Searching for a source of difference in graphical models

Searching for a source of difference in graphical models

Vera Djordjilović[a,*], Monica Chiogna[b]

[a]Department of Economics, Ca' Foscari University of Venice, Italy
[b]Department of Statistical Sciences "Paolo Fortunati", University of Bologna, Italy

Abstract

We look at a two-sample problem within the framework of decomposable graphical models. When the global hypothesis of equality of two distributions is rejected, the interest is usually in localizing the source of difference. Motivated by the idea that diseases can be seen as system perturbations, and by the need to distinguish between the origin of perturbation and components affected by the perturbation, we introduce the concept of a minimal seed set, and its graphical counterpart a graphical seed set. They intuitively consist of variables driving the difference between the two conditions. We propose a simple testing procedure, linear in the number of nodes, to estimate the graphical seed set from data. We illustrate our approach in the context of gene set analysis, where we show that is possible to zoom in on the origin of perturbation in a gene network.

Keywords: Decomposable graphical models, Decomposition, Gaussian graphical models, Graphical log-linear models, Strong meta Markov models, Two sample problem.
2020 MSC: Primary 62H22, Secondary 62F03

1. Introduction

1.1. Motivation

The present work is motivated by the problem of identifying the origin of perturbation in gene regulatory networks. In biological networks, diseases can be modelled as perturbations that affect certain targets, which, once perturbed, propagate the perturbation through network connections [6]. In practice, we often collect and compare observations from healthy individuals and observations from patients after the disease related perturbation has already taken place. On the basis of this comparison, it is of interest to identify the site of original perturbation, i.e., the source of difference, and distinguish it from the elements of the network that were affected through the process of network propagation.

1.2. Statement of the problem and some notation

Let $\mathcal{F} = \{P_\theta; \theta \in \Theta\}$, $\Theta \subset R^d$, be a family, parametrized by θ, of probability distributions for the random vector X_V, indexed by a set V, $|V| = p$, with support $\mathcal{X}_V$. In what follows, to unburden the notation and when no ambiguity can arise, we adopt the notation of [5] and, allowing for a slight abuse of notation, we write θ instead of P_θ to denote individual distributions belonging to $\mathcal{F}$. For $A, B \subseteq V$, we will further write θ_A to denote (the parameters of) the marginal distribution of variables in A and, similarly, $\theta_{A|B}$ to denote a collection of conditional distributions $\left\{\theta_{A|X_B=y}, y \in \mathcal{X}_B\right\}$ indexed by y, where X_B, $B \subseteq V$, is a subvector of X_V and $\mathcal{X}_B$ is the associated support. Different experimental conditions will be distinguished by use of superscripts.

Consider a random vector $X_V \sim P_\theta$. Within the context of two sample problems, the interest is often in testing the null hypothesis of equality of distributions $H_0 : \theta^{(1)} = \theta^{(2)}$. If that hypothesis is rejected, one usually aims at localizing the source of difference.

A common approach to tackle the question in genomics applications is to focus on the p univariate marginal distributions, see for instance [16] for a particularly popular method choice. Marginally speaking, a variable X_v, $v \in V$, can be considered relevant to the aim at hand if its marginal distribution is different in $P_{\theta^{(1)}}$ and $P_{\theta^{(2)}}$.

*Corresponding author. Email address: vera.djordjilovic@unive.it

The (index) set of the relevant variables is then taken to be

$$R = \left\{ v \in V : \theta_v^{(1)} \neq \theta_v^{(2)} \right\}.$$

Whether a variable belongs to R depends solely on its marginal distribution.

Although simple and computationally feasible, the marginal approach might fail to point to the true source of difference whenever an interplay between variables plays a role in differentiating the two distributions [10]. In that case, we propose to privilege a conditional perspective and exploit an approach which takes into account the entire p-dimensional joint distribution and flags a variable relevant only if the difference in its marginal distribution cannot be explained by the remaining variables. We define the set of conditionally relevant variables D as follows.

Definition 1 (Seed set)**.** Consider $\theta^{(1)}, \theta^{(2)} \in \mathcal{F}$. We call the set $D \subseteq V$ the seed set, if the collections of conditional laws $\theta^{(1)}_{V\setminus D|D}$ and $\theta^{(2)}_{V\setminus D|D}$ coincide. Furthermore, we say that D is a minimal seed set, if no proper subset of it is itself a seed set.

To facilitate the understanding of the above definition, it is helpful to consider that, by employing the factorization $p(x;\theta) = p(x_D;\theta_D)p(x_{\bar{D}} \mid x_D;\theta_{\bar{D}|D})$, where $\bar{D} = V \setminus D$, the likelihood ratio $p(x;\theta^{(1)})/p(x;\theta^{(2)})$ simplifies to $p(x_D;\theta^{(1)}_D)/p(x_D;\theta^{(2)}_D)$. The likelihood ratio thus depends only on variables in D. When comparing the two distributions, the variables outside of D are either irrelevant or redundant and D can be seen as the minimal subset of variables explaining the difference between the two distributions. It should be stressed that there is no relation between R and D; in general neither $R \subseteq D$ nor $D \subseteq R$.

In practice, to identify the seed set, D needs to be estimated from data. One could perform a number of tests of equality of conditional distributions, but when p is large, this testing problem becomes extremely challenging, and represents an open area of research, see for instance [25] and references therein. In this paper, we assume that the dependence structure among the p variables in the joint distribution can be well represented by an undirected graph. We then address the problem of identifying D within the framework of graphical models, where we exploit the structural modularity of decomposable graphical models [5, 8]. To this aim, we assume that $\mathcal{F}$ is a strong meta Markov model with respect to a given undirected decomposable graph $G = (V, E)$, where $E \subseteq V \times V$ is a set of edges. Let us denote by $\mathcal{M}(G)$ a family of distributions satisfying the global Markov property relative to G. According to the definition introduced by [5], $\mathcal{F} \subseteq \mathcal{M}(G)$ is a strong meta Markov model if for any decomposition (A, B) of G, parameters θ_A and $\theta_{B|A}$ are variation independent in $\mathcal{F}$ [2, p.26]. In other words, all possible values of θ_A are logically compatible with all possible values of $\theta_{B|A}$.

Under this assumption, there is a close relationship between the parametric model structure and the underlying graph, and we show that the problem of identifying D can be formulated as the problem of testing equality of lower dimensional conditional distributions induced by the structure of G. We further show that the associated test statistics are functions of the quantities pertaining to the lower dimensional marginal distributions. The key advantage is that inference on marginal distributions is significantly less challenging than inference on conditional distributions. Beside the computational gain, we argue that the proposed approach addresses the issue of exploiting information on the structure of dependence in an efficient and elegant way.

2. Decomposition of the global hypothesis of equality of two Markov distributions

A major appeal of decomposable graphs in graphical modelling is that they allow for a clique-grained decomposition of the statistical model. Let $C_1, \ldots, C_k$ be a sequence of cliques of G satisfying a running intersection property (see Section 1 of Supplementary material), and let $S_2, \ldots, S_k$ be an associated sequence of (possibly non-unique) separators. Then, if the distribution of X_V is Markov relative to G, its joint distribution decomposes as:

$$p(x_V) = p(x_{C_1}) \prod_{j=2}^{k} p(x_{R_j} \mid x_{S_j}),$$

where $R_j = C_j \setminus S_j$, $j \in \{2, \ldots, k\}$. Therefore, each distribution $\theta \in \mathcal{F}$ can be uniquely decomposed into k lower dimensional components: $\theta_{C_1}, \theta_{R_2|S_2}, \ldots, \theta_{R_k|S_k}$; uniqueness ensures that θ can be reconstructed back from its components. As a consequence, the global hypothesis of equality $H : \theta^{(1)} = \theta^{(2)}$ also decomposes along the perfect ordering

as $H = \cap_{j=1}^{k} H_j$, where $H_1 : \theta^{(1)}_{C_1} = \theta^{(2)}_{C_1}$ and $H_j : \theta^{(1)}_{R_j|S_j} = \theta^{(2)}_{R_j|S_j}$, $j \in \{2, \ldots, k\}$. Since $\mathcal{F}$ is a strong meta Markov model, the components of θ are variation independent and there are no logical relations among the H_j. The following result states that the log-likelihood ratio for H decomposes analogously and that all component test statistics can be computed in clique-induced marginal models.

Theorem 1. *Let $X^{(1)}_{V,1}, \ldots, X^{(1)}_{V,n_1}$ and $X^{(2)}_{V,1}, \ldots, X^{(2)}_{V,n_2}$ be two independent random samples from, respectively, $\theta^{(1)}$ and $\theta^{(2)}$, $\theta^{(l)} \in \mathcal{F}$, $l \in \{1, 2\}$, where $\mathcal{F}$ is strong meta Markov model relative to G. $H : \theta^{(1)} = \theta^{(2)}$ and its decomposition $H = \cap_{j=1}^{k} H_j$, where $H_1 : \theta^{(1)}_{C_1} = \theta^{(2)}_{C_1}$ and $H_j : \theta^{(1)}_{R_j|S_j} = \theta^{(2)}_{R_j|S_j}$, $j \in \{2, \ldots, k\}$. Let $\lambda(V)$ denote the log likelihood ratio criterion for testing H against a general alternative and let $\lambda(A)$ denote the log likelihood ratio criterion for testing equality of distributions induced by $A \subseteq V$. The following equality holds*

$$\lambda(V) = \lambda(C_1) + \sum_{j=2}^{k} \left\{ \lambda(C_j) - \lambda(S_j) \right\}, \tag{1}$$

where $\left\{ \lambda(C_j) - \lambda(S_j) \right\}$ represents the log likelihood ratio for testing H_j. Moreover, the k terms on the right hand side of (1) *are asymptotically independent under the null hypothesis.*

Proof of Theorem 1: The joint density of any random sample of size n from $\theta \in \mathcal{F}$ factorizes as

$$p(x_{(n)}; \theta) = p\left(x_{C_1,(n)}; \theta_{C_1}\right) \prod_{j=2}^{k} p(x_{R_j,(n)} \mid x_{S_j,(n)}; \theta_{R_j|S_j}), \tag{2}$$

where $x_{(n)}$ stands for $x_1, \ldots, x_n$. Each component can be maximized separately to obtain maximum likelihood estimates $\hat{\theta}_{C_1}$ and $\hat{\theta}_{R_j|S_j}$. Note that maximum likelihood estimate of θ_{C_1} is the same whether based on $x_{(n)}$ or $x_{C_1,(n)}$.

The likelihood ratio for testing H is

$$L(x_{(n_1+n_2)}) = \frac{p\left(x_{(n_1+n_2)}; \hat{\theta}\right)}{p\left(x^{(1)}_{(n_1)}; \hat{\theta}^{(1)}\right) p\left(x^{(2)}_{(n_2)}; \hat{\theta}^{(2)}\right)},$$

where $x_{(n_1+n_2)}$ denotes a pooled sample, $\hat{\theta}$ is the maximum likelihood estimate of $\theta^{(1)} = \theta^{(2)}$ under the null hypothesis, and $\hat{\theta}^{(l)}$, $l \in \{1, 2\}$, is the maximum likelihood estimate of $\theta^{(l)}$ under the alternative. Factorizing each density as in (2), L is decomposed into k components corresponding to the local hypotheses H_j, $j \in \{1, \ldots, k\}$. Using the equality $\theta_{R_j|S_j}(x_{R_j} \mid x_{S_j}) = \theta_{C_j}(x_{C_j})/\theta_{S_j}(x_{S_j})$, we obtain the expression $\lambda(V) = \lambda(C_1) + \sum_{j=2}^{k} \left\{ \lambda(C_j) - \lambda(S_j) \right\}$. Finally, given the modular structure of the joint distribution, the number of degrees of freedom associated to $\lambda(V)$ is exactly the sum of the degrees of freedom of the summands on the right hand-side, which is a sufficient condition for the asymptotic independence of Chi-square random variables [19]. □

In what follows, we give explicit expressions for the decomposition for two important parametric families of distributions.

2.1. Gaussian graphical models

Consider a subfamily of $\mathcal{M}(G)$ composed of Gaussian graphical models. In this case, $\theta = (\mu, \Sigma)$, with $\mu \in \mathbb{R}^p$ and Σ a symmetric positive definite matrix such that $\Sigma^{-1} \in S^+(G)$, where $S^+(G)$ denotes the set of all symmetric $p \times p$ positive definite matrices with zeros corresponding to the missing edges of G. For $A, B \subset V$, let Σ_{AB} denote the corresponding submatrix of Σ and let Σ_A stand for Σ_{AA}.

For a given perfect clique ordering, the global hypothesis of equality $H : \theta^{(1)} = \theta^{(2)}$ decomposes as $H = \cap_{j=1}^{k} H_j$, with $H_1 : \mu^{(1)}_{C_1} = \mu^{(2)}_{C_1}, \Sigma^{(1)}_{C_1} = \Sigma^{(2)}_{C_1}$ and $H_j : \theta^{(1)}_{R_j|S_j} = \theta^{(2)}_{R_j|S_j}$, $j \in \{2, \ldots, k\}$, where

$$\theta_{A|B} = (\mu_A - \Sigma_{AB}\Sigma_B^{-1}\mu_B, \Sigma_{AB}\Sigma_B^{-1}, \Sigma_A - \Sigma_{AB}\Sigma_B^{-1}\Sigma_{BA}),$$

for $A, B \subset V$, denotes parameters of the conditional law.

Given two independent random samples of sizes n_1 and n_2 from $\theta^{(1)}$ and $\theta^{(2)}$, respectively, the log likelihood ratio $\lambda(A), A \subseteq V$, for testing the associated null hypothesis of equality is

$$\lambda(A) = \sum_{l=1}^{2} n_l \log \frac{|\hat{\Sigma}_A|}{|\hat{\Sigma}_A^{(l)}|},$$

where $|\hat{\Sigma}|$ is determinant of the maximum likelihood estimate of Σ under H, and $\hat{\Sigma}^{(l)}, l \in \{1, 2\}$, are maximum likelihood estimates of $\Sigma^{(l)}$ under the general alternative [1, p.416]. Since $(\hat{\Sigma})^{-1}, (\hat{\Sigma}^{(l)})^{-1} \in S^+(G)$, $l \in \{1, 2\}$, and the determinant of every Ω for which $\Omega^{-1} \in S^+(G)$ can be decomposed with respect to the graph as $|\Omega| = \prod_{i=1}^{k} |\Omega_{C_i}| / \prod_{i=2}^{k} |\Omega_{S_i}|$ [13, p.145], the log likelihood ratio $\lambda(V)$ can be equivalently written as $\lambda(V) = \sum_{i=1}^{k} \lambda(C_i) - \sum_{i=2}^{k} \lambda(S_i)$, from which equality of Theorem 1 follows. It is important to stress that when subgraph induced by A is complete, which is the case with cliques C_i and separators S_i, then maximum likelihood estimate of Σ_A is unconstrained. In particular, if for ease of notation we temporarily drop the index A in $x_A^{(l)}$, $l \in \{1, 2\}$, and write $x^{(l)}$ instead, we have

$$\hat{\Sigma}_A = \frac{1}{n_1 + n_2} \left[\sum_{i=1}^{n_1} (x_i^{(1)} - \bar{x})(x_i^{(1)} - \bar{x})^T + \sum_{j=1}^{n_2} (x_j^{(2)} - \bar{x})(x_j^{(2)} - \bar{x})^T \right],$$

where $\bar{x} = (n_1 \bar{x}_1 + n_2 \bar{x}_2)/(n_1 + n_2)$, whereas $\hat{\Sigma}_A^{(1)}$ and $\hat{\Sigma}_A^{(2)}$ are unconstrained estimates of Σ_A computed in the two samples, i.e.

$$\hat{\Sigma}_A^{(l)} = \frac{1}{n_l} \sum_{i=1}^{n_l} (x_i^{(l)} - \bar{x}_l)(x_i^{(l)} - \bar{x}_l)^\top, \quad \bar{x}_l = \frac{1}{n_l} \sum_{i=1}^{n_l} x_i^{(l)}, \quad l \in \{1, 2\}.$$

In other words, it is possible to compute $\lambda(V)$ from test statistics computed in clique-induced marginal models in which maximum likelihood estimation is unconstrained.

2.2. Graphical log-linear models

Consider a subfamily $\mathcal{P} \subset \mathcal{M}(G)$ of graphical log-linear models. Each X_v is now a categorical random variable with a finite set of possible values or levels $\mathcal{I}_v$. Here, $\mathcal{X}_V = \times_{v \in V} \mathcal{I}_v$. We refer to the elements of $\mathcal{X}_V$ as table cells [13, Chapter 4]). Let $X_{V,1}, \ldots, X_{V,n}$ be $n \in \mathbb{N}$ independent realizations of X_V. Cell counts are defined as

$$n(h) = \sum_{i=1}^{n} I\{X_{V,i} = h\}, \quad h \in \mathcal{X}_V,$$

where $I\{\cdot\}$ denotes the indicator function.

For $A \subset V$, table cells $h_A \in \mathcal{I}_A = \times_{v \in A} \mathcal{I}_v$ are obtained by classifying observations only with respect to the variables in A. Marginal cell counts are $n(h_A) = \sum_{i=1}^{n} I\{X_{A,i} = h_A\}$, where $X_{A,i}$ is a subvector of $X_{V,i}$ induced by A.

Under a multinomial sampling scheme, the probability of the observed cell counts is

$$\Pr(N(h) = n(h), h \in \mathcal{X}_V) = \frac{n!}{\prod_{h \in \mathcal{X}_V} n(h)!} \prod_{h \in \mathcal{X}_V} p(h)^{n(h)},$$

where $p(h)$ is the probability for cell $h \in \mathcal{X}_V$. In this case, $\theta = \{p(h)\}_{h \in \mathcal{X}_V}$ satisfies the constraint $\sum_{h \in \mathcal{X}_V} p(h) = 1$ and decomposes as $\theta_{C_1} = \{p(h_{C_1})\}_{h_{C_1} \in \mathcal{X}_{C_1}}$, which refers to the marginal table induced by C_1, and $\theta_{R_j|S_j}$ for $j \in \{2, \ldots, k\}$, where $\theta_{A|B} = \{p(h_A \mid h_B)\}_{h_{A \cup B} \in \mathcal{X}_{A \cup B}}$ refers to the parameters of the h_B-slice of the table, i.e., a table in which objects are classified with respect to the variables in A for a given fixed level of the variables in B.

Consider now $\theta^{(1)}, \theta^{(2)} \in \mathcal{P}$ and the null hypothesis of equality of probabilities in the marginal table induced by $A \subseteq V$. Given two independent random samples with observed cell counts $n^{(1)}$ and $n^{(2)}$ from $\theta^{(1)}$ and $\theta^{(2)}$, respectively, the log likelihood ratio $\lambda(A)$ is

$$\lambda(A) = 2 \left\{ \sum_{h_A \in \mathcal{X}_A} \sum_{l=1}^{2} n^{(l)}(h_A) \log \left(\frac{\hat{p}^{(l)}(h_A)}{\hat{p}(h_A)} \right) \right\},$$

where $\hat{p}_A$ is the maximum likelihood estimate of p_A under the null hypothesis; and $\hat{p}_A^{(1)}$ and $\hat{p}_A^{(2)}$ are maximum likelihood estimates of $p_A^{(1)}$ and $p_A^{(2)}$ under a general alternative. Using the structural decomposition reflected in the maximum likelihood estimator $\hat{p}$:

$$\hat{p}(h) = \frac{\prod_{j=1}^{k} \hat{p}(h_{C_j})}{\prod_{j=2}^{k} \hat{p}(h_{S_j})}, \quad h \in \mathcal{X}_V, \tag{3}$$

we obtain the decompisiton of $\lambda(V)$ featured in Theorem 1. Degrees of freedom associated to $\lambda(V)$ can be computed from the formula $f(V) = f(C_1) + \sum_{j=2}^{k} \left\{ f(C_j) - f(S_j) \right\}$, where $f(A)$ denotes degrees of freedom in a model induced by $A \subseteq V$. Since marginal models induced by cliques and separators are saturated, their degrees of freedom are obtained as $f(C_j) = \prod_{v \in C_j} |\mathcal{I}_v| - 1$, and analogously for separators.

3. Estimation

3.1. The graphical seed set

Before we show how the result of the previous section can be used to make inference about the seed set, we need to introduce the concept of the graphical seed set. Namely, by employing a clique-grained decomposition, we are not always able to identify the minimal seed set; in those cases we can identify its superset that we denote by D_G. Relation between the two sets, that depends on both D and G, is the subject of this section.

Definition 2 (Graphical seed set)**.** Let D be a minimal seed set for $\theta^{(1)}$ and $\theta^{(2)}$, two graphical distributions Markov with respect to G. Let $\mathcal{S} = \{S : S \text{ is a separator in } G\}$ be the collection of separators in G. Then we call the set

$$D_G = \{v \in V \mid \forall S \in \mathcal{S}, \text{ either } v \in S \text{ or } S \text{ does not separate } v \text{ from } D \text{ in } G\} \tag{4}$$

a graphical seed set.

In the above definition, we allow for non-empty intersection between S and D, as well as $S = D$. When $v \in D$, the condition (4) is trivially satisfied (v cannot be separated from D by any set), and therefore $D_G \supseteq D$. The graphical seed set D_G is thus the smallest set containing the seed set D that can be identified by means of set operations on cliques and separators of G.

When the minimal seed set is a separator, we can set $S = D$ in (4), to obtain $D = D_G$. In general, D and D_G will coincide whenever D can be expressed as an intersection of two or more cliques. In other instances, D_G will be a seed set, but not a minimal one. For an illustrative example, see Section 2 of the Supplementary material.

3.2. The graphical seed set estimator

We have seen above that the global hypothesis of equality can be decomposed according to a specified perfect ordering into a set of local hypotheses. However, the perfect ordering is not unique. In fact, there are multiple decompositions of the global hypothesis, each corresponding to a different factorization of the same distribution. It is this multiplicity that we exploit when estimating the graphical seed set.

For a given graph, the enumeration of all decompositions might resemble the problem of enumerating its junction trees [20], but a closer look reveals that it is a far simpler task. Given the uniqueness of the sequence of separators, it is not difficult to show that there is exactly one decomposition for each choice of the root clique – the clique labeled C_1 – leading to a total of k decompositions.

Before we show how these different decompositions relate to the graphical seed set in Proposition 1, we introduce some notation and restate the global testing problem in decision theory terms. Let $\Theta \times \Theta$ be the unrestricted parameter space of $(\theta^{(1)}, \theta^{(2)})$; let $\Theta_0 = \{(\theta, \theta); \theta \in \Theta\}$ denote the space restricted by $H : \theta^{(1)} = \theta^{(2)}$, and let $\Theta_1 = (\Theta \times \Theta) \setminus \Theta_0$. We want to test $H : (\theta^{(1)}, \theta^{(2)}) \in \Theta_0$ against a general alternative $(\theta^{(1)}, \theta^{(2)}) \in \Theta_1$. Let the decision taken on H be denoted by d, where $d = 0$ means that the null hypothesis is not rejected and $d = 1$ means that the null hypothesis is rejected. A test ϕ is a mapping from the sample space to the set $\{0, 1\}$ (we rule out the trivial case that the test makes no decisions). Let d^* denote the correct decision (the truth) for H. As seen in the previous Section, the null hypothesis can be decomposed into a set of independent local hypotheses, i.e., $H = \bigcap_{j=1}^{k} H_j$, and we denote by d_j^* the correct decision for H_j, $j \in \{1, \dots k\}$, so that $d^* = (d_1^*, \dots, d_k^*)$. To identify the $i-$th decomposition, obtained when C_i is set

as the root clique, we let $C_{i,1}, \ldots, C_{i,k}$ denote a sequence of cliques satisfying the running intersection property. Let $S_{i,2}, \ldots, S_{i,k}$ be an associated sequence of separators, and set $S_{i,1} = \varnothing$, $i \in \{1, \ldots, k\}$. In this notation, $H_{i,j}$ will denote the $j-$th null hypothesis in decomposition i, $\phi_{i,j}$ the corresponding test, and $d^*_{i,j}$ the associated correct decision.

We now show the connection between the graphical seed set and the decompositions obtained from the graph G.

Proposition 1. *Let $d^*_i = \left(d^*_{i,1}, \ldots, d^*_{i,k}\right)$ be the vector of correct decisions for the hypotheses $H_{i,j}$ of equality of collections of conditional distributions of $X_{R_{i,j}} \mid X_{S_{i,j}}$ in the $i-$th decomposition. Then*

$$D_G = \bigcap_{i=1}^{k} \bigcup_{\left\{j:\, d^*_{i,j}=1\right\}} C_{i,j}.$$

Proof: See Appendix. □

The above proposition gives an oracle procedure for recovering the graphical seed set from the knowledge of the two joint distributions. In practice, we need to rely on statistical tests. Let $\phi_i = (\phi_{i,1}, \ldots, \phi_{i,k}) \in \{0,1\}^k$ be a vector indicating the results of the statistical tests performed in the i-th decomposition, $i \in \{1, \ldots k\}$, with $\phi_{i,j} = 1$ when the hypothesis $H_{i,j}$ is rejected, and $\phi_{i,j} = 0$ otherwise. The following definition naturally follows.

Definition 3 (Graphical seed set estimator)**.** The random set $\hat{D}_G$, defined as

$$\hat{D}_G = \bigcap_{i=1}^{k} \bigcup_{\{j:\, \phi_{i,j}=1\}} C_{i,j} \tag{5}$$

is an estimator of D_G.

3.3. Asymptotic behavior

Estimator $\hat{D}_G$ is different from classical estimators in that its values depend on data through the results of sequences of tests. Properties of the estimator will ultimately depend on the properties of the tests which are used. A treatment of these properties in the limit of infinite data benefits from the introduction of a more general notion of consistency of tests, that we give in general terms as follows (see Definition 1 in [17] for a similar treatment).

Definition 4. A sequence of tests $\phi(n)$ for the hypothesis $H : (\theta^{(1)}, \theta^{(2)}) \in \Theta_0$ vs $H_1 : (\theta^{(1)}, \theta^{(2)}) \in \Theta_1$ is consistent if for each $(\theta^{(1)}, \theta^{(2)}) \in \Theta \times \Theta$ there exists a sequence of significance levels α_n s.t.

(1) for each $(\theta^{(1)}, \theta^{(2)}) \in \Theta_0$, $\lim_{n\to\infty} \mathbb{P}_{(\theta^{(1)},\theta^{(2)})}(\phi(n) = 1) = 0$;

(2) for each $(\theta^{(1)}, \theta^{(2)}) \in \Theta_1$, $\lim_{n\to\infty} \mathbb{P}_{(\theta^{(1)},\theta^{(2)})}(\phi(n) = 0) = 0$.

In other words, a sequence of tests is consistent if, at least asymptotically, it reports a correct decision. Let us now consider testing $H_{i,j}$ in the above given framework. Let $n = n_1 + n_2$ and assume that as $n \to \infty$, $n_l/n \to \gamma_l$ such that $0 < \gamma_l < 1$, $l \in \{1, 2\}$ and $\gamma_1 + \gamma_2 = 1$. Moreover, let the test statistic $\phi_{i,j}(n)$ be defined as

$$\phi_{i,j}(n) = \begin{cases} 0 & \lambda_{i,j;n} < q_n \\ 1 & \lambda_{i,j;n} > q_n \end{cases}$$

where $\lambda_{i,j;n}$ is the log likelihood ratio for $H_{i,j}$ and q_n a suitable sequence of quantiles. Standard results assure that, under the null hypothesis, the sequence $\lambda_{i,j;n}$ converges to a chi-square distribution with f degrees of freedom, where f is the difference between the dimensions of the unrestricted parameter space and the restricted parameter space implied by the hypothesis of equality of the distributions of $X_{R_{i,j}} \mid X_{S_{i,j}}$ in the two groups. Then, the test that rejects the null hypothesis if $\lambda_{i,j;n}$ exceeds the upper α-quantile of the chi-square distribution is asymptotically of level α. We can state the following proposition.

Proposition 2. *In the framework stated above, for each $H_{i,j}$, there exists a sequence of significance levels α_n, s.t. the sequence of tests $\phi_{i,j}(n)$ is consistent.*

Proof: See Appendix. □

Theorem 2. *The estimator $\hat{D}_G$ is a pointwise consistent estimator of D_G, i.e., $\mathbb{P}_{(\theta^{(1)},\theta^{(2)})}(\hat{D}_G = D_G) \to 1$.*

Proof: See Appendix. □

3.4. Finite sample type I error control

With finite samples, it is customary to assign a bound to the probability of incorrectly rejecting the null hypothesis by imposing conditions such as $\mathbb{P}_{(\theta^{(1)},\theta^{(2)})\in\Theta_0}(\phi_{i,j}(n) = 1) \leq \alpha$. Estimation of D_G requires performing a collection of $k + \sum_{i=1}^{k} \nu(C_i)$ tests, where $\nu(C_i)$ denotes the number of separators contained within the clique C_i. Finite sample behavior of $\hat{D}_G$ thus hinges on the proper control of the multiplicity issue.

We focus on the requirement that the probability that $\hat{D}_G$ contains a false positive should be bounded by a given $\alpha \in (0, 1)$, i.e. $\Pr(\exists v \in V : v \in \hat{D}_G \wedge v \notin D_G) \leq \alpha$. But, if there is such a node v, then given Definition 3 of $\hat{D}_G$, necessarily one of the true null hypotheses in the collection of hypotheses $\mathcal{H} = \left\{H_{ij}, i, j \in \{1, \ldots, k\}\right\}$ was erroneously rejected. This implies that the control of familywise error rate for $\mathcal{H}$, i.e. the probability of rejecting at least one true null hypothesis, results in the control of probability of including a false positive in $\hat{D}_G$.

The simplest approach to control the familywise error rate is to apply the Bonferroni correction with a factor of $k + \sum_{i=1}^{k} \nu(C_i)$. However, the Bonferroni correction can be overly conservative when there is high dependence among p-values. This is the case here, since although local test statistics are independent within a single decomposition (see Theorem 1), considering alternative decompositions leads to logical relations among hypotheses and typically results in a high positive dependence between the associated p-values. To address this issue, we employ the max$T$ method of [22], which uses permutations to obtain the joint distribution of the p-values and, by accounting for the dependence among p-values, attenuates the conservativeness of the Bonferroni procedure. In our setting, the condition of subset pivotality is satisfied, and the Westfall and Young procedure controls the familywise error rate in the strong sense.

In many applications, familywise error rate control is considered too stringent and false discovery rate is considered instead. Unfortunately, no such simple relation exists between controlling false discovery rate for $\mathcal{H}$ and the inclusion of false positives in $\hat{D}_G$. In other words, it is unclear how controlling false discovery rate for $\mathcal{H}$ translates to the type I error guarantees for $\hat{D}_G$. For this reason, we restricted our attention to the familywise error rate.

4. Simulation studies

4.1. Simulation study 1

To study the finite sample behavior of $\hat{D}_G$, we considered a randomly generated graph G consisting of 100 nodes grouped in 37 cliques (the largest clique containing 15 nodes). The code to reproduce all numerical experiments, as well as real data analysis featured in Section 5, is available at `https://github.com/veradjordjilovic/Seed-set`. A plot of the graph is shown in Fig. 5 in Supplementary Material. The minimal seed set was set to $D = \{2, 5\}$. In the chosen graph, the graphical seed set does not coincide with the minimal seed set since there is no separator in G that separates a node number 17 from D. We thus have $D_G = \{2, 5, 17\}$.

We will work in the Gaussian setting. We set the parameters of the first, i.e. control, condition in the following way. The means of 100 variables were drawn randomly from a normal distribution centered at 0.5 (standard deviation 1). The covariance matrix was obtained by starting from a matrix with all off-diagonal elements equal to 0.4 and all diagonal elements equal to 1 and modifying it so that its inverse has zeros corresponding to the missing edges of G. For the second or the perturbed condition, we considered perturbations that alter the means of the two seed set variables linearly. In particular, the means were multiplied by λ that varied in the range $\{1.2, 1.25, \ldots, 1.6, 1.65\}$. The variance of seed set variables was also manipulated and decreased by 50%. We held the sample size fixed and equal for the two conditions: $n_1 = n_2 = 50$. For each λ, we generated 1000 pairs of samples.

Note that this perturbation affecting X_2 and X_5, indirectly affected all the marginal distributions of $(X_1, \ldots, X_{100})^\top$. For an illustration of this effect, see Fig. 6, Supplementary Material, that compares the parameters associated to the first ten variables, i.e., $X_1, \ldots, X_{10}$, in the first and in the second condition for $\lambda = 1.7$.

We computed $\hat{D}_G$ with the `SourceSet` R package, which implements the proposed approach (available from CRAN). The familywise error rate was controlled at 5% by the step-down max$T$ method [22]. To evaluate the performance of our procedure, we computed the empirical power, defined as the frequency with which the estimated graphical seed set $\hat{D}_G$ coincided with the true graphical seed set D_G, and the empirical familywise error rate, defined as the frequency with which $\hat{D}_G$ contained a false positive. The results are shown in Fig. 1.

Results show that the familywise error rate is controlled at the nominal level for all values λ, which is in line with finite sample theoretical type I error guarantees described in Section 3.4. With regards to power, for the lowest level of perturbation $\lambda = 1.2$, corresponding to an increase of 20% in variables X_2 and X_5, we see that the power to identify

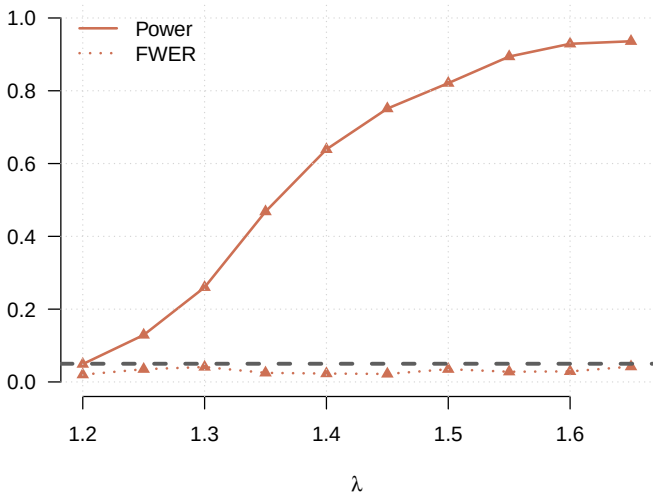

Fig. 1: Simulation study 1: Empirical power and familywise error rate of the graphical seed set estimating procedure as a function of perturbation strength λ. Dashed horizontal line $y = 0.05$, representing the nominal familywise error rate, was added for reference.

D_G is very low. With increasing λ, the power is fast increasing and reaches 80% already for $\lambda = 1.5$. Note that given our definition of power, the maximum attainable power is bounded by the complement of the familywise error rate, i.e. $1 - \mathrm{P}(\exists v \in V : v \in \hat{D}_G \wedge v \notin D_G) \approx 1 - \alpha$, rather than 1.

Unbalanced sample sizes. We further studied the impact an unbalanced sample size can have on the performance of the seed set estimating procedure. To this end, we fixed parameters of the perturbed condition by setting $\lambda = 1.3$ and then varied the sample size of the pooled sample $n = n_1 + n_2$ in the set $\{75, 100, 125, 150, 200, 250, 300, 350\}$. We computed the empirical power and familywise error rate in two scenarios featuring:

- balanced samples: $n_1 = n_2$ when n is even, or $n_1 = \lfloor n/2 \rfloor$ and $n_2 = n_1 + 1$, when n is odd;
- unbalanced samples: $n_1 = 50$ and $n_2 = n - n_1$.

Results, shown in Fig. 2, indicate that the familywise error rate is controlled well in both scenarios. With regards to power, when the total sample size is small, the two scenarios are comparable. With increasing sample size, the difference between n_1 and n_2 is also increasing, and the power in the scenario with balanced samples is higher, but the advantage does not seem to be very large.

Robustness to non-normality. An important issue arising in practical applications is the sensitivity of the procedure to the presence of departures from normality. To investigate this issue, we have considered data generated from skew-normal graphical models [3] and studied the power and familywise error rate as a function of skewness. The results of this simulation study, described in Section 3.1, Supplementary material, suggest that when compared to a setting with normal data, the power does not seem to be much affected, while the familywise error rate increases and possibly surpasses the pre-specified level α. Nevertheless, the increase seems to be small enough as to allow us to conclude that the procedure is quite robust to this particular violation of normality.

Competing methods. To the best of our knowledge, no alternative methods aiming to estimate D_G, i.e. the origin of the perturbation affecting both the means and the (co)variances are currently available. However, some recent approaches focus on detecting more specific forms of perturbations: either those affecting exclusively the graphical structure or the vector of means. In the following section, we report the comparison with a method addressing the former, while in Section 3.2, Supplementary material, we provide a comparison with a method addressing the latter.

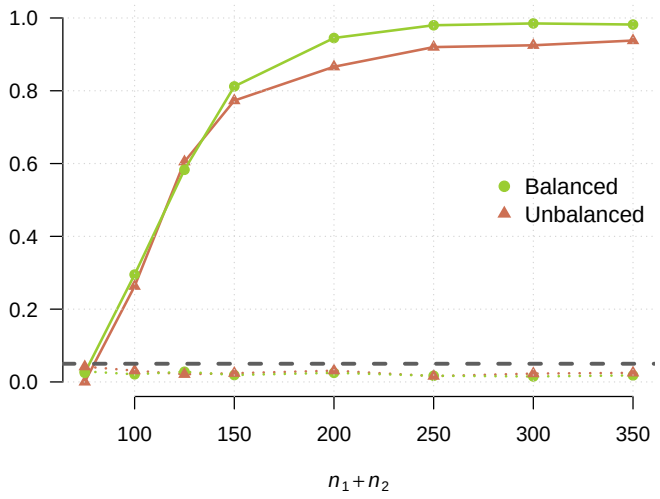

Fig. 2: Simulation study 1: Empirical power and familywise error rate of the graphical seed set estimating procedure as a function of the pooled sample size $n = n_1 + n_2$. In an unbalanced sampling scheme $n_1 = 50$ was fixed, while in a balanced sampling scheme $n_1 = n_2$ if $n_1 + n_2$ was even, and $|n_1 - n_2| = 1$ otherwise. Dashed horizontal line $y = 0.05$, representing the nominal familywise error rate, was added for reference.

4.2. Simulation study 2

To study the behavior of our procedure when the difference between two conditions is driven only by the graphical structure, we considered a small graph consisting of 10 nodes, shown in Fig. 3. The edge between nodes 4 and 6 is present in Condition 1, but absent in Condition 2, i.e., in Condition 2, variables associated to nodes 4 and 6 are conditionally independent given the rest. It is worth noting that, in Condition 2, the graph is not decomposable and that the graphical structure to be used in estimating D is that of condition 1, as it represents the decomposable model common to the two conditions. The minimal seed set is now $D = \{4, 6\}$, and it coincides with the graphical seed set.

Means of the 10 variables were randomly drawn from a normal distribution centered at 0.5 (standard deviation 1) and were the same for Conditions 1 and 2. In each condition, the covariance matrix was obtained from a matrix with all diagonal elements equal to 1 and all off-diagonal elements equal to 0.6, that was modified so that the zero pattern of its inverse corresponds to the missing edges of G. Three different sample sizes were considered, i.e., $n = 200, 300, 500$.

Results, averaged over 500 Monte Carlo runs, are shown in Table 1, where rows labeled 'Seed set' report the percentage of times each node was found to belong to D. Results show that, in this setting, the power, although limited at the smallest sample size, is increasing with increasing sample sizes. This is understandable, since, differently from simulation 1, the difference between the two conditions is relatively sparse, and the smaller this difference, the harder it is to distinguish between the null and the alternative hypothesis.

It is interesting noting that methods for differential networks, such as those in [24] and [23], could also have been used in this setting. For an appreciation of the different results produced by different approaches, we considered the method of [24], for which an implementation is available. The method focuses only on the structure of the covariance; it uses no external information on such structure and it has been developed around estimation consistency. It follows that this method is not directly comparable with our method, and its relative performance is to be interpreted with caution.

The implementation of the differential network method was obtained from the github account of the corresponding author of [24]. Cross validation and L_∞ were chosen as tuning criteria. The output of this method is an estimate of the difference between two precision matrices. To facilitate comparison with our method, we focused on the differential network given by a subset of non zero elements of the estimated difference. A variable was deemed important if the associated node belonged to the estimated differential network, i.e. if at least one edge of the differential network

featured the node in question. In this case, the true differential network consists of a single edge joining nodes 4 and 6. Variables deemed important by this method should thus coincide with the minimal seed set.

Rows labeled 'Differential network' in Table 1, report the percentage of times a variable belonged to the set of important variables according to the differential network method. The method flags nodes 4 and 6 to be relevant also for the smallest sample size (around 85% of times for $n = 200$). However, the rate of a false discovery is much higher, around 40% across the remaining nodes, and does not seem to be decreasing with increasing sample size. Note that this is not in conflict with the consistency of the estimator of [24], since the estimated non-zero elements are getting smaller in absolute value (results not reported here) and converge to zero with increasing sample size.

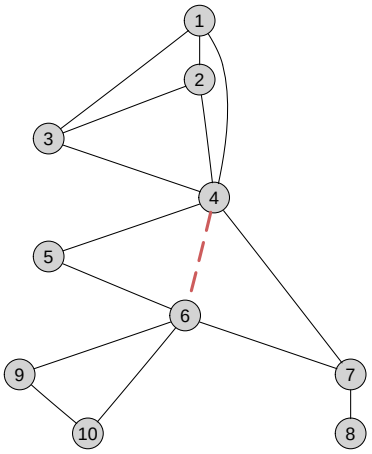

Fig. 3: An undirected graph used in Simulation study 2. Edge (4,6) is present in condition 1, and absent in condition 2.

Table 1: Simulation study 2: percentage of times (%) a node is found to belong to D or a differential network. Monte Carlo standard error of estimates is bounded by 2.2%.

		Node									
		1	2	3	4	5	6	7	8	9	10
$n = 200$	Seed set	1	1	1	22	1	25	3	1	1	1
	Differential network	34	39	40	86	44	85	51	35	40	37
$n = 300$	Seed set	1	1	1	46	1	47	0	0	1	1
	Differential network	39	37	37	93	51	94	50	36	44	42
$n = 500$	Seed set	2	2	2	86	2	86	2	1	0	0
	Differential network	42	42	40	99	50	99	56	36	46	46

5. Biological validation

Genes and gene products cluster into functionally connected pathways, i.e. networks of biological interactions that describe their basic dynamics [12]. A large literature has developed around the problem of detecting statistically significant dysregulations of pathways in different experimental conditions [9, 11, 21], but translating detected dysregulations into claims about their origin is a challenging task. Chromosomal rearrangements offer a possible explanation. Chromosome rearrangements initiate various alterations of the regulation of gene expression through a variety of different mechanisms. For this reason, when comparing populations with and without a given gene rearrangement, sound inferential tools usually flag most pathways including genes with the rearrangement as statistically

different. What we should expect from tools calibrated to detect the source of dysregulation is that they go as close as possible to the rearranged genes. This is the reason why we consider known chromosomal rearrangements as ideal case studies to explore the power of our procedure on real, complex and noisy data.

As an example, consider the BCR/ABL fusion gene, formed by rearrangement of the breakpoint cluster region (BCR) on chromosome 22 with the c-ABL proto-oncogene on chromosome 9. This rearrangement has been postulated to be responsible for the development of leukemia and is present in all chronic myelogenous leukemia patients. It is also identified in some cases of acute lymphocytic leukemia (ALL), in which it is associated with poor prognosis.

We consider a well-known dataset [4] available from an R package `ALL`[14]. Data refer to gene expression signatures of two groups of ALL patients: a first group of 37 subjects with BCR/ABL gene rearrangement, and a second group of 41 subjects without the BCR/ABL gene rearrangement. In what follows, we will consider the Chronic myeloid leukemia pathway, shown in Fig. 7 in Supplementary material, a pathway whose functioning is highly impacted by BCR and ABL genes.

To derive the underlying undirected graph, we used the R package `graphite` [18], which transforms KEGG pathways into graph objects. We moralized and triangulated this graph to obtain a decomposable graph. For graph operations, we relied on the package `gRbase` [7]. The obtained graph consists of three connected components, and for illustration purposes, we restricted our attention to the largest connected component, consisting of 27 nodes and 16 cliques, shown in Fig. 4 (colors can be ignored for now). The number associated to each node is a unique gene identifier from the Entrez Gene database at the National Center for Biotechnology Information [15]. Note that nodes 25 and 613 represent, ABL and BCR genes, respectively.

The global hypothesis of equality of distributions in the two groups is rejected by the likelihood ratio test (p-value $= 2.06 \times 10^{-11}$). To estimate $\hat{D}_G$, we decomposed the graph into a succession of cliques. There are 16 cliques, and thus 16 decompositions of the global null hypothesis, and 41 unique local hypotheses. We controlled the familywise error rate at 5% level by the min$P$ method with $B = 1640$ permutations (the minimal number recommended by the `SourceSet` package). We have thus relied on permutation, rather than asymptotic p-values. Obtained p-values are shown in Table 2, Supplementary material. The threshold found by min$P$ method was 2.4×10^{-3}. The resulting estimate is represented in Fig. 4. Highlighted nodes (either gray or red) belong to cliques that result significantly different in two conditions, while the red nodes form the estimated graphical seed set $\hat{D}_G = \{25, 613, 6776\}$. These three genes, thus, explain the marked difference between the two groups, but their effect does not seem to propagate towards other genes in the network (the majority of white nodes in Fig. 4).

6. Discussion

Two sample testing problem we consider is closely related to the problem of variable selection in a logistic regression. When a predictor is a p-dimensional random vector X and the output is a class label (1 or 2), the minimal seed set coincides with the Markov blanket of the response.

Modularity of graphical modes is usually considered with regards to density factorization or parameter estimation. Theorem 1 mirrors this property in the hypothesis testing setting within the framework of strong meta Markov models, and although conceptually simple, we were unable to find this result in the literature. The strong meta Markov assumption is a strong assumption, however, the two families most often encountered in practical applications, that of Gaussian graphical models and graphical log-linear models, fall within this framework.

The presented approach estimates the graphical seed set which might be larger than the minimal seed set. An open question regards a potential two-step procedure, in which clique grained decomposition is followed by additional tests aiming at identifying $\hat{D} \subseteq \hat{D}_G$. Statistical properties of such a procedure are far from trivial, and we leave this question for future research.

Our approach is based on the assumption that the graphical structure is known, either derived from relevant subject matter considerations or estimated from previous studies. When this is not the case, finding ways to combine learning of the graphical structure with the presented approach in an efficient way, while controlling the desired error rate, represents a methodological challenge that awaits further research.

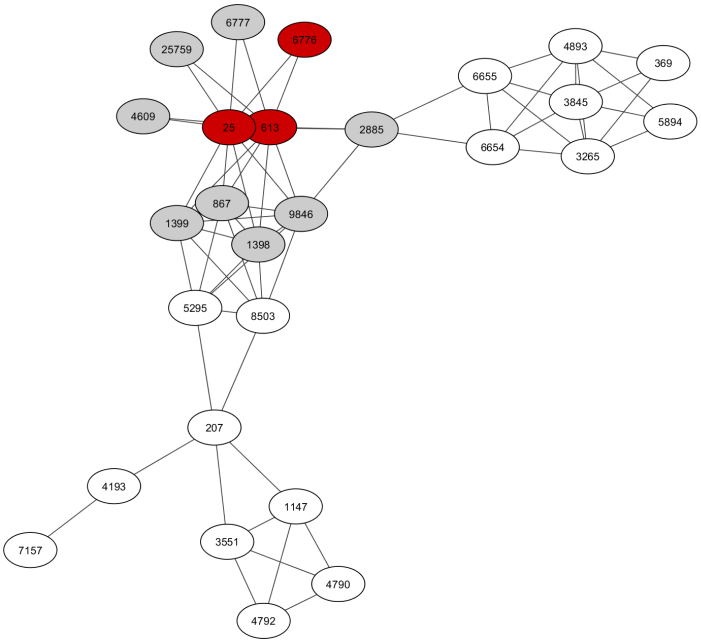

Fig. 4: An undirected graph representing the Chronic myeloid leukemia pathway. Genes belonging to cliques for which the hypothesis of equality of distributions is rejected are highlighted. Genes belonging to the estimated graphical seed set are colored red.

Acknowledgments

We thank the Editor, Associate Editor and the Referees for useful comments and suggestions on an earlier version of the manuscript, which led to this improved version. Insight and expertise that greatly contributed to the development of this paper and the research behind was generously received by a number of colleagues, most notably Chiara Romualdi, Maria Sofia Massa and Elisa Salviato.

Appendix: Proofs

Proof of Proposition 1: Let $P = \bigcap_{i=1}^{k} \bigcup_{\{j:\, d^*_{i,j}=1\}} C_{i,j}$. Then if $v \in P$, for each decomposition i, there is at least one clique $C_{i,j}$ containing v such that $d^*_{ij} = 1$. If $C_{i,l}$ denotes the first clique in the i-th decomposition containing v, we know that v belongs to $R_{i,l}$, otherwise $C_{i,l}$ would not be the first clique containing v. Consider a tree of cliques constructed from the perfect ordering $C_{i,1}, \ldots, C_{i,k}$ in the following fashion. The perfect ordering property guarantees that for each clique $C_{i,j}$, the intersection with the union of predecessor cliques is contained within a single clique, that is

$$C_{i,j} \cap \bigcup_{m=1}^{j-1} C_{i,m} \subset C_{i,n}, \quad \text{for some } n \in \{1, \ldots, j-1\}. \tag{6}$$

Then the set $C_{i,n}$ is set as a parent of $C_{i,j}$ in the clique tree. Parent clique might not be unique, but without loss of generality, we take the first clique satisfying the assumption (6). Then all cliques containing v other than $C_{i,l}$ must be descendants of $C_{i,l}$. We further notice that if $d^*_{i,l} = 0$, then $d^*_{i,m} = 0$ for all its descendants. This implies that necessarily $d^*_{i,l} = 1$ and $S_{i,l}$ does not separate v from D. Since this is true for all decompositions, there can be no separator that separates v from D, implying that v belongs to D_G.

We have proven $v \in P \Rightarrow v \in D_G$, but all considered implications remain valid if reversed, so that $v \in P \Leftrightarrow v \in D_G$. □

Proof of Proposition 2: Choose $\alpha_n = (1 - F_U(n^d))$, with $0 < d < 1/2$, $U \sim \chi^2_f$, and let $q_n = F_U^{-1}(\alpha_n)$. Under the null hypothesis, $\lambda_{i,j;n} \xrightarrow{d} \lambda$, with $\lambda \sim \chi^2_f$. Thanks to the Slutsky theorem, we can write

$$\mathbb{P}_{(\theta^{(1)},\theta^{(2)})\in\Theta_0}(\phi_{i,j}(n) = 1) = \mathbb{P}_{(\theta^{(1)},\theta^{(2)})\in\Theta_0}\left(\frac{\lambda_{i,j;n}}{n^d} > 1\right) \longrightarrow 0.$$

Furthermore, for each $(\theta^{(1)}, \theta^{(2)}) \in \Theta_1$, it is known that the log likelihood ratio test is degenerate with the order $O(\sqrt{n})$. With the choice of α_n above,

$$\mathbb{P}_{(\theta^{(1)},\theta^{(2)})\in\Theta_1}(\phi_{i,j}(n) = 0) = \mathbb{P}_{(\theta^{(1)},\theta^{(2)})\in\Theta_1}\left(\frac{\lambda_{i,j;n}}{n^d} < 1\right) \longrightarrow 0.$$

□

Proof of Theorem 2: For a fixed i, we have that $\phi_i(n) = (\phi_{i,1}(n), \ldots, \phi_{i,k}(n)) \to d_i^* = (d_{i,1}^*, \ldots, d_{i,k}^*)$, since the inequality

$$\mathbb{P}_{(\theta^{(1)},\theta^{(2)})}(\phi_i(n) = d_i^*) \geq 1 - \sum_{j=1}^{k} \mathbb{P}_{(\theta^{(1)},\theta^{(2)})}(\phi_{i,j}(n) \neq d_{i,j}^*)$$

in conjunction with Proposition 2 implies $\mathbb{P}_{(\theta^{(1)},\theta^{(2)})}(\phi_i(n) = d_i^*) \longrightarrow 1$. Convergence of $\hat{D}_G$ to D_G follows straightforwardly. □

References

[1] T. W. Anderson, An Introduction to Multivariate Statistical Analysis, Wiley, New Jersey, 2003.
[2] O. Barndorff-Nielsen, Information and Exponential Families in Statistical Theory, John Wiley & Sons, New York., 2014.
[3] A. Capitanio, A. Azzalini, E. Stanghellini, Graphical models for skew-normal variates, Scandinavian Journal of Statistics 30 (2003) 129–144.
[4] S. Chiaretti, X. Li, R. Gentleman, A. Vitale, K. S. Wang, F. Mandelli, R. Foa, J. Ritz, Gene expression profiles of B-lineage adult acute lymphocytic leukemia reveal genetic patterns that identify lineage derivation and distinct mechanisms of transformation, Clinical Cancer Research 11 (2005) 7209–7219.
[5] A. Dawid, S. Lauritzen, Hyper Markov laws in the statistical analysis of decomposable graphical models, The Annals of Statistics 21 (1993) 1272–1317.
[6] A. Del Sol, R. Balling, L. Hood, D. Galas, Diseases as network perturbations, Current Opinion in Biotechnology 21 (2010) 566–571.
[7] C. Dethlefsen, S. Højsgaard, A common platform for graphical models in R: The gRbase package, Journal of Statistical Software 14 (2005) 1–12.
[8] M. Frydenberg, S. L. Lauritzen, Decomposition of maximum likelihood in mixed graphical interaction models, Biometrika 76 (1989) 539–555.
[9] J. J. Goeman, S. A. Van De Geer, F. De Kort, H. C. Van Houwelingen, A global test for groups of genes: testing association with a clinical outcome, Bioinformatics 20 (2004) 93–99.
[10] N. J. Hudson, A. Reverter, B. P. Dalrymple, A differential wiring analysis of expression data correctly identifies the gene containing the causal mutation, PLoS Comput Biol 5 (2009) e1000382.
[11] M. Hummel, R. Meister, U. Mansmann, GlobalANCOVA: exploration and assessment of gene group effects, Bioinformatics 24 (2008) 78–85.
[12] M. Kanehisa, S. Goto, KEGG: Kyoto Encyclopedia of Genes and Genomes, Nucleic Acids Research 28 (2000) 27–30.
[13] S. L. Lauritzen, Graphical models, Clarendon Press, Oxford, 1996.
[14] X. Li, ALL: A data package, 2009. R package version 1.16.0.
[15] D. Maglott, J. Ostell, K. D. Pruitt, T. Tatusova, Entrez Gene: gene-centered information at NCBI, Nucleic Acids Research 33 (2005) D54–D58.
[16] M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth, limma powers differential expression analyses for rna-sequencing and microarray studies, Nucleic acids research 43 (2015) e47–e47.
[17] J. M. Robins, R. Scheines, P. Spirtes, L. Wasserman, Uniform consistency in causal inference, Biometrika 90 (2003) 491–515.
[18] G. Sales, E. Calura, C. Romualdi, graphite: GRAPH Interaction from pathway Topological Environment, 2016. R package version 1.20.1.
[19] W. Tan, On the distribution of quadratic forms in normal random variables, Canadian Journal of Statistics 5 (1977) 241–250.
[20] A. Thomas, P. J. Green, Enumerating the junction trees of a decomposable graph, Journal of Computational and Graphical Statistics 18 (2009) 930–940.
[21] C.-A. Tsai, J. J. Chen, Multivariate analysis of variance test for gene set analysis, Bioinformatics 25 (2009) 897–903.
[22] P. H. Westfall, S. S. Young, Resampling-Based Multiple Testing: Examples and methods for p-value adjustment, volume 279, John Wiley & Sons, New York., 1993.
[23] Y. Xia, T. Cai, T. T. Cai, Testing differential networks with applications to the detection of gene-gene interactions, Biometrika 102 (2015) 247–266.
[24] S. D. Zhao, T. T. Cai, H. Li, Direct estimation of differential networks, Biometrika 101 (2014) 253–268.
[25] Y. Zhu, J. Bradic, Two-sample testing in non-sparse high-dimensional linear models, arXiv preprint arXiv:1610.04580 (2016).