University of California, Berkeley
U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2009 *Paper* 245

Joint Multiple Testing Procedures for Graphical Model Selection with Applications to Biological Networks

Houston N. Gilbert* Mark J. van der Laan†
Sandrine Dudoit‡

*Division of Biostatistics, University of California - Berkeley, houston@stat.berkeley.edu

†University of California - Berkeley, laan@berkeley.edu

‡Division of Biostatistics and Department of Statistics, University of California, Berkeley, sandrine@stat.berkeley.edu

Joint Multiple Testing Procedures for Graphical Model Selection with Applications to Biological Networks

Houston N. Gilbert, Mark J. van der Laan, and Sandrine Dudoit

Abstract

Gaussian graphical models have become popular tools for identifying relationships between genes when analyzing microarray expression data. In the classical undirected Gaussian graphical model setting, conditional independence relationships can be inferred from partial correlations obtained from the concentration matrix (= inverse covariance matrix) when the sample size n exceeds the number of parameters p which need to estimated. In situations where $n < p$, another approach to graphical model estimation may rely on calculating unconditional (zero-order) and first-order partial correlations. In these settings, the goal is to identify a lower-order conditional independence graph, sometimes referred to as a '0-1 graphs'. For either choice of graph, model selection may involve a multiple testing problem, in which edges in a graph are drawn only after rejecting hypotheses involving (saturated or lower-order) partial correlation parameters. Most multiple testing procedures applied in previously proposed graphical model selection algorithms rely on standard, marginal testing methods which do not take into account the joint distribution of the test statistics derived from (partial) correlations. We propose and implement a multiple testing framework useful when testing for edge inclusion during graphical model selection. Two features of our methodology include (i) a computationally efficient and asymptotically valid test statistics joint null distribution derived from influence curves for correlation-based parameters, and (ii) the application of empirical Bayes joint multiple testing procedures which can effectively control a variety of popular Type I error rates by incorpo- rating joint null distributions such as those described here (Dudoit and van der Laan, 2008). Using a dataset from Arabidopsis thaliana, we observe that the use of more sophisticated, modular approaches to multiple testing allows one to identify

greater numbers of edges when approximating an undirected graphical model using a 0-1 graph. Our framework may also be extended to edge testing algorithms for other types of graphical models (e.g., for classical undirected, bidirected, and directed acyclic graphs).

1 Introduction

With the advent of high-throughput biological assay technology, a common and broadly-defined analysis goal has been *network identification*. Given a set of biological measurements (e.g., microarray expression data), one might wish to infer sets of possibly interacting variables (genes). Clustering methods and naïve pairwise zero-order correlation approaches represent some of the earliest applications of statistical methods to this problem in the functional genomics literature (see e.g., Golub et al. (1999), Butte et al. (2000), and numerous other publications). Such an approach typically relies on the assumption that coexpression among (sets of) genes infers some type of underlying coregulation mechanism.

The last several years has literally seen an explosion of other methods which seek to more accurately infer gene-gene associations from a collection of biological data. Gaussian graphical models (Whittaker, 1990; Cox and Wermuth, 1993, 1996; Lauritzen, 1996; Edwards, 2000) have been particularly popular in this respect, as they are well-suited for estimating more complicated relationships among genes from continuous expression data (see e.g., Butte et al., 2000; Friedman et al., 2000; Waddell and Kishino, 2000a,b; Hartemink et al., 2001; Toh and Horimoto, 2002a,b; Husmeier, 2003; Wang et al., 2003; Wu et al., 2003; Friedman, 2004; Matsuno et al., 2006; Aburatani et al., 2007; Drton and Perlman, 2007; Ma et al., 2007; Ma and Bohnert, 2008, and other work cited below). (For work in which discrete genotype, i.e., single-nucleotide polymorphism (SNP) information was also used for network reconstruction, the reader is referred to Zhu et al. (2004), Bing and Hoeschele (2005), Chessler et al. (2005), and Lee et al. (2006), among others.) In particular, rather than separately examining marginal correlations among pairs of variables (the naïve approach), correlation-based associations obtained from several of these newer methods represent relationships between two (sets of) variables *conditional on one or more other variables*.

The graphical model (defined below) dictates the types of independence relationships that exist between two or more (sets of) variables. In many applications, because the graph is not given and itself must be inferred from the data, the genetic network identification problem becomes one of *graphical model selection* (Wong et al., 2003; Drton and Perlman, 2004; Yuan and Lin, 2007; Drton and Perlman, 2008). Depending on the type of graph (e.g., bidirected, undirected, directed acyclic, see below), model selection typically relies on somehow estimating the (inverse) variance-covariance matrix from which the (conditional) independence relationships can be inferred via the resulting estimators of the (partial) correlation coefficients (N.B., parentheses in previous statement match). A common choice of graphical model is the undirected graphical model (sometimes called the *covariance selection* or *concentration graph* model). In this model, the *saturated* partial correlations between two variables obtained when conditioning on all other variables may be calculated from the inverse-covariance matrix (i.e., concentration) matrix.

If two variables are conditionally independent given all other variables, then the saturated partial correlation will be zero, and no edge will be drawn which connects them in the graph.

As in most statistical endeavors, the model selection problem is confronted with a tradeoff between power and Type I error. That is, one wishes to identify as many true associations as possible while still minimizing the number of inferred spurious relationships (i.e., incorrect edges) drawn in a graph. To ensure proper estimation capability, much of the work which utilizes direct, traditional applications of Gaussian graphical model techniques in biology focuses on situations in which the number of variables is smaller than the sample size. Given the obvious limitations of such an approach in a genomic setting, other researchers have sought means around the '$n < p$' problem. Some have applied sparsity restrictions and shrinkage estimation approaches resulting in regularized graphical models (Dobra et al., 2004; Schäfer and Strimmer, 2004, 2005a,b; Meinshausen and Bühlmann, 2006; Li and Gui, 2006). Others have focused on smaller portions of the data to examine limited or *lower-order partial correlations* for the purposes of model selection (de la Fuente et al., 2004; Magwene and Kim, 2004; Wille et al., 2004; Castelo and Roverato, 2006; Wille and Bühlmann, 2006).

Unlike the saturated partial correlations used in the undirected Gaussian graphical model, lower-order partial correlations do not require a full conditioning set containing all other variables in order to be estimated. As such, partial correlations obtained from conditioning on just a few variables can be readily estimated with modest sample sizes. The *lower-order conditional independence graph* is most likely a hybrid of the graphs obtained using naïve correlation approaches (e.g., as in "relevance networks", (Butte et al., 2000)) and the one estimated from the true undirected Gaussian graphical model (the concentration graph). In the case where only zero-order (unconditional) and first-order partial correlations are considered for the purposes of model selection, the resulting lower-order conditional independence graph is sometimes called a '0-1' graph (de Campos and Huete, 2000; Wille and Bühlmann, 2006). Wille and Bühlmann (2006), however, have shown that the 0-1 graph can serve as a good approximation to the concentration graph with edges inferred from saturated partial correlations.

Regardless of one's selection of type of graph or even of one's underlying approach to model selection, we shift our attention to another issue which may arise in network reconstruction algorithms. For many graphical model selection procedures, making inferences about the (saturated or lower-order partial) correlations for the purposes of edge inclusion into the graph can be formulated as a multiple testing problem (de la Fuente et al., 2004; Magwene and Kim, 2004; Wille et al., 2004; Schäfer and Strimmer, 2005b; Wille and Bühlmann, 2006; Drton and Perlman, 2007). In many instances, quick, easy-to-implement multiple testing procedures (MTPs) are employed at this step in the analysis for hypotheses concerning correlation parameters, e.g., the Benjamini-Hochberg procedure for control of the false discovery rate, (FDR;

Benjamini and Hochberg, 1995). Often, these procedures are *marginal multiple testing procedures*, that is, MTPs which only rely on information about the marginal distributions of the test statistics being used to make inference decisions. We recall, however, that the test statistics obtained from a collection of estimated (partial) correlation coefficients may themselves exhibit some degree of correlation. By exploiting the dependencies among test statistics, we may obtain more powerful multiple testing results, which in turn would generate richer, less-sparse graphical models.

We propose a framework useful for testing hypotheses concerning a rich collection of correlation parameters for the purposes of graphical model selection. Central features of our work include (i) the formulation of an asymptotically valid correlation coefficient test statistics joint null distribution based on influence curves, and (ii) an empirical Bayes joint multiple testing framework allowing one to powerfully control generalized Type I error rates (Dudoit and van der Laan, 2008; Dudoit et al., 2008). We focus our attention on the lower-order correlation approaches taken in the literature (de la Fuente et al., 2004; Magwene and Kim, 2004; Wille et al., 2004; Wille and Bühlmann, 2006), but note that our methods could also easily accomodate or be combined with, for example, the shrinkage approaches of Schäfer and Strimmer (2005b) or the concentration graph work of Drton and Perlman (2007).

The next section reviews some relevant aspects of graphical models and establishes the connection between model selection and multiple testing in the graphical model context. Section 3 introduces our methodological contribution by describing both an appropriate test statistics joint null distribution for use with correlation parameters as well as empirical Bayes joint MTPs. Finally, our methods will be highlighted with an application to an *Arabidopsis thaliana* dataset. Specifically, we will attempt to approximate an undirected graph by identifying edges in a lower-order conditional independence graph (0-1 graph). We see evidence that utilizing a proper estimate of the test statistics joint null distribution in combination with more sophisticated joint MTPs can result in the inclusion of more edges in the lower-order conditional independence genetic network. Moreover, the use of null distributions derived from influence curves can reduce the computational burden associated with most (resampling-based) joint multiple testing methods, making their use even more attractive in this setting.

2 Graphical Models

2.1 Statistical Model and Parameters of Interest

Let $\mathcal{X}_n \equiv \{X_i : i = 1, \ldots, n\}$ denote a *simple random sample* of n independent and identically distributed (IID) random variables from a *data generating distribution* P, i.e. $X_i \sim P$, $i = 1, \ldots, n$. In large-scale biological problems, the data may represent J-dimensional random vectors, or J-vectors,

$X = (X(j) : j = 1, \ldots, J) \sim P \in \mathbb{R}^J$, where the elements $X(j)$ may correspond to data collected on individual covariates such as gene-specific microarray expression measures and where P now specifies their typically unknown *joint data generating distribution*. Suppose that P is an element of a particular *statistical model* $\mathcal{M}$, i.e., a set of possibly nonparametric distributions. In addition, we may define the sets $V \subseteq \{1, \ldots, J\}$ and $\mathcal{K} \subseteq \{1, \ldots, J\}$, whose elements denote the indices of variables comprising the data vectors collected in $\mathcal{X}_n$. Finally, we let $X(\mathcal{K}) = (X(k) : k \in \mathcal{K})$.

When the data $\mathcal{X}_n$ are assumed to be normally distributed, a number of potentially useful parameters of interest become available for consideration. In this case, correlation coefficients may be used to test hypotheses involving statements about independence. If the correlation between two variables is zero, then those variables behave independently from each other. Allowing $\Sigma(P) = \sigma = (\sigma(j, j') : j, j' = 1, \ldots, J)$ to denote the $J \times J$ covariance matrix, we define the correlation coefficient parameter $\text{Cor}[X(j), X(j')] \equiv \rho(j, j')$ as

$$\rho(j, j') = \frac{\sigma(j, j')}{\sqrt{\sigma(j, j)}\sqrt{\sigma(j', j')}}. \tag{1}$$

Given a suitable estimator $\hat{\Sigma}(P_n) = \sigma_n = (\sigma_n(j, j') : j, j' = 1, \ldots, J)$ of the $J \times J$ covariance matrix using the empirical distribution P_n, a common method of estimating the amount of correlation between two variables is through the *empirical correlation coefficients*. These parameter estimates are given by

$$\rho_n(j, j') = \frac{\sigma_n(j, j')}{\sqrt{\sigma_n(j, j)}\sqrt{\sigma_n(j', j')}}. \tag{2}$$

Another parameter of interest may be the correlation between two variables $X(j)$ and $X(j')$ in their conditional distribution given all other measured variables, that is $\text{Cor}[X(j), X(j')|X(V\backslash\{j, j'\})] \equiv \rho(j, j'|V\backslash\{j, j'\})$. Conditional correlation parameters are generally referred to as *partial correlations*. For clarity, we will refer to partial correlations as *saturated* when the conditioning set contains all other variables $X(V\backslash\{j, j'\})$ as written above.

Letting $\Sigma^{-1}(P) = \sigma' = (\sigma'(j, j') : j, j' = 1, \ldots, J)$ denote the $J \times J$ inverse-covariance (i.e., *precision* or *concentration*) matrix, saturated partial correlation parameters (cf. Lauritzen, 1996, p. 130, cited in Drton and Perlman (2007)) can be shown to equal

$$\rho(j, j'|V\backslash\{j, j'\}) = \frac{-\sigma'(j, j')}{\sqrt{\sigma'(j, j)}\sqrt{\sigma'(j', j')}}. \tag{3}$$

Given a suitable estimator $\hat{\Sigma}^{-1}(P_n) = \sigma'_n = (\sigma'_n(j, j') : j, j' = 1, \ldots, J)$ of the $J \times J$ inverse-covariance matrix using the empirical distribution P_n, a common method of estimating the amount of correlation between two variables conditional on all other variables indexed by $V\backslash\{j, j'\}$ is through the *empirical*

saturated partial correlation coefficients. These parameter estimates are given by

$$\rho_n(j, j'|V\backslash\{j, j'\}) = \frac{-\sigma'_n(j, j')}{\sqrt{\sigma'_n(j, j)}\sqrt{\sigma'_n(j', j')}}. \quad (4)$$

Interesting to note in this formulation of saturated partial correlation is that the parameter only implicitly depends on the variables represented in $V\backslash\{j, j'\}$ through the matrix inversion operation. A related observation is that standard estimation of these partial correlation coefficients also requires an invertible, positive definite sample variance-covariance matrix of full rank. Saturated partial correlation coefficient estimation is therefore best restricted to cases where the number of observations exceeds the number of parameters to be estimated, a condition often referred to in the literature as the '$n > p$' setting.

As alluded to above, partial correlations need not necessarily be saturated. Other possible parameters of interest, *lower-order partial correlations* are defined as the correlation between two variables $X(j)$ and $X(j')$ when conditioning on some subset of the variables indexed in $V\backslash\{j, j'\}$. Typically, lower-order partial correlations restrict the cardinality of the conditioning set to a much smaller number, i.e., zero, one, or two. In the case where the conditioning set is the empty set $\emptyset$, one recovers the familiar correlation coefficients of Equation (1). These unconditional parameters are sometimes equivalently called *zero-order partial correlations*. When conditioning on just one variable $X(k)$, $k \in V\backslash\{j, j'\}$, one may refer to the parameter as a *first-order partial correlation*, and so on. The benefit of using lower-order partial correlations is that they can be more accurately estimated from data with even modest sample sizes such as those available in most biological studies.

Lower-order partial correlations can be derived from the 'bottom-up' either through regression techniques or via a recursion using correlation coefficients of the preceding order. Specifically, in the case of the first-order partial correlation $\mathrm{Cor}[X(j), X(j')|X(k)] \equiv \rho(j, j'|k)$ between variables $X(j)$ and $X(j')$ given a distinct third variable $X(k)$, $k \in V\backslash\{j, j'\}$, the recursion is defined as

$$\rho(j, j'|k) = \frac{\rho(j, j') - \rho(j, k)\rho(j', k)}{\sqrt{1 - \rho^2(j, k)}\sqrt{1 - \rho^2(j', k)}}, \quad (5)$$

where $\rho(j, j')$, $\rho(j, k)$, and $\rho(j', k)$ represent the corresponding unconditional zero-order correlations for which estimators have already been presented. A general proof of the recursion formula is given by Anderson (2003, Section 2.5.3). In contrast to the case of saturated partial correlations, Equation (5) highlights the fact that lower-order partial correlations depend *explicitly* on the conditioning variable indexed by k. The regression-based formulation of first-order partial regressions will be discussed in more detail in Section 3.1.4.

2.2 Gaussian Graphical Models

Graphical models provide a convenient framework for handling large, multivariate data structures and for investigating complex relationships between variables. A graph $G = (V, E)$ consists of a set of vertices (variables) $V \subseteq \{1, \ldots, J\}$ connected by a series of edges in an edge set $E \subseteq \{1, \ldots, J\} \times \{1, \ldots, J\}$. In the Gaussian graphical model context, the data are assumed to be distributed according to a multivariate normal distribution, i.e., $X \sim P \equiv \mathrm{N}_J(\mu, \Sigma)$. By assigning elements $X(j)$ to particular vertices in V, the connection between statistical and graphical models is made. The elements $e(j, j')$ of the edge set E impose conditional independencies between variables $X(j)$ and $X(j')$ via *(pairwise and global) Markov properties* for different types of graphs, e.g., bidirectional, undirected, and directed acyclic graphs (DAGs). A recent review of this material has also been included in a paper by Drton and Perlman (2007), and key, relevant material from that discussion will also be presented in the following sections.

Examples of three types of commonly considered graphical models – bidirected, undirected, and directed acyclic graphs - are displayed in Figure 1. The graphs all consist of four nodes connected by three edges. For the purposes of the discussion to follow, we will assume the variables represented by the nodes in these graphs follow some four-dimensional data generating probability distribution, i.e., $X = (X(j) : j = 1, \ldots, 4) \sim P \in \mathbb{R}^4$.

2.2.1 Bidirected Gaussian Graphical Models

Bidirected graphs $G = (V, E)$ contain edges $E \subseteq \{1, \ldots, J\} \times \{1, \ldots, J\}$ taking the form $j \leftrightarrow j'$ (Figure 1, Graph A). An upper bound on the maximum number of edges is given by $\binom{J}{2}$. The *pairwise bidirected Markov property* of G confers *marginal independence* for all pairs (j, j') of variables in which the edge $j \leftrightarrow j'$, $1 \leq j < j' \leq J$ is absent in the graph (Drton and Perlman, 2007). Explicitly, from Figure 1, Graph A, we have $X(1) \perp\!\!\!\perp X(4)$, $X(2) \perp\!\!\!\perp X(3)$, $X(2) \perp\!\!\!\perp X(4)$. In this special case, the Markov conditioning variable is taken to be the empty set $\emptyset$.

For a Gaussian data generating distribution P, from Equation (1), we can now define the relationship

$$X(j) \perp\!\!\!\perp X(j') \quad \Longleftrightarrow \quad \rho(j, j') = 0. \tag{6}$$

Because of the central role of the variance-covariance matrix in estimating $\rho(j, j')$, bidirected graphical models have also been called *covariance graph models* where some authors prefer dashed edges to two-sided arrows (Cox and Wermuth, 1993, 1996, cited in Drton and Perlman, 2007).

This is not to say that the bidirected graph is useful only for identifying marginal independence relationships between pairs of variables. Conditional independencies may also be conferred by the edges of the bidirected graph. The *global bidirected Markov property* states that, for disjoint subsets A, B,

and C of the vertices in V,

$$X_A \perp\!\!\!\perp X_B | X_C \tag{7}$$

if every path from a vertex in A to a vertex in B must pass through a non-endpoint vertex not contained in C (from Drton and Perlman (2007)). From Figure 1, Graph A, it follows that $X(2) \perp\!\!\!\perp X(4)|X(3)$ because the path from variable $X(2)$ to variable $X(4)$ must through $X(1)$, which is a variable not indexed by the conditioning set $\{3\}$. Similar reasoning holds for $X(2) \perp\!\!\!\perp X(4)|X(1)$.

In functional genomics, bidirected graphical models are (knowingly or unknowingly) applied when probing coexpression patterns between genes. Again, in microarray expression studies, for example, the biological question is typically framed under the tacit assumption that coexpression implies coregulation at the transcript level.

2.2.2 Undirected Gaussian Graphical Models

Undirected graphs $G = (V, E)$ contain edges $E \subseteq \{1, \ldots, J\} \times \{1, \ldots, J\}$ taking the form $j - j'$ (Figure 1, Graph B). Similar to bidirected graphical model, an upper bound on the maximum number of edges is given by $\binom{J}{2}$. The *pairwise undirected Markov property* of G confers *conditional independence* for all pairs (j, j') of variables in which the edge $j - j'$ is absent from the graph (Drton and Perlman, 2007). In order to account for the effects of all other variables in G, the conditioning set is taken to be $V \backslash \{j, j'\}$. Specifically, from Figure 1, Graph B, it follows that $X(1) \perp\!\!\!\perp X(4)|(X(2), X(3))$, $X(2) \perp\!\!\!\perp X(3)|(X(1), X(4))$, and $X(2) \perp\!\!\!\perp X(4)|(X(1), X(3))$.

For a Gaussian data generating distribution P, from Equation (3), we can now define the relationship

$$X(j) \perp\!\!\!\perp X(j')|X(V \backslash \{j, j'\}) \quad \Longleftrightarrow \quad \rho(j, j'|V \backslash \{j, j'\}) = 0. \tag{8}$$

Due to the central role of the inverse variance-covariance matrix in Equation (3), undirected graphical models have also been called *concentration graph models* (Cox and Wermuth, 1996) as well as *covariance selection models* (Dempster, 1972) and *Markov random fields* (Jordan, 2003).

In addition to the pairwise Markov properties associated with undirected graphical models, there are also *global Markov properties* associated with these graphs. Global Markov properties are based on graph-theoretic arguments of naïve separation (Lauritzen, 1996; Jordan, 2003; Drton and Perlman, 2007). For disjoint subsets of vertices A, B, and C, the relationship $X_A \perp\!\!\!\perp X_B|X_C$ holds if every path connecting the vertices in A with the vertices in B must pass through those in C. For example, from Figure 1, Graph B, this means that $X(2) \perp\!\!\!\perp X(3)|X(1)$, and $X(2) \perp\!\!\!\perp (X(3), X(4))|X(1)$.

Shrinkage approaches and methods imposing sparsity restrictions on the graph (Dobra et al., 2004; Schäfer and Strimmer, 2004, 2005a,b; Meinshausen and Bühlmann, 2006; Li and Gui, 2006), have all been applied to undirected

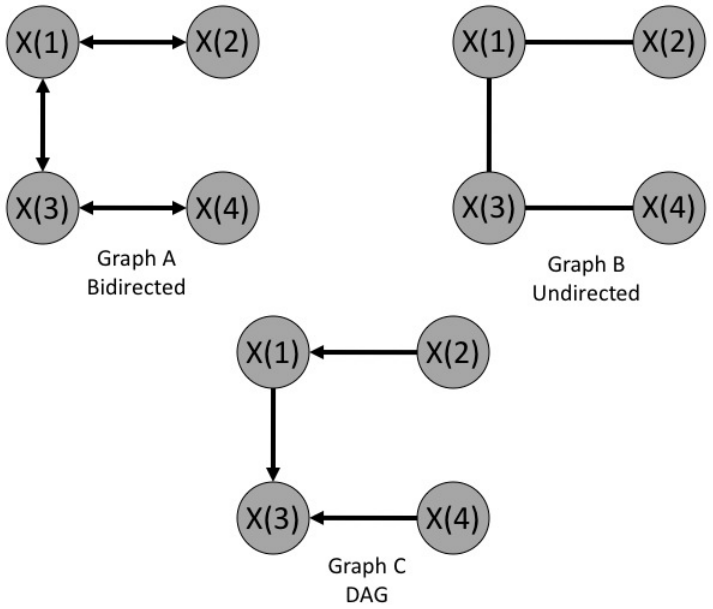

Figure 1: *Examples of bidirected (A), undirected (B) and directed acyclic graphical models (C).* As described in the text, the edge set E for each type of graph structure will dictate a set of (conditional) independence relationships among variables indexed by the vertex set V. For bidirected Gaussian graphical models of the form in subfigure A, one interpretation of the missing edge between $X(2)$ and $X(4)$ is that the two variables are marginally independent, with pairwise correlation equal to zero. In contrast, for undirected Gaussian graphical models of the form in subfigure B, one interpretation of the missing edge between $X(2)$ and $X(4)$ is that the two variables are conditionally independent given all other variables (i.e., $X(1)$ and $X(3)$), with pairwise saturated partial correlation equal to zero. As mentioned in the text, for directed acyclic Gaussian graphical models of subfigure C, the conditional independence relationships will depend on a well-numbering of the vertices in the V. In this particular case, the absent edge implies that $X(2)$ and $X(4)$ are conditionally independent given $X(1)$. Generally, partial correlations from these graphs will be of some order depending on a submatrix of the covariance matrix $\Sigma(P) = \sigma$ (see references in text).

graphical models in the more familiar situation confronted in modern biology where $p \gg n$. Other researchers have instead used *lower-order partial correlation coefficients* to quantify dependencies among variables in the graph G (de la Fuente et al., 2004; Magwene and Kim, 2004; Wille et al., 2004; Castelo and Roverato, 2006; Wille and Bühlmann, 2006).

By restricting the analysis to lower-order partial correlations, higher-level dependencies between sets of variables may be missed. In this sense, a restricted analysis is generally no longer firmly grounded in the world of undirected graphical models. Wishing to explore the probabilistic and estimation

properties of graphs using only zero- and first-order partial correlations, Wille and Bühlmann (2006) found that such a method may still provide a good approximation for the concentration graph model. Wille and Bühlmann (2006) also show that the two graphs coincide in some situations. We may adopt the shorthand notation adopted by Wille and Bühlmann (2006) from de Campos and Huete (2000) in which these graphs are referred to as '0-1' graphs. The '0-1' moniker here refers to the fact that only *zero-* and *first-* order partial correlations are considered when constructing the graph.

2.2.3 Directed Acyclic Graphical Models

Popular in the machine learning and epidemiologic causal inference literature, a directed acyclic graph (DAG) is characterized by unidirectional edges $j \rightarrow j'$ (Figure 1, Graph C). Whereas the edge sets in bidirected and undirected graphical models may be considered *unordered*, the edge set $E \subseteq \{1, \ldots, J\} \times \{1, \ldots, J\}$ associated with a DAG defines a *partial ordering* of the vertices in V in which no path exists connecting a vertex node j back to itself. As a result, the maximum number of edges in a DAG is typically some number less than $\binom{J}{2}$. The Markov properties of DAGs are therefore also somewhat more complicated than the above cases. Specifically, conditional independence assertions made regarding two variables with a missing edge between them must also adhere to a *well-numbering or topological ordering* of V. In a related concept, unlike the undirected graphical model case, rather than relying on graph theoretic properties of naïve separation, both the *directed pairwise Markov property* and the *directed global Markov property* rely on a slightly more complicated concept of d-separation ("directed separation", Lauritzen, 1996; Jordan, 2003, Chapter 3 and Chapter 16, respectively). Both the terminology as well as the theoretical underpinnings associated with DAGs extend beyond the current discussion. The reader is therefore referred to the work cited in this section for further clarification.

DAGs are often used to represent relationships of cause and effect. While this causal property may be appealing, the use of DAGs in studying biological networks is often argued to be conceptually difficult. It is well known, for example, that biological pathways may rely on positive and negative feedback loops for autoregulation. The cyclicity of such mechanisms precisely characterizes those relationships which, by definition, are disallowed by a DAG structure. Given the statistical challenges associated with DAGs, their distributional properties and estimation remain an active area of research (Kalisch and Bühlmann, 2007, for example). Examples with biological applications can be found in Drton and Perlman (2007) and Opgen-Rhein and Strimmer (2007). A paper investigating the use of lower-order conditional independencies in the DAG context was recently published by Lèbre (2009). Again, in this case, the Benjamini-Hochberg (Benjamini and Hochberg, 1995) method was used to control the FDR at fairly stringent siginificance levels, i.e., $\alpha = 0.01$ (Lèbre, 2009).

2.3 Inferring Lower-Order Graphs

2.3.1 Choice of Test Statistics

Recall from our discussion in Section 2.1 that two possible parameters of interest were given by the (unconditional) correlations of Equation (1) and the first-order partial correlations defined in Equation (5). We denote these parameters by $\rho(j, j')$ and $\rho(j, j'|k)$, respectively.

Zero-order Correlations (Bidirected Graphical Models). While in general a lack of correlation does not imply independence, in the Gaussian bidirected graphical model, statements about marginal independence between nodes in G and unconditional correlation are equivalent. An absent edge implies marginal independence, which in turn implies zero correlation. Contrapositively, if one rejects the null hypothesis that $\rho(j, j') = 0$, then one may draw an edge between the two nodes j and j'. In this case, one may turn to familiar methods for testing hypotheses of lack of correlation in order to reconstruct the graph G from observed data. A suitable choice of test statistic is given by the empirical correlation coefficient defined in Equation (2). For n copies of normal IID data, i.e., $X \sim \mathrm{N}_J(\mu, \Sigma)$, and assuming the population correlation coefficient $\rho(j, j') = 0$, the test statistic

$$T_n(j, j') = \sqrt{n-2}\frac{\rho_n(j, j')}{\sqrt{1-\rho_n^2(j, j')}} \tag{9}$$

follows a t-distribution with $n - 2$ degrees of freedom.

Relying on asymptotic properties of the sample correlation coefficient, Fisher's z transformation, may also be used for testing hypotheses involving correlation parameters. Fisher's $z = \zeta_n$ is defined as

$$T_n(j, j') = \zeta_n(j, j') = \frac{1}{2}\log\frac{1+\rho_n(j, j')}{1-\rho_n(j, j')} = \tanh^{-1}\rho_n(j, j'). \tag{10}$$

Let

$$\zeta(j, j') = \frac{1}{2}\log\frac{1+\rho(j, j')}{1-\rho(j, j')} \tag{11}$$

denote the value of Fisher's z at the true value of the population correlation coefficient $\rho(j, j')$. Without loss of generality, i.e., even in the noncentral case in which $\rho(j, j') \neq 0$, a normal approximation holds such that

$$\sqrt{n-3}(\zeta_n(j, j') - \zeta(j, j')) \xrightarrow{d} \mathrm{N}(0,1), \quad \text{as} \quad n \to \infty. \tag{12}$$

Fisher's z-transform has asymptotic variance independent of the population correlation and is a *variance-stabilizing transformation* (Anderson, 2003, Section 4.2.3).

Partial Correlations (Undirected Graphical Models and Lower-Order Conditional Independence Graphs). Similar to bidirectional graphical models, undirected graphical models imply relationships between conditional independence statements and the values of (saturated) partial correlation coefficients. A nice result for test statistics of hypotheses of partial correlations is that the empirical partial correlations are distributed similarly to their unconditional correlation counterparts, with only slight modifications to account (penalize) for the cardinality of the conditioning set (Drton and Perlman, 2007; Anderson, 2003, Theorem 4.3.5). In the saturated case, one may substitute the empirical partial correlations of Equation (4) into the test statistics formula in Equation (9) to generate test statistics which following a t-distribution with $n - |V\backslash\{j, j'\}| - 2$ degrees of freedom. Similarly, replacing $\rho(j, j')$ and $\rho_n(j, j')$ with $\rho(j, j'|V\backslash\{j, j'\})$ and $\rho_n(j, j'|V\backslash\{j, j'\})$ to produce $\zeta(j, j'|V\backslash\{j, j'\})$ and $\zeta_n(j, j'|V\backslash\{j, j'\})$ will produce the following augmentation to Fisher's z-transform. As $n \to \infty$,

$$\sqrt{n - |V\backslash\{j, j'\}| - 3}(\zeta_n(j, j'|V\backslash\{j, j'\}) - \zeta(j, j'|V\backslash\{j, j'\})) \xrightarrow{d} \mathrm{N}(0,1). \quad (13)$$

When exploring 0-1 graphs, we again consider normally distributed data $X \sim \mathrm{N}_J(\mu, \Sigma)$. Test statistics $T_n(j, j')$ for zero-order correlations are given above (Equations (9) and (10)). In the case of first-order partial correlations, we are specifically interested in the parameter $\rho(j, j'|k)$ corresponding to the partial correlation of variables $X(j)$ and $X(j')$ conditional on a single third variable $X(k)$, $k \neq j \neq j'$. Similar to the formulas above, the cardinality of the conditioning set is simply equal to one. It follows that, depending on the choice of test statistic, the first-order partial correlation test statistics $T_n(j, j'|k)$ may either be distributed t_{n-3} or (asymptotically) N(0,1) with a Fisher's z standard deviation of $\sqrt{n-4}$.

Another option for testing hypotheses concerning correlations is through maximum likelihood estimation. Likelihood approaches have been described in Anderson (2003, Section 4.2.2), Wille et al. (2004), and Wille and Bühlmann (2006). They are often optimal for hypotheses based on the normal distribution (Anderson, 2003, Section 4.2.2). When testing first-order partial correlations of the form $\rho(j, j'|k)$, the (log) likelihood can be easily obtained and used for testing elements of the lower-dimensional information matrix Σ^{-1}, specifically $H_0(j, j'|k) = \Sigma^{-1}(j, j') \equiv \mathrm{I}(\rho(j, j'|k) = 0)$. Log-likelihood ratio statistics may then be used to test the hypotheses defined above. In this case, the test statistics are asymptotically χ^2-distributed.

2.3.2 Edge Inclusion Algorithms

We focus on the frequentist algorithm for estimating lower-order conditional independence graphs described by Wille et al. (2004) and slightly modified later by Wille and Bühlmann (2006) to include the 0-1 graph case. The main difference between the two estimation procedures appears to be whether to include the unconditional pairwise correlation coefficients in the algorithm (see

below). For completeness and with minor changes from the original, the algorithm proceeds as given below. While Wille et al. (2004) and Wille and Bühlmann (2006) propose calculating p-values from likelihood ratio test statistics, we have left that portion of the following algorithm more general to reflect the discussion of other choices of test statistics described earlier.

0-1 GRAPH ALGORITHM OF WILLE AND BÜHLMANN

1. Our parameters of interest are $\rho(j, j|k)$, i.e., the zero-order and first-order pairwise partial correlation coefficients corresponding to variables $X(j)$ and $X(j')$, conditional on each possible third variable indexed by $k \in \emptyset \cup V\backslash\{j, j'\}$. The case in which $k = \emptyset$ refers to unconditional, zero-order correlation coefficients. We wish to test hypotheses of the form $H_0(j, j'|k) = \mathrm{I}(\rho(j, j'|k) = 0)$. For all $j, j' \in \{1, \ldots, J\}$, $j \neq j'$ and $k \in \emptyset \cup V\backslash\{j, j'\}$, $|k| = 1$, calculate the empirical first-order partial correlation coefficients $\rho_n(j, j'|k)$ and their corresponding test statistics $T_n(j, j'|k)$. Given a suitable choice of marginal null distribution $Q_0(j, j')$, also compute the *unadjusted p-values* $P_{0n}(j, j'|k)$ for each test statistic.

2. For each of the (j, j')-pairs, compute the *maximum p-values* $P_{0n,max}(j, j')$ (alternatively, *minimum (absolute) test statistics*, $T_{n,min}(j, j')$), over all k, i.e.,
$$P_{0n,max}(j, j') = \max_{k \in \emptyset \cup \{1,2,\ldots,J\}\backslash\{j,j'\}} P_{0n}(j, j'|k).$$

3. Apply a multiple testing procedure over the collection of $M = \binom{J}{2} = J(J-1)/2$ composite maximum p-values $P_{0n,max}(j, j')$, e.g., the Bonferroni (1936) procedure for control of the family-wise error rate (FWER) or the Benjamini-Hochberg (Benjamini and Hochberg, 1995) procedure for control of the FDR. The *adjusted* maximum p-values are denoted by
$$\widetilde{P}_{0n,max}(j, j').$$

4. Draw an edge between vertex j and j' iff
$$\widetilde{P}_{0n,max}(j, j') < \alpha,$$
for some prespecified significance level (i.e., $\alpha = 0.05$).

The intuition behind the algorithm can be described with the following logic. First, one should note that there exists a two-dimensional testing problem in this scenario. That is, for each of the $M = \binom{J}{2} = J(J-1)/2$ (j, j')-pairs of variables, $1+(J-2) = (J-1)$ zero-order and first-order partial correlations must be computed, for a total of $J(J-1)(J-1)/2$ calculations. Assuming the test statistics are arranged in an $M \times (J-1)$ matrix, obtaining $P_{0n,max}$ amounts to finding row maxima, thereby reducing the testing problem into one dimension (a column vector) through the composite p-value measure. If

a correlation between variables $X(j)$ and $X(j')$ is the possible result of another variable $X(k)$, then one would expect $\rho(j, j'|k)$ to tend towards zero, which implies a small value of the test statistic T_n and a correspondingly large p-value $P_{0n}(j, j'|k)$. By taking the maximum p-value over all k, one assumes there exists at least one other gene which may "explain away" the association between j and j' (Wille et al., 2004). Conversely, for an association to remain significant after calculating row maxima, then the (j, j')-specific p-values calculated over k must all be small, meaning that the association between j and j' has endured even after accounting for the individual possible effects of all other genes. Another by-product of this algorithm is also therefore an M-vector of indices $k_{max} = (k_{max}(m) : m = 1, \ldots, M)$ with elements $k_{max}(m)$ corresponding to the variables (genes) which yielded the respective values of $P_{0n,max}(j, j')$.

The adjusted p-values $\widetilde{P}_{0n,max}$ serve as a measure of significance between genes j and j' and can be used to test for the presence of edges in the 0-1 graph. As noted in Wille et al. (2004) and Wille and Bühlmann (2006), the use of $P_{0n,max}(j, j')$ is a strong, possibly over-conservative, composite measure. Wille and Bühlmann (2006) also state,

> It should be noted, however, that we test the null hypothesis that at least one $[H_0(j, j'|k)]$ is true versus the alternative that none $[H_0(j, j'|k)]$ is true. Therefore, less conservative approaches (Holm, 1979; Simes, 1986) are not applicable.

In fact, this statement appears to blur some issues inherent in a two-dimensional *sequential* testing problem where the focus is on the tests over the M rows. The term sequential refers to the fact that the vector k_{max}, which serves to collapse one of the dimensions of the matrix of test statistics, has been obtained in a data-adaptive manner. The MTP carried out over the vector M-vector $P_{0n,max}$ in Step 3 of the above algorithm, is then *conditional* on k_{max}. That is, once the vector of maximum p-values $P_{0n,max}$ (or minimum test statistics $T_{n,min}$) has been obtained, multiple testing is no longer performed explicitly in the k direction (across columns). Rather, any candidate subsequent MTP is carried out on values in the (j, j')-direction (across rows). Therefore, as long as one is clear that Type I error control can only be guaranteed when holding the variables indexed by k_{max} fixed, then several choices of (single-step or step-wise) MTPs are generally valid. In particular, the Holm procedure cited above (Holm, 1979), like the Bonferroni procedure, is also a marginal MTP which makes no assumptions on the joint distribution of the data or test statistics. Moreover, if FDR control is applicable in this testing situation, then a step-wise procedure for control of the more stringent FWER, is also certainly applicable.

3 Methods

All multiple testing procedures require a null distribution to specify cut-offs and/or adjusted p-values in order for a given Type I error rate to be probabilistically controlled. This section begins with a review of points to consider when selecting a null distribution for use in testing scenarios. With this discussion serving as a springboard, we continue with the derivation of a test statistics joint null distribution estimate useful when testing for the presence of edges in a lower-order conditional independence graph.

After a review of various Type I error rates, the final part of this section describes an empirical Bayes multiple testing methodology controlling generalized error rates defined as parameters of the distribution of functions $g(V_n, S_n)$ of the numbers of false positives V_n and true positives S_n (van der Laan et al., 2005; Dudoit et al., 2008; Dudoit and van der Laan, 2008, Chapter 7). Examples of such error rates include, among others, the FWER and FDR. A central component of this framework involves generating guessed sets of true null hypotheses. These powerful, joint MTPs effectively make use of the test statistics null distribution formulated below.

3.1 Choice of Null Distribution

Many researchers propose the use of (empirical variants of) standard, easy-to-implement *marginal multiple testing procedures* such as the Bonferroni and Benjamini-Hochberg methods to calculate adjusted p-values $\widetilde{P}_{0n,max}$ (Wille et al., 2004; Magwene and Kim, 2004; de la Fuente et al., 2004; Schäfer and Strimmer, 2005b; Wille and Bühlmann, 2006). Marginal multiple testing procedures are those which are based solely on the marginal distributions of the test statistics, i.e., on cut-off rules for the corresponding unadjusted p-values.

Joint multiple testing procedures, e.g., maxT or minP procedures, attempt to improve upon marginal MTPs by accounting for dependencies between the test statistics used to probe each hypothesis. Implementation of joint MTPs can be more complicated and more computationally intensive than conducting multiple hypothesis testing using a marginal procedure. Rather than relying on influence curves (see below), suitable *test statistics joint null distribution* are often more generally specified and estimated via (nonparametric or model-based) resampling procedures. Westfall and Young (1993) provide a framework for resampling-based multiple hypothesis testing which relies on the *subset pivotality* assumption for generating a valid joint null distribution. Tests of correlation coefficients in particular are one scenario in which this assumption fails to hold, in part because the test statistics joint null distribution may depend on the truth or falsehood of hypotheses concerning dependencies among test statistics (p. 43, Westfall and Young, 1993; Dudoit and van der Laan, 2008, Section 2.6.5).

Asymptotically valid test statistics joint null distributions which do not rely on the subset pivotality assumption and which are available for general

data generating distributions have been characterized for use with joint MTPs (Pollard and van der Laan, 2004; Dudoit et al., 2004; van der Laan et al., 2004; van der Laan and Hubbard, 2006; Dudoit and van der Laan, 2008, Chapter 2). These null distributions are based on *null domination conditions*, which state that one must select a test statistics null distribution Q_0 (or estimator thereof, Q_{0n}) that stochastically dominates the unknown true distribution of test statistics $Q_n = Q_n(P)$. Choosing such a Q_0 ensures that one makes more Type I errors under the null distribution than one would have had committed had the true distribution Q_n been available.

The first original proposal of Pollard and van der Laan (2004), Dudoit et al. (2004), and van der Laan et al. (2004), defines the null distribution as the asymptotic distribution of a vector of *null shift and scale-transformed test statistics*, based on user-supplied upper bounds for the means and variances of the test statistics for the true null hypotheses (Dudoit and van der Laan, 2008, Section 2.3). The second and most recent proposal of van der Laan and Hubbard (2006) defines the null distribution as the asymptotic distribution of a vector of *null quantile-transformed test statistics*, based on user-supplied test statistic marginal null distributions (Dudoit and van der Laan, 2008, Section 2.4). These marginal null distributions are often the ones the user would have selected in a univariate testing situation, including marginal permutation distributions. In practice, the joint null distribution Q_0 is also typically unknown and itself must be estimated from the data. Resampling procedures (e.g., non-parametric or model-based bootstrap) are available to conveniently obtain consistent estimators Q_{0n} of the null distribution and of the corresponding test statistic cut-offs, parameter confidence regions, and adjusted p-values (Dudoit and van der Laan, 2008, Procedures 2.3 and 2.4). One limitation of these procedures is that they can often become computationally burdensome, particularly for large numbers of hypotheses M, large numbers B of bootstraps or permutations, and for more complicated test statistics.

For a broad class of testing problems, such as the test of single-parameter null hypotheses using t-statistics, an asymptotically valid null distribution is the M-variate Gaussian distribution $\mathrm{N}(0, \sigma^*)$, with mean vector zero and covariance matrix $\sigma^* = \Sigma^*(P)$ equal to the correlation matrix of the vector influence curve for the estimator ψ_n of the parameter of interest ψ (Dudoit and van der Laan, 2008, Section 2.6). In this case, one may simply simulate from a suitable multivariate normal distribution rather than committing to calculating B vectors of permutation- or bootstrap-based test statistics. Test statistic-specific null distribution estimation approaches may, as in the case of t-statistics, yield continuous null distributions, which have the additional advantage that they may not suffer as much from the discreteness of the bootstrap as previously proposed methods (Dudoit and van der Laan, 2008, Section 2.3.2).

In the case of saturated partial correlation coefficients, a similar approach was taken by Drton and Perlman (2007) in which an estimate of the variance-covariance matrix of partial correlations in the concentration graph model was

used to test for edges in the '$n > p$' setting. Our approach for the 0-1 graph involves deriving a joint null distribution for lower-order partial correlation test statistics which extends the ability to approximate the dependencies between these correlations in the setting where the number of hypotheses may exceed the sample size.

The purpose of the following sections is to describe the formulation of a test statistics joint null distribution useful for testing hypotheses involving (partial) correlation coefficients. The presentation of the material below very closely follows that of Dudoit and van der Laan (2008, Sections 1.2.5, 2.3.2, 2.6.1, 2.6.4, and 2.6.5).

3.1.1 General Set-Up: t-statistics

We consider the two-sided test of M single-parameter null hypotheses $H_0(m) = \mathrm{I}(\psi(m) = \psi_0(m))$ against alternative hypotheses $H_1(m) = \mathrm{I}(\psi(m) \neq \psi_0(m))$, where $\Psi(P) = \psi = (\psi(m) : m = 1, \ldots, M)$ is an M-vector of real-valued parameters $\Psi(P)(m) = \psi(m)$. The value of the null parameter $\psi_0(m)$ is often taken to be zero.

The null hypotheses can be tested using an M-vector of *t-statistics* $T_n = (T_n(m) : m = 1, \ldots, M)$, defined by

$$T_n(m) \equiv \frac{\text{Estimator} - \text{Null value}}{\text{Standard error}} = \sqrt{n}\frac{\psi_n(m) - \psi_0(m)}{\sigma_n(m)}, \tag{14}$$

where $\hat{\Psi}(P_n) = \psi_n = (\psi_n(m) : m = 1, \ldots, M)$ is an *asymptotically linear estimator* of the parameter M-vector $\Psi(P) = \psi$, with M-dimensional vector *influence curve* (IC) $\mathrm{IC}(X|P) = (\mathrm{IC}(X|P)(m) : m = 1, \ldots, M)$, such that

$$\psi_n(m) - \psi(m) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{IC}(X_i|P)(m) + o_P(1/\sqrt{n}), \tag{15}$$

and $\mathrm{E}[\mathrm{IC}(X|P)(m)] = 0$ for $m = 1, \ldots, M$. Let $\Sigma(P) = \sigma = (\sigma(m, m') : m, m' = 1, \ldots, M)$ denote the $M \times M$ parameter covariance matrix. Assume that one can obtain consistent estimators $\sigma_n(m, m')$ of the covariances $\sigma(m, m') = \mathrm{E}[\mathrm{IC}(X|P)(m)\,\mathrm{IC}(X|P)(m')]$ as well as consistent estimators $\sigma_n^2(m)$ of the variances $\sigma^2(m) = \sigma(m, m) = \mathrm{E}[\mathrm{IC}^2(X|P)(m)]$, $m = 1, \ldots, M$.

The influence curve of a given estimator can be derived as its mean-zero-centered functional derivative, i.e., as a function of the empirical distribution P_n for the entire sample of size n, applied to the empirical distribution for a sample of size one (Gill, 1989; Gill et al., 1995). A vector influence curve is therefore estimated for each of n observations.

Let $Q_n = Q_n(P)$ denote the finite sample joint distribution of T_n, under the true, unknown data generating distribution P. Large absolute values of the t-statistic $T_n(m)$ are assumed to provide evidence against the corresponding null hypothesis $H_0(m) = \mathrm{I}(\psi(m) = \psi_0(m))$.

By the central limit theorem, for the test of single-parameter null hypotheses using t-statistics, a *t-statistic-specific null distribution* $Q_0^t = Q_0^t(P)$ is the M-variate Gaussian distribution $\mathrm{N}(0, \sigma^*)$, where $\sigma^* = \Sigma^*(P)$ is the correlation matrix of the M-dimensional vector influence curve, $\mathrm{IC}(X|P) = (\mathrm{IC}(X|P)(m) : m = 1, \ldots, M)$, for an asymptotically linear estimator ψ_n of the parameter M-vector ψ.

In this case, one can estimate Q_0^t by $Q_{0n}^t = \mathrm{N}(0, \sigma_n^*)$, where $\sigma_n^* = \hat{\Sigma}^*(P_n)$ is a consistent estimator of the correlation matrix σ^*. For example, one could use the correlation matrix σ_n^* corresponding to the following estimator of the $M \times M$ influence curve covariance matrix,

$$\sigma_n = \hat{\Sigma}(P_n) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{IC}_n(X_i)\, \mathrm{IC}_n^\top(X_i), \tag{16}$$

where $\mathrm{IC}_n(X) = (\mathrm{IC}_n(X)(m) : m = 1, \ldots, M)$ is an estimator of the M-vector influence curve $\mathrm{IC}(X|P)$.

3.1.2 Tests of Means

Influence curves can be derived straightforwardly for simple parameters such as means. For example, in the one-sample case, when estimating the mean vector $\psi = \mathrm{E}[X]$, for a random M-vector $X \sim P$, using the empirical mean vector $\psi_n = \bar{X}_n$, the influence curves are $\mathrm{IC}(X|P)(m) = X(m) - \psi(m)$ and the corresponding estimators are $\mathrm{IC}_n(X|P)(m) = X(m) - \psi_n(m)$, where $\psi_n(m) = \bar{X}_n(m) = \sum_i X_i(m)/n$, $m = 1, \ldots, M$. Then, σ_n^* is simply the empirical correlation matrix. Note that the components of the vector influence curve for each subject reflect the concept of the difference between a sample of size one ($X(m)$) and a sample of size n ($\bar{X}_n(m)$).

3.1.3 Tests of Correlation Coefficients

A common testing problem occurs when the parameter of interest is the $J \times J$ *correlation* matrix for a random J-vector $X \sim P$, that is, $\Psi(P) = \psi = (\psi(j, j') : j, j' = 1, \ldots, J) = \mathrm{Cor}[X]$, with elements $\psi(j, j') = \Psi(P)(j, j') = \mathrm{Cor}[X(j), X(j')]$. Suppose one is interested in testing the $M = \binom{J}{2} = J(J-1)/2$ null hypotheses $H_0(j, j') = \mathrm{I}(\psi(j, j') = 0), j = 1, \ldots, J-1, j' = j+1, \ldots, J$. To reiterate, this testing problem may correspond to testing for coexpression in a microarray study, or, in the graphical model context, to testing for edges in a bidirectional graph (covariance model) where the Markov properties imply marginal independence among unconnected nodes.

Common test statistics for this problem are $T_n(j, j') = \sqrt{n}\psi_n(j, j')$, where $\psi_n(j, j') = \hat{\Psi}(P_n)(j, j')$ are the *empirical correlation coefficients*. The influence curves for the empirical correlation coefficients $\psi_n(j, j')$ can be obtained by applying the multivariate δ-method with the function:

$$
\begin{aligned}
f(\xi(j,j')) = \psi(j,j') &= \mathrm{Cor}[X(j), X(j')] \qquad (17)\\
&= \frac{\mathrm{Cov}[X(j), X(j')]}{\sqrt{\mathrm{Var}[X(j)]}\sqrt{\mathrm{Var}[X(j')]}}\\
&= \frac{\mathrm{E}[X(j)X(j')] - \mathrm{E}[X(j)]\,\mathrm{E}[X(j')]}{\sqrt{\mathrm{E}[X^2(j)] - (\mathrm{E}[X(j)])^2}\sqrt{\mathrm{E}[X^2(j')] - (\mathrm{E}[X(j')])^2}},
\end{aligned}
$$

defined in terms of a 5×1 parameter vector $\xi(j,j') = \Xi(P)(j,j') = [\mathrm{E}[X(j)], \mathrm{E}[X(j')], \mathrm{E}[X^2(j)], \mathrm{E}[X^2(j')], \mathrm{E}[X(j)X(j')]]^\top$, $j, j' = 1, \ldots, J$. Let $f'(\xi)$ denote the 1×5 gradient row vector of $f(\xi)$. Then,

$$
\psi_n(j,j') - \psi(j,j') = f'(\xi(j,j'))(\xi_1(j,j') - \xi(j,j')) + o_P(1/\sqrt{n}), \qquad (18)
$$

where $\xi_n(j,j') = \hat{\Xi}(P_n)(j,j') = [X(j), X(j'), X^2(j), X^2(j'), X(j)X(j')]^\top$ is a 5×1 column vector based on the empirical moments. In the end, the influence curve for the estimator $\psi_n(j,j') \equiv \rho_n(j,j')$ defined above in Equation (2) is given by

$$
\mathrm{IC}(X|P)(j,j') = f'(\xi(j,j'))(\xi_1(j,j') - \xi(j,j')) \qquad (19)
$$

$$
= \frac{1}{\sqrt{\mathrm{Var}[X(j)]}\sqrt{\mathrm{Var}[X(j')]}}
\begin{bmatrix}
\mathrm{E}[X(j)]\frac{\mathrm{Cov}[X(j),X(j')]}{\mathrm{Var}[X(j)]} - \mathrm{E}[X(j')] \\
\mathrm{E}[X(j')]\frac{\mathrm{Cov}[X(j),X(j')]}{\mathrm{Var}[X(j')]} - \mathrm{E}[X(j)] \\
-\frac{1}{2}\frac{\mathrm{Cov}[X(j),X(j')]}{\mathrm{Var}[X(j)]} \\
-\frac{1}{2}\frac{\mathrm{Cov}[X(j),X(j')]}{\mathrm{Var}[X(j')]} \\
1
\end{bmatrix}^\top
$$

$$
\times
\begin{bmatrix}
X(j) - \mathrm{E}[X(j)] \\
X(j') - \mathrm{E}[X(j')] \\
X^2(j) - \mathrm{E}[X^2(j)] \\
X^2(j') - \mathrm{E}[X^2(j')] \\
X(j)X(j') - \mathrm{E}[X(j)X(j')]
\end{bmatrix}.
$$

Therefore, for each individual, one obtains a general formula for the components of the M-dimensional influence curve. A plug-in estimator $\mathrm{IC}_n(X|P)$ of the vector influence curve is available by replacing each of the elements in Equation (19) with their empirical analogs (moments, means, variances and covariances). From these curves, one can estimate Q_0^t by $Q_{0n}^t = \mathrm{N}(0, \sigma_n^*)$, where $\sigma_n^* = \hat{\Sigma}^*(P_n)$ is an estimator of the correlation matrix for the empirical correlation coefficients derived from the influence curves as described in Equation (16).

3.1.4 Tests of First-Order Partial Correlation Coefficients

As in the 0-1 graph algorithm of Wille and Bühlmann (Section 2.3.2), one may encounter the testing problem in which the parameter of interest is the $J \times J$ *first-order partial correlation* matrix for a random J-vector $X \sim P$, conditional on a vector of indices $k_{max}(j, j')$, whose corresponding elements each represent the single variable thought to best "explain away" the potential association between variables $X(j)$ and $X(j')$. That is, $\Psi(P) = \psi = (\psi(j, j'|k) : j = 1, \ldots, J-1, j' = j+1, \ldots, J, k \equiv k_{max}(j, j') \in \{1, \ldots, J\} \backslash \{j, j'\})$, with elements $\psi(j, j'|k) = \Psi(P)(j, j'|k) = \text{Cor}[X(j), X(j')|X(k)]$. Suppose one is interested in testing the $M = J(J-1)/2$ null hypotheses $H_0(j, j'|k) = \text{I}(\psi(j, j'|k) = 0), j = 1, \ldots, J-1, j' = j+1, \ldots, J, k = k_{max}(j, j')$.

Common test statistics for this problem are $T_n(j, j'|k) = \sqrt{n}\ \psi_n(j, j'|k)$, where $\psi_n(j, j'|k) = \hat{\Psi}(P_n)(j, j'|k)$ are the *first-order empirical partial correlation coefficients* as defined in Equation (5). The parameter function defining these first-order partial correlation coefficients is given by:

$$\begin{aligned} f(\xi(j, j'|k)) &= \psi(j, j'|k) \\ &= \text{Cor}[X(j), X(j')|X(k)] \\ &= \frac{\text{Cov}[X(j), X(j')|X(k)]}{\sqrt{\text{Var}[X(j)|X(k)]}\sqrt{\text{Var}[X(j')|X(k)]}}, \end{aligned} \tag{20}$$

where

$$\begin{aligned} \text{Cov}[X(j), X(j')|X(k)] &= \text{E}[(X(j) - \text{E}[X(j)|X(k)]) \\ &\qquad \times (X(j') - \text{E}[X(j')|X(k)])|X(k)] \\ &= \text{E}[X(j)X(j')|X(k)] - \text{E}[X(j)|X(k)]\,\text{E}[X(j')|X(k)]. \end{aligned}$$

One feature of Equation (20) is that it is defined in terms of a parameter vector $\xi(j, j'|k) = \Xi(P)(j, j'|k) = [\text{E}[X(j)|X(k)], \text{E}[X(j')|X(k)], \text{E}[X^2(j)|X(k)], \text{E}[X^2(j')|X(k)], \text{E}[X(j)X(j')|X(k)]]^\top, j, j', k = 1, \ldots, J, j \neq j' \neq k$ containing five different conditional expectations, each requiring a method for estimation in their own right.

The conditional expectations in Equation (20) represent regression functions. Influence curves for regression coefficient parameters have been derived (Dudoit and van der Laan, 2008, Section 2.6.6), although their use in this scenario will lead to even more complicated parameterizations of the vector influence curve for correlation coefficients, particularly with continuous $X(k)$. Specifically, at the parameter level, the influence curve will have two components; one for the correlation parameter, and a second, nested influence curve for the regressions. Having observed that working with the above formula requires an underlying mechanism for estimating the conditional expectations,

we note that first-order partial correlations may also be estimated in terms of the residuals obtained from linear regression.

In the least squares context, for example, we can regress both $X(j)$ and $X(j')$ on $X(k)$ and compute the correlation between the n-dimensional error vectors $\epsilon(j)$ and $\epsilon(j')$, i.e., $\mathrm{Cor}[\epsilon(j), \epsilon(j')]$. In this case, the formula for the vector influence curve for estimates of first-order partial correlation coefficients between pairs of error vectors $\epsilon(j)$ and $\epsilon(j')$, and, therefore, between pairs of variables $X(j)$ and $X(j')$ conditional on $X(k)$, reduces down to

$$\begin{aligned} \mathrm{IC}(X|P)(j,j'|k) &= f'(\xi(j,j'|k))(\xi_1(j,j'|k) - \xi(j,j'|k)) \qquad (21) \\ &= \frac{1}{\sqrt{\mathrm{Var}[\epsilon(j)]}\sqrt{\mathrm{Var}[\epsilon(j')]}} \begin{bmatrix} -\frac{1}{2}\frac{\mathrm{Cov}[\epsilon(j),\epsilon(j')]}{\mathrm{Var}[\epsilon(j)]} \\ -\frac{1}{2}\frac{\mathrm{Cov}[\epsilon(j),\epsilon(j')]}{\mathrm{Var}[\epsilon(j')]} \\ 1 \end{bmatrix}^\top \\ &\quad \times \begin{bmatrix} \epsilon(j)^2 - \mathrm{Var}[\epsilon(j)] \\ \epsilon(j')^2 - \mathrm{Var}[\epsilon(j')] \\ \epsilon(j)\epsilon(j') - \mathrm{E}[\epsilon(j)\epsilon(j')] \end{bmatrix}. \end{aligned}$$

To clarify, we assume the following linear models for the conditional expected values of the variables $X(j)$ and $X(j')$ conditional on $X(k)$, $j, j', k \in \{1, \ldots, J\}, j \neq j' \neq k$,

$$\mathrm{E}[X(j)|X(k)] = \beta_{0,j} + \beta_{1,j}X(k), \qquad (22)$$

with errors given by

$$\epsilon(j) = X(j) - \mathrm{E}[X(j)|X(k)], \qquad \mathrm{E}[\epsilon(j)] = 0, \quad \epsilon(j) \perp\!\!\!\perp X(k). \qquad (23)$$

Letting $\hat{\beta}$ denote a regression coefficient estimate of the β terms in Equation (22), the error for the ith observation may be estimated by its corresponding residual, i.e.,

$$e_i(j) = X_i(j) - (\hat{\beta}_{0,j} + \hat{\beta}_{1,j}X_i(k)). \qquad (24)$$

From this point, the first-order partial correlation estimate $\rho_n(j, j'|k)$ can be obtained from the Pearson product-moment correlation coefficient calculated from the residual n-vectors $e(j)$ and $e(j')$. Specifically, in addition to the recursion formula in Equation (5), we also have

$$\rho_n(j,j'|k) = \frac{n\sum_{i=1}^n e_i(j)e_i(j') - \sum_{i=1}^n e_i(j)\sum_{i=1}^n e_i(j')}{\sqrt{n\sum_{i=1}^n e_i^2(j) - (\sum_{i=1}^n e_i(j))^2}\sqrt{n\sum_{i=1}^n e_i^2(j') - (\sum_{i=1}^n e_i(j'))^2}} \qquad (25)$$

Note that we have said nothing about the estimation procedure used to obtain the residuals. Ordinary least squares and/or robust alternatives are both potentially valid candidate estimation methods.

To obtain a plug-in estimator $\mathrm{IC}_n(X|P)$ of the vector influence curve, one simply replaces each moment by its empirical analog and then replaces $\epsilon(j)$ with $e(j)$. Equations (22)-(24) are written generally for any choice of variable $X(k)$. In our specific case, recalling the recursion of (5), the number of linear models which would need to be estimated for the 0-1 graph edge inclusion algorithm reduces dramatically to $2M$ once the vector k_{max} has been obtained. To reiterate for clarity purposes, for the $M = J(J-1)/2$ pairs of variables (j, j'), there exists another M-vector of variables indexed by k_{max} which represents the variables $k \in \emptyset \cup \{1, \ldots, J\} \backslash \{j, j'\}$ yielding the partial correlation coefficient with the smallest (absolute) value of the test statistics T_n, or, similarly, the largest corresponding p-value. Elements $k_{max}(m)$ include $\emptyset$ in the case where the unconditional correlation coefficient produced the largest p-value. The formulas in Equations (19) and (21) represent the m-th element of the vector influence for the i-th individual. Unconditional zero-order partial correlations can also be calculated using Equation (21) by noting that linear model estimators of $\mathrm{E}[X(j)|X(k) = \emptyset]$ and $\mathrm{E}[X(j')|X(k) = \emptyset]$ are simply the means of $X(j)$ and $X(j')$, i.e., a linear model fit with only an intercept term.

3.2 Choice of Type I Error Rate

A variety of Type I error rates may be relevant when testing for edge inclusion in the graph G. The selection of a Type I error rate is a decision which is intimately linked to the data and to the biological and statistical questions at hand. In order to inform these decisions, we present the material below.

We define a *Type I error rate* as a parameter $\theta_n = \Theta(F_{V_n,R_n})$ of the joint distribution of the numbers of Type I errors V_n and rejected hypotheses R_n. Examples of Type I error rates such that $\Theta(F_{V_n,R_n}) \in [0, 1]$ are given below.

- The *family-wise error rate* (FWER) is the probability of at least one Type I error V_n,

$$FWER \equiv \Pr(V_n > 0) = 1 - F_{V_n}(0). \tag{26}$$

- The *generalized family-wise error rate* (gFWER(k)) is the probability of at least $k + 1$ Type I errors for a user-supplied integer k,

$$gFWER(k) \equiv \Pr(V_n > k) = 1 - F_{V_n}(k). \tag{27}$$

- The *false discovery rate* (FDR) is the expected value of the proportion of Type I errors among the rejected hypotheses R_n, with $V_n/R_n = 0$ if $R_n = 0$.

$$FDR \equiv E\left[\frac{V_n}{R_n}\right] = \int q dF_{V_n/R_n}(q). \tag{28}$$

- The *tail probability for the proportion of false positives* (TPPFP(q)) is the probability that the random variable V_n/R_n exceeds a user-supplied bound $q \in (0,1)$. Again, $V_n/R_n = 0$ if $R_n = 0$.

$$TPPFP(q) \equiv \Pr\left(\frac{V_n}{R_n} > q\right) = 1 - F_{V_n/R_n}(q). \tag{29}$$

This is to say the choice of which Type I error rate to control may extend beyond the decision of whether to select the FWER or FDR (Benjamini and Hochberg, 1995), or, more generally whether to control a *tail probability* or *expected value error rate*. The gFWER(k), for example, can be thought of a relaxed version of the FWER, where some discrete number of false positives can be tolerated and probabilistically controlled. The TPPFP(q), on the other hand, can be viewed as an alternative to FDR, where rather than controlling the long-run average of the *proportion of false positives* (PFP) *among the rejections made* V_n/R_n, a bound on the PFP is probabilistically controlled. (The PFP is also sometimes referred to in the literature as the *false discovery proportion* or FDP.) Controlling the TPPFP(q) has therefore also been referred to in the literature as "exceedance control" of the random variable V_n/R_n (Genovese and Wasserman, 2004a,b), particularly under testing conditions in which the variance of V_n/R_n may increase.

3.3 Empirical Bayes Joint Multiple Testing Procedures

The empirical Bayes multiple testing procedures (EBMTPs) have been formulated and characterized elsewhere (van der Laan et al., 2005; Dudoit et al., 2008; Dudoit and van der Laan, 2008, Chapter 7). In this section, we reintroduce the methodology for illustation purposes. In particular, the mechanics of the EBMTPs will be presented in such a manner as to include the test statistics joint null distribution outlined above in Section 3.1. A similar presentation of this material can be found in Dudoit and van der Laan (2008, Procedure 7.1) and in the supplementary web material accompanying Dudoit et al. (2008).

EBMTPs are intended to control generalized tail probability and expected value Type I error rates which can be defined as a parameter of the distribution of a function $g(V_n, S_n)$ of the numbers of false positives V_n and true positives S_n (Dudoit and van der Laan, 2008). Examples of such Type I error rates include the FWER, gFWER(k), TPPFP(q) and FDR defined in Section 3.2. A central feature of the EBMTPs is the estimation of the distribution of *guessed sets of true null hypotheses* $Q_{0n}^{\mathcal{H}}$. An ubiquitous choice of model for generating such guessed sets is the *common marginal non-parametric mixture model*, where the M test statistics are assumed to follow a *common marginal non-parametric mixture distribution*,

$$T_n(m) \sim f \equiv \pi_0 f_0 + (1-\pi_0) f_1, \qquad m = 1, \ldots, M, \tag{30}$$

where π_0 denotes the *prior probability of a true null hypothesis*, f_0 the *marginal null density of the test statistics*, and f_1 the *marginal alternative density of the*

test statistics, i.e., $\pi_0 \equiv \Pr(H_0(m) = 1)$, $T_n(m)|\{H_0(m) = 1\} \sim f_0$, and $T_n(m)|\{H_0(m) = 0\} \sim f_1$.

Applying Bayes' rule to the elements comprising the test statistics mixture distribution in Equation (30) results in another parameter of interest, the *local q-value function*, i.e., the posterior probability function for a true null hypothesis $H_0(m)$, given the corresponding test statistic $T_n(m)$,

$$\pi_0(t) \equiv \Pr(H_0(m) = 1|T_n(m) = t) = \frac{\pi_0 f_0(t)}{f(t)}, \qquad m = 1, \ldots, M. \tag{31}$$

One convenient feature of this equation is that it does not require specification of the alternative distribution f_1.

In practice, the local q-value function $\pi_0(t)$ in Equation (31) is unknown, as it depends on the unknown true null hypothesis prior probability π_0, test statistic marginal null density f_0, and test statistic marginal density f. In many testing situations, the marginal null density is assumed to be known *a priori* and can be applied directly. In the case of unconditional correlation coefficients, for an estimate f_{0n} of the null density f_0, one may use the t distribution with $n - 2$ degrees of freedom or appeal to a standard normal distribution when using a Fisher transformation or normal approximation. An estimate f_n of the full density f may be obtained using (nonparametric) density estimation techniques. Finally, as in Dudoit et al. (2008), a trivial estimator π_{0n} of the prior probability π_0 of a true null hypothesis is the conservative value of one, i.e., $\pi_{0n} = 1$. Alternatively, under the assumption that the null hypotheses $H_0(m)$ are conditionally independent of the data $\mathcal{X}_n$ given the corresponding test statistics $T_n(m)$, the proportion of true null hypotheses $h_0/M \equiv \pi_0$ may be estimated less conservatively via the sum of the estimated local q-values,

$$\frac{h_{0n}^{QV}}{M} = \frac{1}{M} \sum_{m=1}^{M} \pi_{0n}(T_n(m)). \tag{32}$$

Having calculated the local q-value for each element in the vector of observed test statistics T_n, one can guess whether a given hypothesis is true by generating 0/1 *Bernoulli realizations* of the corresponding posterior probabilities. Given a vector of null test statistics T_{0n}, a corresponding vector H_{0n} of guessed true null hypotheses will partition T_{0n} into two sets of test statistics over which to count the numbers of guessed false positives V_n and guessed true positives S_n for a given cut-off c. The purpose of local q-values having been established, we can now proceed with the presentation of the EBMTP algorithm (Figure 2).

EMPIRICAL BAYES JOINT MULTIPLE TESTING PROCEDURE

1. Collect data $\mathcal{X}_n \equiv \{X_i : 1, \ldots, n\}$. The data may represent J-dimensional random vectors such that $X = (X(j) : j = 1, \ldots, J) \sim P \in \mathbb{R}^J$.

2. Calculate observed test statistics $T_n = (T_n(m) : m = 1, \ldots, M)$. Note the possible change in dimension. For example, for tests of unconditional correlation coefficients, $M = J(J-1)/2$.

3. Randomly draw B vectors of null test statistics T_{0n} from a suitable test statistics (joint) null distribution Q_0 (or estimator thereof Q_{0n}) as given, for example, by $\mathrm{N}(0, \sigma_n^* = \hat{\Sigma}^*(P_n))$ in Section 3.1. Create a $M \times B$ matrix T_{0n}^B of null test statistics.

4. Apply kernel density estimation over the vector of observed test statistics T_n to obtain f_n, an estimate of the test statistics density f.

5. Armed with additional estimators (i) f_{0n} of the test statistics marginal null distribution f_0 (e.g., $\mathrm{N}(0,1)$, t_ν or a kernel density estimate obtained by pooling the elements of T_{0n}^B), and (ii) π_{0n} of the prior probability of a hypothesis being true (see above), evaluate the local q-value function (Equation (31)) at T_n to obtain *local q-values* $\pi_{0n}(m)$. Bound these probabilities by taking $\min(\pi_{0n}(m), 1)$.

6. Given an M-vector of estimated local q-values π_{0n}, generate M *Bernoulli realizations*, i.e., draws from $Bern(\pi_{0n}(m))$, indicating *guessed sets of true and false null hypotheses*. For simplicity, do this B times.

   N.B. – The end products of Steps (3) and (6) comprise B pairs $(T_{0n}^b, \mathcal{H}_{0n}^b)$ of null test statistics and guessed sets of true null hypotheses. Generating these pairs completes the portion common to all EBMTP applications. Steps (1)-(6) do not depend on choice of Type I error rate or on choice of summary measure (i.e., cut-off or adjusted p-value). To emphasize this point, these steps are shown in gray and are found to the left of the dashed vertical line in Figure 3.3.

7. For each choice of candidate cut-off c (or, if one wishes to calculate *adjusted p-values*, for each observed test statistic $T_n(m)$), calculate for each pair $(T_{0n}^b, \mathcal{H}_{0n}^b)$, the number of *guessed false positives* V_n. Similarly, for each pair $(T_n, \mathcal{H}_{0n}^b)$, also calculate the number of *guessed true positives* S_n. Specifically, these numbers are defined as

$$V_n^b(c) = \sum_{m \in \mathcal{H}_{0n}^b} \mathrm{I}(T_{0n}^b(m) > c), \text{ and} \tag{33}$$

$$S_n^b(c) = \sum_{m \notin \mathcal{H}_{0n}^b} \mathrm{I}(T_n(m) > c).$$

   In practice, for a rich collection of candidate cut-offs (observed test statistics), this step generates matrices V_n^B and S_n^B, with rows equal to the number of cut-offs and B columns.

8. Apply the g-function $g(V_n, S_n)$, corresponding to choice of Type I error rate. For the FDR, this function equals $V_n/(V_n+S_n)$. For tail probability error rates, this step involves applying an indicator function whose value is one when a threshold is crossed (e.g., $k \geq 0$ for gFWER(k) or $q \in (0,1)$ for TPPFP(q)). In the simplest case of FWER, this is $\mathrm{I}(g(V_n, S_n) \geq 0) = \mathrm{I}(V_n \geq 0)$.

9. Average the values obtained from applying the g-function over the B samples. These row averages correspond to the estimated Type I error rate for a given cut-off. We select as *fine-tuned common cut-off* the least conservative cut-off which controls Type I error at *nominal level* α. For adjusted p-values, the row means are estimates of the smallest Type I error rate (e.g., FWER, gFWER(k), TPPFP(q), FDR) of the EBMTP at which one would reject the corresponding null hypothesis. Because there is no guarantee that these adjusted p-values are monotonic, we enforce monotonicity by taking the minimum adjusted p-value over sequentially smaller nested subsets of hypotheses ordered in terms of *evidence against the null hypothesis*.

4 Application: *Arabidposis thaliana* Isoprenoid Pathway Genetic Networks

The following section highlights the application of our methods to a dataset from the plant model organism *Arabidopsis thaliana*. The data were used by Wille et al. (2004) and Wille and Bühlmann (2006) to study lower-order conditional independence graphs. The data are publicly available electronically from the publisher of Wille et al. (2004).

4.1 *A. thaliana* Isoprenoid Pathway Dataset

The *A. thaliana* dataset is targeted towards examining the regulatory network within and genetic crosstalk between two isoprenoid *biochemical pathways* found in higher plants. Note that we distinguish the term biochemical pathway from any terms such as network, graph, or even a path within a graph G. The biochemical pathway represents the ordered sequence of key molecules known to be present and/or produced at each step in a series of intracellular chemical reactions.

Isoprenoids serve various central roles in plant biochemistry, functioning as membrane components, photosynthetic pigments, or hormones (Rodriguez-Concepcion and Boronat, 2002, cited in Wille et al. (2004)). Two distinct isoprenoid pathways exist. The mevalonate (MVA) pathway occurs in the cytosol, while the non-mevalonate (MEP) pathway is confined to the chloroplast.

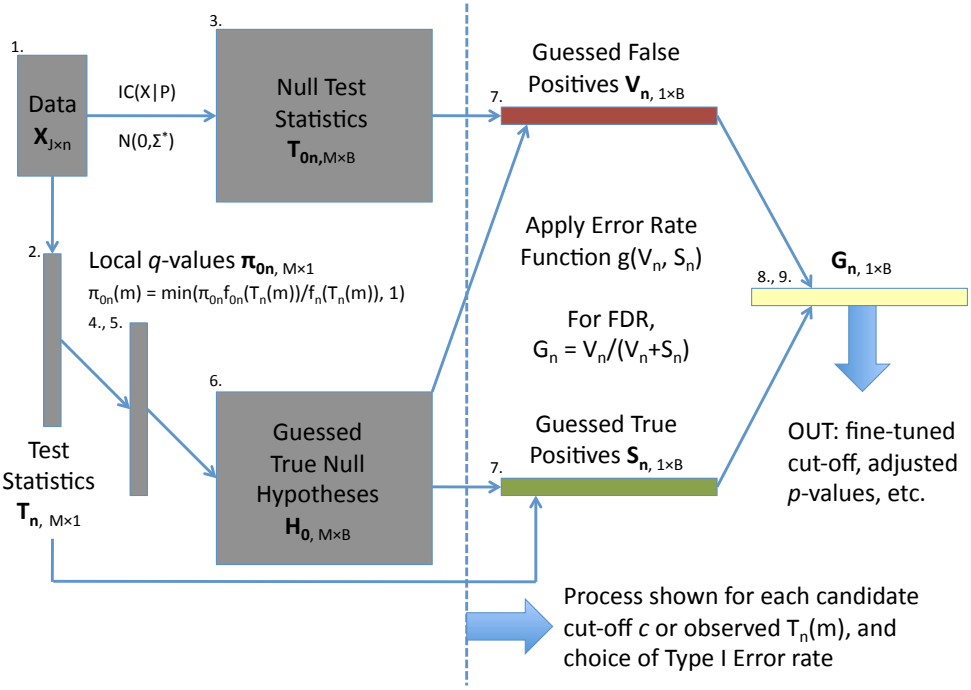

Figure 2: *Flowchart highlighting the algorithm behind joint empirical Bayes multiple testing procedures.* The diagram has been tailored to incorporate the test statistics joint null distributions based on influence curves proposed in Section 3.1. Generalized EBMTPs are modular in that they can be adapted to control a variety of Type I error rates. The special case of the false discovery rate is depicted. The numbers on the figure correspond the steps of the algorithm described in Section 3.3 in the text.

Both pathways operate independently under normal conditions, although interaction between the pathways has also been reported (Laule et al., 2003; Rodriguez-Concepcion et al., 2004, cited in Wille et al. (2004)).

The data consist of expression measurements for $J = 40$ genes assayed on $n = 118$ Affymetrix GeneChip microarrays. The 118 chips represent expression measurements from plants grown under various experimental conditions (see supplementary data, Wille et al. (2004)). Of these 40 genes, 16 were assigned to the cytosolic MVA pathway, 19 to the plastidial MEP pathway, and the remaining 5 represent mitochondrial proteins. One protein associated with the MEP pathway, GGPPS7, is absent from the data file and from the figures in Wille et al. (2004) and Wille and Bühlmann (2006), and was therefore excluded from our analysis. The data have been preprocessed elsewhere, and they are complete, i.e., there are no missing expression measurements. Depending on

the abundance of missing data one may encounter in other settings, one may wish to impute missing observations for the calculation of variances, covariances and correlations, e.g., by k-nearest neighbors imputation (Troyanskaya et al., 2001) available in the R package `impute` (Hastie et al., 2001). Finally, the data were $\log_2$-transformed before proceding with the analysis.

4.2 Multiple Testing Procedures

To continue with the analysis, we will proceed by roughly adopting the algorithm of Wille and Bühlmann (2006) presented in Section 2.3.2 for detecting edges in a 0-1 graph. The main contribution and focus of our analysis will be our handling of Step 3 in the algorithm, where an estimate of the test statistics joint null distribution will be employed in conjunction with the EBMTPs introduced in Section 3.3. As stated above, the multiple testing procedures used in this analysis illustrate a data-adaptive sequential testing scenario in which the role of the M-vector k_{max} is to effectively reduce the dimensionality of the testing problem from two to one dimensions. Therefore, Type I error control is guaranteed only when holding the variables indexed by k_{max} fixed.

4.2.1 Null and Alternative Hypotheses

The hypotheses of interest concern the $M \equiv \binom{J}{2} = J(J-1)/2$ lower-order partial correlation coefficients $\psi(m) = \psi(j, j'|k) = \rho(j, j'|k)$. Given a vector of indices $k_{max} = (k_{max}(m) = k_{max}(j, j') : k \in \emptyset \cup \{1, \ldots, J\} \backslash \{j, j'\})$ described above, the index triplets range over $\{(j, j'|k) : j = 1, \ldots, J-1, j' = j+1, \ldots, J, k = k_{max}(j, j')\}$.

Consider two-sided tests of the $M = J(J-1)/2$ null hypotheses $H_0(m) = \mathrm{I}(\psi(m) = \psi_0(m))$ vs. the alternative hypotheses $H_1(m) = \mathrm{I}(\psi(m) \neq \psi_0(m))$. For simplicity and without loss of generality, the null value $\psi_0(m)$ is taken to be zero. The null value $\psi_0(m) = 0$ is also biologically meaningful in our context.

4.2.2 Test Statistics

The M null hypotheses can be tested based on the following t-statistics,

$$T_n(m) \equiv \sqrt{n - \mathrm{I}(k_{max}(m) \in V \backslash \{j, j'\}) - 2} \frac{\psi_n(m)}{\sqrt{1 - \psi_n(m)}}, \qquad m = 1, \ldots, M, \tag{34}$$

where $\psi_n = (\psi_n(m) : m = 1, \ldots, M)$ denotes the M-vector of empirical partial correlation coefficients $\rho(j, j'|k)$ for the pair $(X(j), X(j'))$ conditional on $X(k) = X(k_{max}(j, j'))$.

For Gaussian data generating distributions, the t-statistics of Equation (34) will have marginal t-distributions with $(n-2)$ degrees of freedom when $k_{max}(m)$ is $\emptyset$ and $(n-3)$ degrees of freedom when $k_{max}(m)$ corresponds to a true variable index, i.e., when computing a first-order partial correlation.

Additionally, one may appeal to Fisher's z-transformation as given in Equation (13) to calculate transformed test statistics $Z_n = (Z_n(m) : m = 1, \ldots, M)$ which have asymptotic N(0,1) marginal distributions. In the case of the 0-1 graph, using Fisher's transformation has the added benefit of guaranteeing from the outset that the marginal distribution of the test statistics remains constant over M.

4.2.3 Test Statistics Null Distributions

Marginal null distribution for t-statistics $T_n(m)$ and transformed test statistics $Z_n(m)$ have been described above. For large values of n, one may also appeal to the central limit theorem and choose N(0,1) as a marginal null distribution for t-statistics of the form T_n given in Equation (34).

We propose two estimates of the test statistics joint null distribution. The first estimate is the one obtained using influence curves as set forth in Section 3.1.4. In this case, the test statistics joint null distribution is multivariate Gaussian and estimated by $Q_{0n}^t = \mathrm{N}(0, \sigma_n^*)$, where σ_n^* is the correlation matrix corresponding to the variance-covariance matrix of the vector influence curve, $\sigma_n = \hat{\Sigma}(P_n)$ as defined in Equation (16). An $M \times B$ matrix with columns representing B realizations from $\mathrm{N}(0, \sigma_n^*)$ serves (i) as an estimate Q_{0n}^t of the test statistics joint null distribution Q_0^t and (ii) as draws of null test statistics T_{0n} for the EBMTPs to follow.

A second choice of test statistics null distribution may be achieved by building upon the first joint null distribution described in the above paragraph. Here, one may choose to use a matrix of null quantile-transformed test statistics Z_n^B (van der Laan and Hubbard, 2006; Dudoit and van der Laan, 2008, Section 2.4). Applying this transformation will preserve the estimation of the joint dependencies from the multivariate normal distribution (down columns), but will map the null test statistics back to a user-specified marginal null distribution (across rows) that one may have chosen in the univariate testing scenario, i.e., t_{n-2} or t_{n-3}. To clarify, this choice of test statistics null distribution will transform the multivariate null distribution given above by $\mathrm{N}(0, \sigma_n^*)$ to have marginal t-distributions.

In all, three different combinations of test statistics and their respective estimated joint null distributions were used in the analysis: (i) Fisher's z-statistics with $\mathrm{N}(0, \sigma_n^*)$, (ii) t-statistics with $\mathrm{N}(0, \sigma_n^*)$, and (iii) t-statistics with Z_n^B, the null-quantile transformed version of $\mathrm{N}(0, \sigma_n^*)$ with t_{n-3} chosen as marginal null distribution (Table 1). Differences between the three cases may provide insight into the finite-sample behavior of the procedures with respect to this particular dataset. Case (i) represents normal test statistics paired with a normal test statistics joint null distribution. Case (ii) represents the use of t-statistics paired with a normal approximation for choice of (marginal and joint) null distribution, which, with $n = 118$ samples, may be appropriate. Finally, case (iii) illustrates an approach where one may still wish to use t-statistics to test lower-order partial correlation hypotheses, and in which

one desires an estimate of the test statistics joint null distribution to maintain marginal t-distributions.

4.2.4 FWER- and FDR-Controlling Multiple Testing Procedures

We chose to control both the family-wise error rate (FWER) and the false discovery rate (FDR). Marginal MTPs such as the Bonferroni procedure for FWER control (Bonferroni, 1936) and the Benjamini-Hochberg (BH) procedure (Benjamini and Hochberg, 1995) for FDR control have been well described in the literature. We will assume the reader is familiar with these MTPs and move on to describe the joint MTPs used in the analysis. Both the Bonferroni and Benjamini-Hocberg procedures are easily implemented using the function `mt.rawp2adjp()` available in the R/Bioconductor package `multtest` (Pollard et al., 2005, `http://www.bioconductor.org`).

FWER-Controlling Procedures

Single-step maxT procedure. Adjusted p-values for a single-step maximum test statistic (ss maxT) procedure (Dudoit and van der Laan, 2008, Procedure 3.5) may be calculated using either null distribution described above. Specifically, adjusted p-values are obtained from the empirical distributions of the B maxima of (absolute) null test statistics $\{\max_m T_{0n}^B(m,b) : b = 1\ldots,B\}$ or null-transformed test statistics $\{\max_m Z_n^B(m,b) : b = 1\ldots,B\}$. For a test at nominal FWER level α, one rejects null hypotheses with adjusted p-values less than or equal to α.

Empirical Bayes single-step maxT FWER-controlling procedures. The single-step maxT procedure can be accommodated to work with the EBMTP framework described in Section 3.3. For FWER control, the function $g(V_n, S_n) = V_n$ is selected for the distribution of the number of false positives, and the tail probability error rate $\Pr(V_n > 0)$ is controlled at level α. Rather than calculating maxima over all M hypotheses, the FWER-controlling empirical Bayes procedures restrict calculating the B maxima over those null test statistics $T_{0n} \in \mathcal{H}_{0n}^b$ which correspond to null hypotheses guessed as belonging to the set of true null hypotheses $\mathcal{H}_0$. In this sense, the EBMTPs controlling the FWER are expected to be more powerful than their vanilla joint MTP counterparts as more extreme test statistics are more likely to be assigned as belonging to the set of true alternative hypotheses $\mathcal{H}_1$. Specifically, adjusted p-values are obtained from the empirical distributions of the B restricted maxima of (absolute) null test statistics $\{\max_{m\in\mathcal{H}_{0n}^b} T_{0n}^B(m,b) : b = 1\ldots,B\}$ or null-transformed test statistics $\{\max_{m\in\mathcal{H}_{0n}^b} Z_n^B(m,b) : b = 1\ldots,B\}$.

We will denote the empirical Bayes FWER-controlling procedure which uses the most conservative estimate $\pi_{0n} = 1$ of the probability that a null hypothesis is true as EB-FWER(1). As in Equation (32), choosing a less conservative plug-in estimate π_{0n} of the prior probability π_0 may produce a more

powerful MTP as maxima will be calculated over even further restricted subsets of null test statistics. We will denote this procedure by EB-FWER(LQV), as the less conservative estimate of π_0 is a function of the local q-values obtained through density estimation.

FDR-Controlling Procedures

Empirical Bayes common cut-off FDR-controlling procedures. For empirical Bayes FDR control, the function $g(V_n, S_n) = V_n/(V_n + S_n)$, and the expected value error rate $\mathrm{E}[g(V_n, S_n)]$ is controlled at level α. A common cut-off may be derived to control the Type I error rate. For a rich vector of candidate cut-offs, e.g., $c \in \{2.00, 2.05, \ldots, 4.00\}$, the common cut-off c_n is the *smallest* (i.e., least conservative) value in c which satisfies the expected value error constraint. That is,

$$c_n \equiv \inf \left\{ c \in \mathbb{R} : \frac{1}{B} \sum_{b=1}^{B} \frac{V_n^b(c^{(M)})}{V_n^b(c^{(M)}) + S_n^b(c^{(M)})} < \alpha \right\}, \tag{35}$$

where $c^{(M)}$ denotes the cut-off M-vector with all elements equal. Adjusted p-values may be obtained in a similar manner as above when treating the vector of observed test statistics as cut-offs themselves. As mentioned in Section 3.3, a monotonicity constraint is placed on the adjusted p-values by considering row minima over sequentially smaller subsets of test statistics ordered in terms of their evidence against the null hypothesis.

We will denote the empirical Bayes FDR-controlling procedure which uses the most conservative estimate $\pi_{0n} = 1$ of the probability that a null hypothesis is true as EB-FDR(1). Because V_n^b and S_n^b both depend on the set of guessed true null hypotheses $\mathcal{H}_{0n}^b$, again, as in Equation 32, choosing a less conservative plug-in estimate π_{0n} of the prior probability π_0 may produce a more powerful MTP. We will denote this procedure by EB-FDR(LQV), as the less conservative estimate of π_0 is a function of the local q-values obtained through density estimation.

4.3 Results

4.3.1 Edge Identification in the 0-1 Graph

When identifying the elements of the vector k_{max}, none of the M least significant test statistics were obtained using the unconditional empirical correlation coefficients. This means that if one wishes to use t-statistics to identify edges in the lower-order conditional graph, using the algorithm proposed by Wille and colleagues will result in probing only first-order partial correlations with t_{n-3} distributions. In this case, the common-cut-off procedures described above remain valid MTPs for use in this setting. (Of course, combining t_{n-2} and t_{n-3}

distributions may not have produced very different end-results, although procedures based on minimum p-values or other *common-quantile* methods may have been more appropriate (Dudoit and van der Laan, 2008).)

When performing density estimation in the EBMTP analysis, the estimated proportion of true null hypotheses h_{0n}/M (i.e., the estimated prior probability π_{0n} of a null hypothesis being true) was calculated to be 0.52. Given that this estimate is sufficiently less than the conservative estimate of one used in the EB-FWER(1) and EB-FDR(1) procedures, it may be justifiable (and even desirable) to relax the prior and also apply the less conservative EB-FWER(LQV) and EB-FDR(LQV) procedures in this setting.

Joint distribution estimates were generated using $B = 5000$ samples from $\mathrm{N}(0, \sigma_n^*)$. Results did not change appreciably when repeating the analysis several times nor for larger values of B, specifically $B = 10,000$ and $B = 25,000$ (data not shown).

As expected, minor differences were observed in terms of the number of rejections across the various combinations of test statistics and test statistic joint null distributions used in the analysis (Table 1). Focusing on control of the FWER, slightly more hypotheses were rejected using the joint multiple testing procedures than the marginal Bonferroni correction. The empirical Bayes procedures performed similarly to the ss maxT procedure, with the EB-FWER(LQV) procedure sometimes rejecting just one more hypothesis than either of the other joint multiple testing procedures. In the case of FDR control, across all columns of Table 1, the EB-FDR(1) procedure rejected more hypotheses than BH, while the EB-FDR(LQV) procedure resulted in the largest number of edges called significant (Table 1).

Focusing on the case in which Fisher's z-statistics were combined with the test statistics joint null distribution obtained from sampling repeatedly from $\mathrm{N}(0, \sigma_n^*)$, it is possible to compare our results with those reported by Wille et al. (2004) and Wille and Bühlmann (2006). Figure 3 is adapted from the papers of Wille and colleagues and demonstrates high levels of reproducibility between the two analyses using Fisher z-statistics and likelihood ratio tests as was done in the other analyses. The additional edges called significant by the EB-FDR(1) procedure (relative to BH) include, in rank-order: GGPPS11 — UPPS1, CMK — DXR, GGPPS2 — GGPPS9, AACT1 — IPPI1, and GGPPS11 — MECPS (Figure 3, solid red edges). Specifically, the edges GGPPS11 — UPPS1 and GGPPS2 — GGPPS9 show evidence of cross-talk between proteins in the plastidial MEP pathway and the mitochondrion, while the edge AACT1 — IPPI1 shows an association between a late product in the MEP pathway with an early product in the cytosolic MVA pathway. The other edges CMK — DXR and GGPPS11 — MECPS connect members of the same (plastidial) MEP pathway. The additional edge called significant by the EB-FDR(LQV) procedure (relative to EB-FDR(1)) was GGPPS6 — GGPPS12 (Figure 3, dashed red edge). This edge connects two later products also located within the MEP pathway.

Test statistics	Fisher's z	t-statistics	t-statistics
Joint null dist.	$\mathrm{N}(0, \sigma_n^*)$	$\mathrm{N}(0, \sigma_n^*)$	Z_n^B
Marginal null dist.	$\mathrm{N}(0, 1)$	$\mathrm{N}(0, 1)$	t_{n-3}

Procedure	# rejections	# rejections	# rejections
Bonferroni	26	29	26
ss maxT	27	29	29
EB-FWER(1)	29	29	29
EB-FWER(LQV)	29	30	30
BH	47	49	46
EB-FDR(1)	52	53	51
EB-FDR(LQV)	53	54	53

Table 1: *Numbers of rejected hypotheses (significant edges) for the lower-order conditional independence graph of isoprenoid pathway genes in A. thaliana.* Both the FWER and FDR were controlled at nominal level $\alpha = 0.05$. Results are shown for various combinations of test statistics and test statistics marginal and joint distributions. The rank order of the test statistics did not change from procedure to procedure, so the number of rejections represents the point on the list of ordered test statistics where one stops rejecting hypotheses (i.e., including edges in G). To reiterate, the joint distribution given by $\mathrm{N}(0, \sigma_n^*)$ is obtained from influence curves as in Section 3.1.4. The joint distribution Z_n^B is the matrix of null quantile-transformed test statistics obtained from $\mathrm{N}(0, \sigma_n^*)$ with marginal t-distributions. Joint distribution estimates were generated using $B = 5000$ samples from $\mathrm{N}(0, \sigma_n^*)$.

4.3.2 Secondary Parameters of Interest

A byproduct of the '0-1' algorithm of Section 2.3.2 is the vector of indices k_{max} with elements $k_{max}(m) = k_{max}(j, j')$ corresponding to the variable which was ultimately conditioned on for the first-order partial correlation between pairs of variables $X(j)$ and $X(j')$. That is, the variable $X(k_{max}(j, j'))$ was the variable which could presumably best "explain away" any putative association between $X(j)$ and $X(j')$. Given both the statistical and biological importance of $X(k_{max}(j, j'))$, an interesting outcome to consider is the distribution of the number of times a given variable was conditioned upon in the M composite first-order partial correlations.

Among the 53 edges called significant by the EB-FDR(LQV) procedure (Table 1, Column 1), a total of 17 different variables were used for conditioning. Five variables, however, were used for conditioning in approximately half (26) of the most significant test statistics (Table 2). Focusing on the variable that

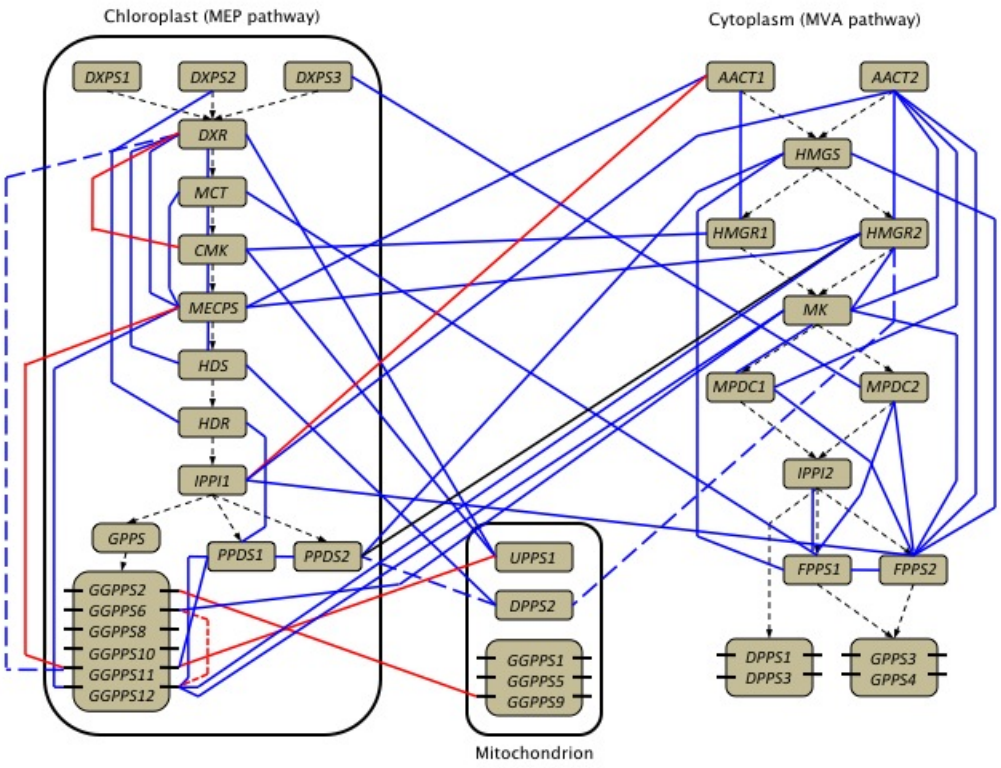

Figure 3: *Edges called significant in the 0-1 graph using the lower-order partial correlation coefficient test statistics joint null distribution of Section 3.1.4 in conjunction with empirical Bayes joint multiple testing procedures (EBMTP) of Section 3.3.* The figures shows the results when using Fisher's z test statistics. The FDR was controlled at nominal level $\alpha = 0.05$ to detect associations within and between the MEP and MVA isoprenoid pathways in *A. thaliana*. Dashed black lines with arrows represent biochemical pathways. Solid blue lines indicate edges statistically reproduced from the original figure of Wille et al. (2004) using the EBMTP framework. Dashed blue lines indicate edges called significant in the EBMTP analysis but missing from the figure of Wille et al. (2004) when controlling the FDR using the BH method (Benjamini and Hochberg, 1995). Solid red lines indicate additional edges called significant using the EBMTP(1) procdure. The dashed red line indicates the additional edges called significant using the relatively less conservative EBMTP(LQV) procedure to account for the estimated proportion of true null hypotheses h_0/M. The solid black line represents the single edge which could not be reproduced from the figure in Wille et al. (2004). Joint distribution estimates were generated using $B = 5000$ samples from $\mathrm{N}(0, \sigma_n^*)$.

was conditioned upon most often, MPDC1 appears to belong to a set of associated edges within the cytosolic pathway, as it is conditioned upon by pairs of other genes for which it itself is significantly associated (AACT2, MK, FPPS2 and MPDC1's biochemical relative, MPDC2) (Table 2, Figure 3). Interestingly, MPDC1 is conditioned upon by variables represented in two edges linking the plastidial and cytosolic pathways (AACT2 — IPPI1; GGPPS6 — MK). MPDC1 is also the conditioning variable in the edge DPPS2 — HMGR2 linking the cytosolic pathway to the mitochondrion. In fact, the next two most significant edges (DPPS1 — GGPPS8, adjusted p-value = 0.0514; DPPS3 — GGPPS1, adjusted p-value = 0.0672) also condition upon MPDC1. Both of these edges link members of the cytosolic pathway with genes assigned to chloroplast and mitochondrion, respectively. No edge exists, however, linking MPDC1 to genes or gene products associated with either cell organelle.

While there is sometimes concern that the '0-1' graph may lack certain statistical properties associated with the concentration graph, it appears there may be potentially useful biological, regulatory, and/or network-associated information contained in the vector of lower-order conditioning variables indexed by $k_{max}(j, j')$.

5 Discussion

We have formulated and implemented a means for inferring lower-order relationships in graphical models. This work has expanded on what others in the field have done by formalizing and then applying state-of-the-art multiple testing approaches in this setting. Central features of our method include (i) an asymptotically valid test statistics joint null distribution which exploits the parametric form of t-statistics useful for probing hypotheses about correlation parameters, and (ii) the application of powerful EBMTPs which effectively make use of the joint distribution in (i). Both of these features are modular in the sense that they need not be implemented in tandem. The influence curve test statistics joint null distribution can be used in conjunction with other joint MTPs, and the EBMTPs can also incorporate other estimates of the test statistics joint null distribution, such as those obtained from resampling-based methods.

The test statistics joint null distributions obtained from influence curves effectively use the data to estimate dependencies between empirical correlation coefficient test statistics in a manner which does not, for example, rely on restrictive assumptions such as subset pivotality. In addition to the cases of bidirected, 0-1, and undirected graphs, with special care being given to ensure a proper vertex well-numbering, an influence curve test statistics null distribution could be adapted to the DAG setting. Moreover, joint null distributions based on influence curves can be estimated more rapidly than null distributions estimated by other means such as permutation-based or bootstrap-based approaches. This influence curve approach therefore represents a methodolog-

Cond. Variable	Variable 1	Variable 2	Fisher's z	Adj. p
MPDC1	AACT2	MK	5.699	0.0000
	FPPS2	MK	4.593	0.0002
	FPPS2	MPDC2	4.421	0.0003
	AACT2	IPPI1	4.007	0.0012
	GGPPS6	MK	3.555	0.0055
	DPPS2	HMGR2	-3.181	0.0156
	AACT2	FPPS2	3.066	0.0213
MECPS	AACT1	HMGR1	4.895	0.0000
	DXR	HDS	4.535	0.0002
	DXR	GGPPS11	3.755	0.0030
	CMK	HMGR1	-3.635	0.0043
	GGPPS12	HMGR2	-3.241	0.0135
AACT2	FPPS2	MPDC1	6.928	0.0000
	MK	MPDC1	3.996	0.0012
	HMGR2	MK	3.031	0.0229
	FPPS2	IPPI1	3.008	0.0237
	AACT1	IPPI1	2.900	0.0304
DXR	CMK	MCT	6.952	0.0000
	HDS	MECPS	5.927	0.0000
	CMK	UPPS1	-3.259	0.0131
	GGPPS11	UPPS1	-2.981	0.0253
	GGPPS11	MECPS	2.853	0.0343
CMK	DXR	UPPS1	-6.368	0.0000
	DXR	MCT	5.726	0.0000
	FPPS1	MCT	4.422	0.0003
	MCT	MECPS	3.389	0.0092

Table 2: *Significant edges grouped by the variables which they were conditioned upon in '0-1' graph algorithm.* The frequency with which a variable was conditioned upon among the full set of hypotheses differs slightly when compared to the frequency among the list of rejections (data not shown). Of the three next-most significant edges (adjusted p-value $\in (0.05, 0.07)$), two conditioned on MPDC1 and one more conditioned on MECPS. The full list of rejected hypotheses sorted by conditioning variable is available in the supplementary material.

ical advance in the field as practitioners often cite the computational burden of resampling methods as a barrier to using more elaborate joint MTPs.

Joint MTPs, i.e., testing procedures which can effectively make use of a test statistics joint null distribution (or estimator thereof), further allow one to more effectively glean information from the data during the analysis phase. The generalized EBMTP framework presented in this paper is an example of

such a family of powerful methods designed for use with complex multivariate distributions encountered in genomic and other high-dimensional settings. As stated above, EBMTPs have the added benefit that they are modular in design. The generalized EBMTPs, for example, can be applied to quickly control a variety of Type I error rates without the need for additional rounds of (re)sampling for each new choice of error rate, i.e., one sample of B draws from $\mathrm{N}(0, \sigma^*)$ would suffice for controlling any or all of the FWER, gFWER(k), TPPFP(q) or FDR. The EBMTPs could also be combined with other estimates of the (inverse) variance-covariance matrix such as those obtained from shrinkage approaches or test statistics null distributions for concentration graph models in the $n > p$ setting (Drton and Perlman, 2007, for example).

6 Software and Supplementary Material

The multiple testing methods described above, including null distributions based on influence curves and the empirical Bayes joint multiple testing procedures, have been included in the April 2009 release of the R/Bioconductor package `multtest` (Pollard et al., 2005, `http://www.bioconductor.org`). Similar to the original main user-level function `MTP` (for performing resampling-based multiple hypothesis testing, such as the `ss maxT` procedure), the empirical Bayes procedures have been implemented in the function `EBMTP`, which main output is returned in the form of adjusted p-values. For several choices of tests involving t-statistics, both functions allow for the use of influence curve null distributions by setting the argument `nulldist='ic'`. Note that in `multtest`, for the density estimation Steps (4) and (5) of the EBMTP algorithm of Section 3.3, the full density of the test statistics is obtained by applying kernel density estimation over the (absolute value of the) vector of observed test statistics, while the null density is obtained by applying kernel density estimation over the (absolute values of the) pooled elements of the matrix of null test statistics T_{0n}^B with column vectors sampled from $\mathrm{N}(0, \sigma_n^*)$.

Arguments for calculating unconditional pairwise correlation test statistics are available in `multtest`, however, first-order partial correlation test statistics have not yet been implemented. A file with code implementing the '0-1' algorithm of Section 2.3.2 is available at `http://www.stat.berkeley.edu/~houston/CorrTestStatsICNulldist.R`. The file's main function, `IC.corr`, returns a list containing, among other objects, a vector of observed test statistics (either t-statistics or Fisher's z-statistics) and an estimate of the test statistics null distribution derived from influence curves as in Equations (19) and (21). An example illustrating how to interface these objects with the `multtest` functionality is given in the code comments.

Test statistics joint null distributions based on influence curves may be obtained faster than those estimated via other resampling-based approaches such as permutation or bootstrap methods. To illustrate this point, the non-parametric bootstrap was used to obtain an estimate Q_{0n} of the joint null

distribution Q_0. In this set-up, the data (i.e., J-vectors of expression measures) are resampled with replacement $B = 5000$ times. For each bootstrap sample, the test statistics for the first-order partial correlations conditional on the variables indexed by k_{max} (fixed, see above) were calculated using the recursion given in Equation (5). The raw bootstrap test statistics were then transformed using the null quantile transformation of van der Laan and Hubbard (2006) and Dudoit and van der Laan (2008, Procedure 2.4). Run time comparisons were performed on a Macintosh laptop computer running R 2.7.2 (R Development Core Team, 2008, `http://www.R-project.org`) on Mac OS X Version 10.5.5 with 2GB of RAM and a 2.4GHz Intel Core 2 Duo processor.

Once the vector of indices k_{max} was obtained, the influence curve approach, written entirely in R, required 12.90 seconds to return a 741×5000 matrix of null test statistics. Of this time, 3.46 seconds were dedicated to a call to `mvrnorm()` in the MASS library for generating correlated test statistics (Venables and Ripley, 2002). In contrast, the nonparametric bootstrap approach, conditional on the same vector of indices in k_{max}, which utilized optimized R internals driving the `cor()` function, required 92.89 seconds to obtain a similar matrix of null test statistics, of which only 8.31 seconds were dedicated to the quantile transform. Of course, the more appropriate resampling scheme would involve calculating a new vector k^b_{max} for each bootstrap sample. In this case, approximately 1 hour and 45 minutes were required to obtain the matrix of null test statistics. In many practical applications, test statistics may not be so easily obtainable. Without the aid of the recursion, for example, a bootstrap-based estimate of Q_0 could have been obtained by calculating the correlation between residuals obtained from OLS models using the `lm()` command as described in Section 3.1.4. Implemented entirely in R, this approach would have required several days on the same machine to produce another 741×5000 matrix of null test statistics.

The full text version of Table 2 can be found at `http://www.stat.berkeley.edu/~houston/kMaxResTab.txt`.

Acknowledgments

The authors wish to thank Mathias Drton and Peter Bühlmann for helpful discussions and correspondence.

References

S. Aburatani, F. Sun, S. Saito, M. Honda, S. Kaneko, and K. Horimoto. Gene systems network inferred from expression profiles in hepatocellular carcinogenesis by graphical Gaussian model. *EURASIP J. Bioinform. Syst. Biol.*, page 47214, 2007.

T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley

Series in Probability and Statistics, Wiley-Interscience, Hoboken, NJ, 3rd edition, 2003.

Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57(1):289–300, 1995.

N. Bing and I. Hoeschele. Genetical genomics analysis of a yeast segregant population for trnascription network analysis. *Genetics*, 170:533–542, 2005.

C. E. Bonferroni. Teoria statistica delle classi e calcolo delle probabilita. In *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, pages 3–62. 1936.

A. J. Butte, P. Tomayo, D. Slonim, T. R. Golub, and I. S. Kohane. Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proc Nat Acad Sci USA*, 97(22): 12182–12186, 2000.

R. Castelo and A. Roverato. A robust procedure for Gaussian graphical model search from microarray data with p larger than n. *Journal of Machine Learning Research*, 7:2621–2650, 2006.

E. J. Chessler, L. Lu, S. Shon, Y. Qu, J. Gu, J. Wang, H. C. Hsu, J. D. Mountz, N. E. Baldwin, M. A. Langston, D. W. Threadgill, K. F. Manly, and R. W. Williams. Complex trait analysis of gene expression networks that modulate nervous system function. *Nat. Genetics*, 2005.

D. R. Cox and N. Wermuth. Linear dependencies represented by chain graphs. *Statistical Science*, 8:204–18, 1993.

D. R. Cox and N. Wermuth. *Multivariate Dependencies*. Chapman & Hall, London, 1996.

L. de Campos and J. Huete. A new approach for learning belief networks using independence criteria. *International Journal of Approximate Reasoning*, 24: 11–37, 2000.

A. de la Fuente, N. Bing, I. Hoeschele, and P. Mendes. Discovery of meaningful associations in genomic data using partial correlation coefficients. *Bioinformatics*, 20(18):3565–3574, 2004.

A. P. Dempster. Covariance selection. *Biometrics*, 28:157–75, 1972.

A. Dobra, C. Hans, B. Jones, J. R. Nevins, G. Yuo, and M. West. Sparse graphical models for exploring gene expression data. *Journal of Multivariate Analysis*, 90:196–212, 2004.

M. Drton and M. D. Perlman. Multiple testing and error control in Gaussian graphical model selection. *Statistical Science*, 22(3):430–49, 2007.

M. Drton and M. D. Perlman. A SINful approach to Gaussian graphical model selection. *Journal of Statistical Planning and Inference*, 138(4):1179–200, 2008.

M. Drton and M. D. Perlman. Model selection for Gaussian concentration graphs. *Biometrika*, 91(3):591–602, 2004.

S. Dudoit and M. J. van der Laan. *Multiple Testing Procedures and Applications to Genomics*. Springer, 2008.

S. Dudoit, M. J. van der Laan, and K. S. Pollard. Multiple testing. Part I. Single-step procedures for control of general type I error rates. *Statistical Applications in Genetics and Molecular Biology*, 2004.

S. Dudoit, H. N. Gilbert, and M. J. van der Laan. Resampling-based empirical Bayes multiple testing procedures for controlling generalized tail probability and expected value error rates: Focus on the false discovery rate and simulation study. *Biometrical Journal*, 50(5):716–44, 2008. URL `http://www.stat.berkeley.edu/~houston/BJMCPSupp/BJMCPSupp.html`.

D. Edwards. *Introduction to Graphical Modelling*. Springer Verlag, 2000.

N. Friedman. Inferring cellular networks using probabilistic graphical models. *Science*, 303:799–805, 2004.

N. Friedman, M. Linial, I. Nachman, and D. Pe'er. Using Bayesian networks to analyze expression data. *Journal of Computational Biology*, 86:785–801, 2000.

C. R. Genovese and L. Wasserman. Exceedance control of the false discovery proportion. Technical Report 807, Department of Statistics, Carnegie Mellon University, http://www.stat.cmu.edu/tr/tr807/tr807.html, July 2004a.

C. R. Genovese and L. Wasserman. A stochastic process approach to false discovery control. *Annals of Statistics*, 32(3):1035–1061, 2004b.

R. D. Gill. Non- and semi-parametric maximum likelihood estimator and the von Mises method. *Scandinavian Journal of Statistics*, 16(2):97–128, 1989. With a discussion by J.A. Wellner and J. Praestgaard and a reply by the author.

R. D. Gill, M. J. v. d. Laan, and J. A. Wellner. Inefficient estimators of the bivariate survival function for three models. *Annales de l'Institut Henri Poincarè. Probabilitè et Statistiques*, 31(3):545–97, 1995.

T. R. Golub, D. K. Slonim, P. Tomayo, C. Huard, M. Gassenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caliguri, C. D. Bloomfield, and E. S. Lander. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science*, 286:531–537, 1999.

A. J. Hartemink, D. K. Gifford, T. S. Jaakkola, and R. A. Young. Using graphical models and genomics expression data to statistically validate models of genetic regulatory networks. In *Pac Symp Biocomput* PSB01, pages 422–33, 2001.

T. Hastie, R. Tibshirani, B. Narasimhan, and G. Chu. *impute: Imputation for microarray data*, 2001. R package version 1.15.1.

S. Holm. A simple sequentially rejective multiple testing procedure. *Scand. J. Statist.*, 6:65–70, 1979.

D. Husmeier. Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. *Bioinformatics*, 19(l7):2271–82, 2003.

M. I. Jordan. An introduction to probabilisitic graphical models. Unpublished lecture notes for Statistics C241A and Computer Science 281A, "Introduction to Statistical Learning Theory", University of California, Berkeley, Spring 2007, 2003.

M. Kalisch and P. Bühlmann. Estimating high-dimensional directed acyclic graphs with the PC-algorithm. *Journal of Machine Learning Research*, 8: 613–36, 2007.

O. Laule, A. Fürholz, H. S. Chang, T. Zhu, X. Wang, P. B. Heifetz, W. Gruissem, and M. Lange. Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis. *Proc. Natl. Acad. Sci.*, 100:6866–71, 2003.

S. Lauritzen. *Graphical Models*. Oxford University Press, 1996.

S. Lèbre. Inferring dynamic genetic networks with low order dependencies. *Statistical Applications in Genetics and Molecular Biology*, 8(1), 2009. URL `http://www.bepress.com/sagmb/vol8/iss1/art9`.

S. Lee, D. Pe'er, A. M. Dudley, G. M. Church, and D. Koller. Indentifying regulatory mechanisms using individual variation reveals key role for chromatin modification. *Proc Nat Acad Sci USA*, 103(38):14062–14067, 2006.

H. Li and J. Gui. Gradient directed regularization for sparse Gaussian concentration graphs, with applications to inference of genetic networks. *Biostatistics*, 7(2):302–317, 2006.

S. Ma and H. J. Bohnert. Gene networks in *arabidopsis thaliana* for metabolic and environmental function. *Molecular Biosystems*, 4(3):199–204, 2008.

S. Ma, Q. Gong, and H. J. Bohnert. An *arabidopsis* gene network based on the Gaussian graphical model. *Genome Research*, 17(11):1614–25, 2007.

P. M. Magwene and J. Kim. Estimating genomic coexpression networks using first-order conditional independence. *Genome Biology*, 5:R100, 2004.

T. Matsuno, N. Tominaga, K. Arizono, T. Iguchi, and Y. Kohara. Graphical Gaussian modeling for gene association structures based on expression deviation patterns induced by various chemical stimuli. *IEICE Trans Inf and Sys*, E89-D(4):1563–1574, April 2006.

N. Meinshausen and P. Bühlmann. High-dimensional graphs with the Lasso. *Annals of Statistics*, 34:1436–1462, 2006.

R. Opgen-Rhein and K. Strimmer. From correlation to causation networks: A simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC Systems Biology*, 1:37, 2007.

K. S. Pollard and M. J. van der Laan. Choice of a null distribution in resampling-based multiple testing. *Journal of Statistical Planning and Inference*, 125:85–100, 2004.

K. S. Pollard, Y. Ge, S. Taylor, H. N. Gilbert, and S. Dudoit. *multtest: Resampling-based multiple hypothesis testing*, 2005. R package version 1.20.0.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. URL `http://www.R-project.org`. ISBN 3-900051-07-0.

M. Rodriguez-Concepcion and A. Boronat. Elucidation of the methylerythritol phosphate pathway for isoprenoid biosysnthesis in bacteria and plastids. A metabolic milestone achieved through genomics. *Plant Physiology*, 130:1079–89, 2002.

M. Rodriguez-Concepcion, O. Fores, J. F. Martinez-Garcia, V. Gonzalez, M. Phillips, A. Ferrer, and A. Boronat. Distinct light-mediated pathways regulate the biosynthesis and exchange of isoprenoid precursors during *arabidopsis* seedling development. *Plant Cell*, 16:144–56, 2004.

J. Schäfer and K. Strimmer. Learning large-scale graphical Gaussian models from genomic data. In J. F. F. Mendes, S. N. Dorogotsev, A. Povolostky, F. V. Abreu, and J. G. Oliveira, editors, *Science of Complex Networks: From Biology to the Internet and WWW*, volume 776, pages 263–276. American Institute of Physics, August 2004.

J. Schäfer and K. Strimmer. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, 4(1), 2005a. URL `http://www.bepress.com/sagmb/vol4/iss1/art32`. Article 32.

J. Schäfer and K. Strimmer. An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics*, 2005b.

R. J. Simes. An improved Bonferroni procedure for multipel tests of significance. *Biometrika*, 73:751–4, 1986.

H. Toh and K. Horimoto. Inference of a genetic network by a combined approach of cluster analysis and graphical Gaussian modeling. *Bioinformatics*, 18:287–97, 2002a.

H. Toh and K. Horimoto. System for automatically inferring a genetic network from expression profiles. *Journal of Biological Physics*, 28:449–64, 2002b.

O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, and B. Russ. Missing value estimation for DNA microarrays. *Bioinformatics*, 17(6):520–5, 2001.

M. J. van der Laan and A. E. Hubbard. Quantile-function based null distribution in resampling based multiple testing. *Statistical Applications in Genetics and Molecular Biology*, 5(1), 2006. URL `http://www.bepress.com/sagmb/vol5/iss1/art14`. Article 14.

M. J. van der Laan, S. Dudoit, and K. S. Pollard. Multiple testing. Part II. Step-down procedures for control of the family-wise error rate. *Statistical Applications in Genetics and Molecular Biology*, 3(1), 2004.

M. J. van der Laan, M. D. Birkner, and A. E. Hubbard. Empirical Bayes and resampling based multiple testing procedure controlling tail probability of the proportion of false positives. *Statistical Applications in Genetics and Molecular Biology*, 4(1), 2005. URL `http://www.bepress.com/sagmb/vol4/iss1/art29`. Article 29.

W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer, New York, fourth edition, 2002. URL `http://www.stats.ox.ac.uk/pub/MASS4`. ISBN 0-387-95457-0.

P. J. Waddell and H. Kishino. Correspondence analysis of genes and tissue types and finding genetic links from microarray data. *Genome Informatics*, 11:83–95, 2000a.

P. J. Waddell and H. Kishino. Cluster inference methods and graphical models evaluated on NC160 microarray gene expression data. *Genome Informatics*, 11:129–40, 2000b.

J. Wang, O. Myklebost, and E. Hovig. Mgraph: Graphical models for microarray data analysis. *Bioinformatics*, 19(17):2210–11, 2003.

P. H. Westfall and S. S. Young. *Resampling-based Multiple Testing: Examples and Methods for p-value Adjustment*. John Wiley and Sons, 1993.

J. Whittaker. *Graphical Models in Applied Multivariate Statistics*. Wiley, Chichester, 1990.

A. Wille and P. Bühlmann. Low-order conditional independence graphs for inferring genetic networks. *Statistical Applications in Genetics and Molecular Biology*, 5(1), 2006. URL `http://www.bepress.com/sagmb/vol5/iss1/art1`. Article 1.

A. Wille, P. Zimmerman, E. Vranová, A. Fürholz, O. Laule, S. Bleuler, L. Hennig, A. Prelić, P. von Ruhr, L. Thiele, E. Zitzler, W. Gruissem, and P. Bühlmann. Sparse graphical Gaussian modelling of the isoprenoid network in *arabidopsis thaliana*. *Genome Biology*, 5:R92, 2004.

F. Wong, C. K. Carter, and R. Kohn. Efficient estimation of covariance selection models. *Biometrika*, 90:809–30, 2003.

X. Wu, Y. Ye, and K. R. Subramanian. Interactive analysis of gene interactions using graphical Gaussian model. *ACM SIGKDD Workshop on Data Mining in Bioinformatics*, 3:63–9, 2003.

M. Yuan and Y. Lin. Model selection and estimation in the Gaussian graphical model. *Biometrika*, 94(1):19–35, 2007.

J. Zhu, P. Y. Lum, J. Lamb, D. GuhaThakurta, S. W. Edwards, R. Thieringer, J. P. Berge, M. S. Wu, J. Thompson, A. B. Sachs, and E. E. Schadt. An integrative genomics approach to the reconstruction of gene networks in segregating populations. *Cytogenetics Genome Research*, 105:363–374, 2004.