indexing with k and j as in (60), we have double indices s and t in (58). Accounting for these differences, we find the correspondence

$$a = m^2,\ b = m^2,\ x_k \equiv x_{st} = N/2,\ y_j \equiv y_{st} = N/2,$$

$$\xi_k \equiv \xi_{st} = -\frac{m(p-2)+s+t-1}{2} \text{ for } s,t = 1,2,\ldots,m,$$

$$\eta_j \equiv \eta_{st} = -\frac{m(p-2)+s+t-2}{2} \text{ for } s,t = 1,2,\ldots,m,$$

$$C = \frac{\prod_{s,t=1}^{m}\Gamma(b_{st}+1)}{\prod_{s,t=1}^{m}\Gamma(b_{st}+0.5)}. \tag{66}$$

Comparing Lemmas 8 and 9, we further have

$$\beta_k \equiv \beta_{st} = \frac{N}{2}(1-\rho)\ \forall s,t, \quad \epsilon_j \equiv \beta_{st} = \frac{N}{2}(1-\rho)\ \forall s,t. \tag{67}$$

Furthermore, when $h = 0$, the right-side of (58) equals 1, i.e., C is such that $E\{W^0\} = E\{1/\mathcal{L}^0 \,|\, \mathcal{H}_0\} = 1$, as required by Lemma 8. Thus, Lemma 8 is applicable with $W = 1/\mathcal{L}$ and parameters specified in (66).

Using the above values in Lemma 9, we have

$$\begin{aligned} \sum_{k=1}^{a}\xi_k - \sum_{j=1}^{b}\eta_j &= \sum_{s,t=1}^{m}(\xi_{st} - \eta_{st}) \\ &= \sum_{s,t=1}^{m}\frac{-m(p-2)-s-t+1+m(p-2)+s+t-2}{2} \\ &= \sum_{s,t=1}^{m}\frac{-1}{2} = -\frac{m^2}{2}, \end{aligned}$$

Authorized licensed use limited to: Jitendra Tugnait. Downloaded on October 08,2020 at 21:51:17 UTC from IEEE Xplore. Restrictions apply.

$$\sum_{j=1}^{b}\eta_j-\sum_{j=1}^{b}\eta_j^2=\sum_{s,t=1}^{m}\left(\xi_{st}^2-\eta_{st}^2\right)$$
$$=\sum_{s,t=1}^{m}\left(\xi_{st}-\eta_{st}\right)\left(\xi_{st}+\eta_{st}\right)$$
$$=\frac{-1}{2}\sum_{s,t=1}^{m}\frac{-m(p-2)-s-t+1-m(p-2)-s-t+2}{2}$$
$$=\frac{m^3(p-1)}{2}-\frac{m^2}{4}.$$

Using these sums in (61) and (62) and simplifying, we obtain (32) and (34). Similarly, it follows that

$$\sum_{k=1}^{a}\frac{B_{r+1}(\beta_k+\xi_k)}{(\rho x_k)^r}=\sum_{s,t=1}^{m}\frac{B_{r+1}(\beta_{st}+\xi_{st})}{(\rho x_{st})^r}$$
$$=\sum_{s,t=1}^{m}\frac{B_{r+1}((1-\rho)\frac{N}{2}-\frac{m(p-2)+s+t-1}{2})}{(\rho N/2)^r},\tag{68}$$
$$\sum_{j=1}^{b}\frac{B_{r+1}(\epsilon_j+\eta_j)}{(\rho y_j)^r}=\sum_{s,t=1}^{m}\frac{B_{r+1}(\epsilon_{st}+\eta_{st})}{(\rho y_{st})^r}$$
$$=\sum_{s,t=1}^{m}\frac{B_{r+1}((1-\rho)\frac{N}{2}-\frac{m(p-2)+s+t-2}{2})}{(\rho N/2)^r}.\tag{69}$$

Substitution of (68) and (69) in (64) yields (35). We also have $\sum_{k=1}^{a}\mathcal{O}(x_k^{-5})=\sum_{s,t=1}^{m}\mathcal{O}(x_{st}^{-5})=\mathcal{O}(m^2/N^5)$ and $\sum_{j=1}^{b}\mathcal{O}(y_j^{-5})=\sum_{s,t=1}^{m}\mathcal{O}(y_{st}^{-5})=\mathcal{O}(m^2/N^5)$. It then follows from Lemma 9 that (33) holds true. This completes the proof of Theorem 2.

C Proof of Lemma 4

We are interested in determinants of certain matrices. Any elementary row or column operation on a matrix does not change its determinant. Just for this proof, in dealing with $\bar{\mathbf{\Sigma}}$ and $\bar{\mathbf{\Omega}}$, we move the elements associated with the m^2 vertices $\{A_{j,1}\cup A_{k,1}\}$ to the top left corner of the respective matrices, in the order $(j-1)m+1,\ldots,(j-1)m+m,(k-1)m+1,\ldots,(k-1)m+m$. The so modified matrices are still denoted by $\bar{\mathbf{\Sigma}}$ and $\bar{\mathbf{\Omega}}$.

Consider the correspondence of the following conformably partitioned matrices

$$\bar{\mathbf{\Omega}}=\begin{bmatrix}\bar{\mathbf{\Omega}}_{(jkjk)} & \vdots & * \\ \cdots & \cdots & \cdots \\ & \vdots & *\end{bmatrix}=\bar{\mathbf{\Sigma}}^{-1}$$
$$=\begin{bmatrix}\bar{\mathbf{\Sigma}}_{(jkjk)} & \vdots & * \\ \cdots & \cdots & \cdots \\ & \vdots & *\end{bmatrix}^{-1}\tag{70}$$

where $\bar{\mathbf{\Sigma}}_{(jkjk)}$ is defined using $\bar{\mathbf{\Sigma}}$ just as $\bar{\mathbf{\Omega}}_{(jkjk)}$ is defined from $\bar{\mathbf{\Omega}}$ in (43). For ease of notation, redefine some subsets of $\bar{V}$ as

$$V_{-(jk)}:=\bar{V}\backslash\{A_{j,1}\cup A_{k,1}\}$$
$$V_{(jk)}:=\{A_{j,1}\cup A_{k,1}\}$$
$$V_{-(j)}:=\bar{V}\backslash A_{j,1},\quad V_{(j)}:=A_{j,1}$$
$$V_{-(k)}:=\bar{V}\backslash A_{k,1},\quad V_{(k)}:=A_{k,1}.\tag{71}$$

In this notation, we rewrite (36)–(39) as

$$\bar{\mathbf{\Sigma}}_{\bar{V}\backslash\{A_{j,1}\cup A_{k,1}\}}=\bar{\mathbf{\Sigma}}_{V_{-(jk)}}\tag{72}$$
$$\bar{\mathbf{\Sigma}}_{\bar{V}\backslash A_{j,1}}=\bar{\mathbf{\Sigma}}_{V_{-(j)}}\tag{73}$$
$$\bar{\mathbf{\Sigma}}_{\bar{V}\backslash A_{k,1}}=\bar{\mathbf{\Sigma}}_{V_{-(k)}}.\tag{74}$$

From [1, Eqn. (B.2)] and (70), we have (recall the notation in (40))

$$\bar{\mathbf{\Omega}}_{(jkjk)}^{-1}=\bar{\mathbf{\Sigma}}_{V_{(jk)}}-\bar{\mathbf{\Sigma}}_{V_{(jk)},V_{-(jk)}}\bar{\mathbf{\Sigma}}_{V_{-(jk)}}^{-1}\bar{\mathbf{\Sigma}}_{V_{-(jk)},V_{(jk)}}.\tag{75}$$

By [1, Proposition C.5], the right-side of (75) is the conditional covariance

$$\bar{\mathbf{\Sigma}}_{V_{(jk)}|V_{-(jk)}}=\mathbb{E}\{\boldsymbol{x}_{V_{(jk)}}(t)\boldsymbol{x}_{V_{(jk)}}^{\top}(t)|\boldsymbol{x}_{V_{-(jk)}}(t)\}$$
$$=\bar{\mathbf{\Sigma}}_{V_{(jk)}}-\bar{\mathbf{\Sigma}}_{V_{(jk)},V_{-(jk)}}\bar{\mathbf{\Sigma}}_{V_{-(jk)}}^{-1}\bar{\mathbf{\Sigma}}_{V_{-(jk)},V_{(jk)}}.\tag{76}$$

Thus

$$\bar{\mathbf{\Omega}}_{(jkjk)}^{-1}=\bar{\mathbf{\Sigma}}_{V_{(jk)}|V_{-(jk)}}.\tag{77}$$

Similarly,

$$\bar{\mathbf{\Omega}}_{(jj)}^{-1}=\bar{\mathbf{\Sigma}}_{V_{(j)}|V_{-(j)}}.\tag{78}$$

By [1, Eqn. (B.1)] concerning determinant of partitioned matrices, we have

$$|\bar{\mathbf{\Sigma}}_{\bar{V}}|=|\bar{\mathbf{\Sigma}}_{\bar{V}\backslash A_{j,1}}|\,|\bar{\mathbf{\Sigma}}_{A_{j,1}|\bar{V}\backslash A_{j,1}}|$$
$$=|\bar{\mathbf{\Sigma}}_{V_{-(j)}}|\,|\bar{\mathbf{\Sigma}}_{V_{(j)}|V_{-(j)}}|.\tag{79}$$

Similarly, we have

$$|\bar{\mathbf{\Sigma}}_{\bar{V}\backslash A_{k,1}}|=|\bar{\mathbf{\Sigma}}_{V_{-(k)}}|=|\bar{\mathbf{\Sigma}}_{V_{-(jk)}}|\,|\bar{\mathbf{\Sigma}}_{V_{(j)}|V_{-(jk)}}|.\tag{80}$$

Using notation (71), (79) and (80), we have

$$\frac{|\bar{\mathbf{\Sigma}}_{\bar{V}\backslash\{A_{j,1}\cup A_{k,1}\}}|\,|\bar{\mathbf{\Sigma}}_{\bar{V}}|}{|\bar{\mathbf{\Sigma}}_{\bar{V}\backslash A_{j,1}}|\,|\bar{\mathbf{\Sigma}}_{\bar{V}\backslash A_{k,1}}|}=\frac{|\bar{\mathbf{\Sigma}}_{V_{-(jk)}}|\,|\bar{\mathbf{\Sigma}}_{\bar{V}}|}{|\bar{\mathbf{\Sigma}}_{V_{-(j)}}|\,|\bar{\mathbf{\Sigma}}_{V_{-(k)}}|}$$
$$=\frac{|\bar{\mathbf{\Sigma}}_{V_{(j)}|V_{-(j)}}|}{|\bar{\mathbf{\Sigma}}_{V_{(j)}|V_{-(jk)}}|}.\tag{81}$$

Since $\bar{\mathbf{\Omega}}_{(jkjk)}^{-1}=\bar{\mathbf{\Sigma}}_{V_{(jk)}|V_{-(jk)}}$ by (77), we have

$$\bar{\mathbf{\Sigma}}_{V_{(j)}|V_{-(jk)}}=[\bar{\mathbf{\Omega}}_{(jkjk)}^{-1}]_{1:m,1:m}.\tag{82}$$

Using (78), (81) and (82), we have the desired result. ■

REFERENCES

[1] S. L. Lauritzen, *Graphical Models*. Oxford, UK: Oxford Univ. Press, 1996.
[2] J. Whittaker, *Graphical Models in Applied Multivariate Statistics*. New York: Wiley, 1990.

Authorized licensed use limited to: Jitendra Tugnait. Downloaded on October 08,2020 at 21:51:17 UTC from IEEE Xplore. Restrictions apply.

[3] P. Bühlmann and S. van de Geer, *Statistics for High-Dimensional Data*. Berlin: Springer, 2011.
[4] A. O. Hero, III, and B. Rajaratnam, "Foundational principles for large-scale inference: Illustrations through correlation mining," *Proc. IEEE*, vol. 64, no. 104, pp. 93–110, Jan. 2016.
[5] P. Danaher, P. Wang and D. M. Witten, "The joint graphical lasso for inverse covariance estimation across multiple classes," *J. Roy. Statistical Soc., Series B (Methodological)*, vol. 76, pp. 373–397, 2014.
[6] N. Friedman, "Inferring cellular networks using probabilistic graphical models," *Science*, vol. 303, pp. 799–805, 2004.
[7] N. Meinshausen and P. Bühlmann, "High-dimensional graphs and variable selection with the Lasso," *Ann. Statist.*, vol. 34, no. 3, pp. 1436–1462, 2006.
[8] K. Mohan, P. London, M. Fazel, D. Witten and S. I. Lee, "Node-based learning of multiple Gaussian graphical models," *J. Mach. Learn. Res.*, vol. 15, pp. 445–488, 2014.
[9] J. Chiquet, G. Rigaill and M. Sundquist, "A multiattribute Gaussian graphical model for inferring multiscale regulatory networks: An application in breast cancer," Chapter 6 in G. Sanguinetti, and V. Huynh-Thu, (Eds), *Gene Regulatory Networks: Methods and Protocols*, vol. 1883 in series *Methods in Molecular Biology*, pp. 143–160, Humana Press, New York, NY, 2019.
[10] M. Kolar, H. Liu and E. P. Xing, "Markov network estimation from multi-attribute data," in *Proc. 30th Intern. Conf. Mach. Learn.*, Atlanta, GA, 2013.
[11] M. Kolar, H. Liu and E. P. Xing, "Graph estimation from multi-attribute data," *J. Mach. Learn. Res.*, vol. 15, pp. 1713–1750, 2014.
[12] J. K. Tugnait, "Edge exclusion tests for improper complex Gaussian graphical model selection," *IEEE Trans. Signal Process.*, vol. 67, no. 13, pp. 3547–3560, Jul. 1, 2019.
[13] X. Du and S. Ghosal, "Multivariate Gaussian network structure learning," *J. Stat. Plann. Inf.*, vol. 199, pp. 327–342, Mar. 2019.
[14] N. Katenka and E. D. Kolaczyk, "Multi-attribute networks and the impact of partial information on inference and characterization," *Ann. Appl. Statist.*, vol. 6, no. 3, pp. 1068–1094, 2012.
[15] E. Pavez and A. Ortega, "Generalized Laplacian precision matrix estimation for graph signal processing," in *Proc. IEEE ICASSP*, Shanghai, China, Mar. 2016, pp. 6350–6354.
[16] E. Pavez, H. E. Egilmez, and A. Ortega, "Learning graphs with monotone topology properties and multiple connected components," *IEEE Trans. Signal Process.*, vol. 66, no. 9, pp. 2399–2413, May 1, 2018.
[17] G. Fracastoro, D. Thanou and P. Frossard, "Graph transform optimization with application to image compression," *IEEE Trans. Image Process.*, vol. 29, pp. 419–432, 2020.
[18] E. Pavez, H. E. Egilmez, Y. Wang, and A. Ortega, "GTT: Graph template transforms with applications to image coding," in *Proc. Picture Coding Symp.*, pp. 199-203, Cairns, Australia, 2015.
[19] X. Dong, D. Thanou, M. Rabbat and P. Frossard, "Learning graphs from data," *IEEE Signal Process. Mag.*, vol. 36, no. 3, pp. 44–63, 2019.
[20] F. Hu, Z. Lu, H. Wong and T. P. Yuen, "Analysis of air quality time series of Hong Kong with graphical modeling," *Environmetrics*, vol. 27, no. 3, pp. 169–181, May 2016.
[21] R. Dahlhaus, "Graphical interaction models for multivariate time series," *Metrika*, vol. 51, pp. 157–172, 2000.
[22] P.W. F. Smith and J. Whittaker, "Edge exclusion tests for graphical Gaussian models," in *Proc. Learn. Graphical Models*, M. I. Jordan (Ed), pp. 555–574, Cambridge, MA: MIT Press, 1998.
[23] G. Marrelec *et al.*, "Partial correlation for functional brain interactivity investigation in functional MRI," *Neuroimage*, vol. 32, no. 1, pp. 228–237, 2006.
[24] M. Drton and M. D. Perlman, "Multiple testing and error control in Gaussian graphical model selection," *Statist. Sci.*, vol. 22, pp. 430–439, 2007.
[25] S. Ryali, T. Chen, K. Supekar and V. Menon, "Estimation of functional connectivity in fMRI data using stability selection-based sparse partial correlation with elastic net penalty," *Neuroimage*, vol. 59, pp. 3852–3861, 2012.
[26] A. J. Rothman, P. J. Bickel, E. Levina and J. Zhu, "Sparse permutation invariant covariance estimation," *Electron. J. Statist.*, vol. 2, pp. 494–515, 2008.
[27] W. Liu, "Gaussian graphical model estimation with false discovery rate control," *Ann. Statist.*, vol. 41, no. 6, pp. 2948–2978, 2013.
[28] Z. Ren, T. Sun, C-H. Zhang and H. H. Zhou, "Asymptotic normality and optimalties in estimation of large Gaussian graphical models," *Ann. Statist.*, vol. 43, no. 3, pp. 991–1026, 2015.
[29] J. Janková and S. van de Geer, "Inference in high-dimensional graphical models," Chapter 14 in M. H. Maathuis, M. Drton, S. Lauritzen and W. Wainwright (Eds), *Handbook Graphical Models*, pp. 325–348, CRC Press, Boca Raton, FL, 2018.
[30] Y. Ning and H. Liu, "A general theory of hypothesis tests and confidence regions for sparse high dimensional models," *Ann. Statist.*, vol. 45, no. 1, pp. 158–195, 2017.
[31] Y. Xia, T. Cai, and T. T. Cai, "Multiple testing of submatrices of a precision matrix with applications to identification of between pathway interactions," *J. Am. Stat. Assoc.*, vol. 113, no. 521, pp. 328–339, 2018.
[32] H. H. Andersen, M. Hojbjerre, D. Sorensen, and P. S. Eriksen, *Linear and Graphical Models for Multivariate Complex Normal Distribution*, Lecture Notes in Statistics, vol. 101. New York: Springer-Verlag, 1995.
[33] Y. Matsuda, "A test statistic for graphical modelling of multivariate time series," *Biometrika*, vol. 93, no. 2, pp. 399–409, 2006.
[34] R. J. Wolstenholme, and A. T. Walden, "An efficient approach to graphical modeling of time series," *IEEE Trans. Signal Process.*, vol. 64, no. 12, pp. 3266–3276, Jun. 2015.
[35] G. E. P. Box, "A general distribution theory for a class of likelihood criteria," *Biometrika*, vol. 36, pp. 317–246, Dec. 1949.
[36] J. K. Tugnait, "An edge exclusion test for complex Gaussian graphical model selection," in *Proc. IEEE Statistical Signal Process. Workshop*, Freiburg, Germany, Jun. 2018, pp. 678–682.
[37] L. Wasserman, *All of Statistics*, New York, NY: Springer, 2004.
[38] Y. Benjamini and D. Yekutieli, "The control of the false discovery rate in multiple testing under dependency," *Ann. Stat.*, vol. 29, no. 4, pp. 1165–1188, 2001.
[39] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Roy. Statist. Soc., Series B (Methodological)*, vol. 57, pp. 289–300, 1995.
[40] Amsterdam Library of Textures (ALOT), [Online]. Available: https://aloi.science.uva.nl/public_alot
[41] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. New York: Wiley, 1982.
[42] T. W. Anderson, *An Introduction to Multivariate Statistical Anal.*, Third edition. Hoboken, NJ: Wiley, Wiley-Interscience, 2003.
[43] J. K. Tugnait and S. A. Bhaskar, "On testing for impropriety of multivariate complex-valued random sequences," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2988–3003, Jun. 1, 2017.

Jitendra K. Tugnait (Life Fellow, IEEE) received the B.Sc. (with Hons.) degree in electronics and electrical communication engineering from the Punjab Engineering College, Chandigarh, India, in 1971, the M.S. and E.E. degrees from Syracuse University, Syracuse, NY, USA, in 1973 and 1974, respectively, and the Ph.D. degree from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 1978, all in electrical engineering.

From 1978 to 1982, he was an Assistant Professor of electrical and computer engineering with the University of Iowa, Iowa City, IA, USA. He was with the Long Range Research Division of the Exxon Production Research Company, Houston, TX, USA, from June 1982 to September 1989. He joined the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL, USA, in September 1989 as a Professor, where he is currently James B. Davis Professor. His current research interests are in statistical signal processing, wireless physical and secure communications, and multiple target tracking.

Dr. Tugnait has served as an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE SIGNAL PROCESSING LETTERS, and IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and as a Senior Area Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and a Senior Editor for the IEEE WIRELESS COMMUNICATIONS LETTERS.

Authorized licensed use limited to: Jitendra Tugnait. Downloaded on October 08,2020 at 21:51:17 UTC from IEEE Xplore. Restrictions apply.