Tri-Graph: a Novel Graphical Model with Application to Genetic Regulatory Networks

Anja Wille and Peter Bühlmann
ETH Zürich, Switzerland

August 9, 2004

Abstract

Graphical models have become increasingly popular to analyze full conditional (in-)dependencies between random variables. However, full conditional relationships between variables can be only accurately estimated if the number of observations is relatively large in comparison to the number of variables. If this prerequisite is not fulfilled and the graphical model cannot be learned with adequate accuracy, simpler models have to be relied on to determine dependencies between random variables.

Here, we present such a simplified graphical modeling approach, called tri-graph, in which full conditional modeling is carried out in small subgraphs with three vertices only. These subgraphs are then combined into the full model. We analyze the probabilistic properties of the tri-graph and demonstrate in a simulation study that despite its simplicity, the tri-graph is a good estimator of the full conditional independence pattern between random variables.

Possible applications of our approach include the analysis of microarray gene expression profiles which usually comprise many variables (genes) but few observations. As an example, we use the tri-graph to estimate the conditional dependence structure between 40 genes to infer a genetic regulatory network for the isoprenoid biosynthesis in *Arabidopsis thaliana*.

Heading: Modified graphical modeling

1 Introduction

Graphical models (Edwards, 2000; Lauritzen, 1996) form a probabilistic tool to analyze and visualize conditional dependence between random variables. Random variables are represented by vertices of a graph and conditional relationships between them are encoded by edges. Conditional independence between two variables is always in relation to other existing and missing edges. The inclusion or removal of a single edge can change the independence pattern of the complete graph.

Graphical modeling of full conditional dependencies are powerful for a small number of random variables. Based on maximum likelihood methods with a model selection penalty, the independence structure of the graph can be learned. As the number of variables increases, however, only a small subset of the super-exponentially growing number of models can be tested (Wang *et al.*, 2003). More importantly, for many random variables, accurate estimation of conditional dependencies requires many more observations than are sometimes available (for example in gene expression profiling). The corresponding high rate of false positive and false negative edges then makes the interpretation of the graph within the Markov property framework (Edwards, 2000; Lauritzen, 1996) rather difficult (Husmeier, 2003; Waddell & Kishino, 2000).

These problems may be circumvented using a simpler approach with better estimation properties to characterize the dependence structure between random variables. The simplest method would be to model the marginal dependence structure in a so called covariance graph (Cox & Wermuth, 1993, 1996). The covariance structure of random variables can be accurately estimated and easily interpreted even with a large number of variables and a small sample size. However, the covariance graph contains only limited information since the effect of other variables on the relationship between two variables is ignored.

The full conditional independence graph and the covariance graph play dual (opposite) roles with respect to interpretation and estimation properties in graphical modeling. Our aim is to design a simple yet powerful approach in-between both forms of modeling. It should allow to study dependence patterns in a more complex and exhaustive way than with only pairwise correlation-based relationships while maintaining high accuracy even for few observations. For this purpose, we propose not to condition on all variables at a time. Instead, we apply graphical modeling only to small subgraphs with three vertices to explore the dependence between two of the variables conditional on the third one. These subgraphs are then combined for inference on the complete graph, which we call the "tri-graph".

A possible application of our approach would be in gene expression profiling where the activity of thousands of genes is monitored over relatively few samples. Despite the aforementioned problems, graphical models have become increasingly popular for inferring genetic regulatory networks based on the conditional dependence structure of gene expression levels (Wang *et al.*, 2003; Friedman *et al.*, 2000; Hartemink *et al.*, 2001; Toh & Horimoto, 2002). The conditional tri-graph can improve upon the accuracy of the estimated dependencies by focusing on simpler aspects of conditional dependence.

We analyze here probability and estimation properties of the tri-graph and demonstrate its usefulness to discover conditional dependence patterns. As our main interest is to apply our approach in gene expression profiling, we focused on simulated networks with genetic

and metabolic topologies, and we also present an example where we infer a gene network for isoprenoid biosynthesis in *Arabidopsis thaliana*.

2 Definition of the modified graphical model

Consider p random variables $X_1, \ldots, X_p$ which we sometimes denote by the random vector $\mathbf{X} = (X_1, \ldots, X_p)$. Full conditional dependence between two variables X_i and X_j refers to the conditional dependence between X_i and X_j given all other variables $X_k, k \in \{1, \ldots, p\} \setminus \{i, j\}$. Conditional independence between X_i and X_j denoted by $X_i \perp\!\!\!\perp X_j \mid \mathbf{X} \setminus \{X_i, X_j\}$ states that there is no direct relationship between X_i and X_j.

Describing all conditional (in-)dependencies is often a desirable goal. However, estimating and interpreting the conditional dependence pattern of p variables is often a very difficult task, due to the complexity of p-dimensional distributions. In particular, the latter applies when p is relatively large. Therefore, it may be desirable to rely on simpler models than the full conditional dependence structure of all p variables. The tri-graph in which the dependence pattern between two variables is explored by conditioning separately on one of the remaining variables X_k, for each $k \in \{1, \ldots, p\} \setminus \{i, j\}$, represents such a simplified model. Its precise definition is given in Section 2.3.

In graphical modeling, the dependence pattern between variables is associated with a graph in which vertices encode the random variables and edges encode conditional dependence between variables. In the full conditional independence graphs, two variables are connected if and only if these two variables are conditionally dependent given all remaining variables. Figure 2.1 shows an example of the dependence patterns between variables $X_1, \ldots, X_4$ and the corresponding full conditional independence graph. All edges in the graph are undirected.

$$X_1 \perp\!\!\!\perp X_4 \mid (X_2, X_3)$$
$$X_2 \perp\!\!\!\perp X_3 \mid (X_1, X_4)$$
$$X_2 \perp\!\!\!\perp X_4 \mid (X_1, X_3)$$

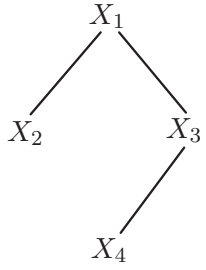

Figure 2.1: Conditional independence model and associated graph

For continuous random variables $\mathbf{X}$ that follow a multivariate normal distribution with mean $\mathbb{E}(\mathbf{X}) = \mu$ and covariance matrix $\mathrm{Cov}(\mathbf{X}) = \Sigma$,

$$\mathbf{X} \sim \mathcal{N}(\mu, \Sigma),$$

we now give the probabilistic definitions for graphical modeling based on the full conditional independence graph, the covariance graph, and our tri-graph.

2.1 The full conditional independence graph

In the full conditional independence graph, an edge between vertex i and j is drawn if and only if X_i and X_j are conditionally dependent given all other variables $\{X_k; k \in \{1, \ldots, p\} \setminus \{i, j\}\}$. Due to the Gaussian assumption, this means that the vertices X_i and X_j $(i \neq j)$ are connected if and only if the partial correlation coefficients

$$\omega_{ij} \neq 0, \quad \omega_{ij} = \frac{\Sigma^{-1}_{ij}}{\sqrt{\Sigma^{-1}_{ii}\Sigma^{-1}_{jj}}} \tag{2.1}$$

where Σ^{-1}_{ij} are the elements of the inverse covariance matrix (precision matrix).

To learn the conditional independence structure of the graph, it is therefore necessary to determine which elements of the precision matrix Σ^{-1} are estimated to be 0. Since this is commonly carried out jointly for all edges in a likelihood approach, super-exponentially $(2^{p(p-1)/2})$ many tests have to be conducted to find the best model for the data. For a large number of variables, this is hardly feasible. Instead, non-exhaustive search algorithms such as backward and forward selection procedures are used to learn the model.

Furthermore, since the computation of the partial correlation coefficients includes matrix inversion of the covariance matrix, a relatively large sample size n is necessary for their accurate estimation (Lauritzen, 1996). For certain applications like genomics, such a sample size is typically not available. Conditional independence graphs learned from such data will then be rather unreliable with a high false positive and high false negative rate. We will show that the much simpler concepts such as the covariance graph and the tri-graph can be estimated with higher accuracy. However, among the latter two, only the tri-graph can capture the conditional independence structure well.

2.2 The covariance graph

In the covariance graph, an edge between vertex i and j $(i \neq j)$ is drawn if and only if the correlation coefficient

$$\rho_{ij} \neq 0, \quad \rho_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}. \tag{2.2}$$

The covariance graph as a representation of the marginal dependence structure between variables is simple to interpret and has the advantage that it can be accurately estimated from finite-sample data even if p is very large in comparison to sample size n, see Section 3.1. However, this graph is often not sufficient to capture more complex conditional dependence patterns.

2.3 Tri-graph: a modified graphical model

Our modified graph construction combines statistical features from the covariance and the conditional independence graph. In this respect, it can be viewed as being between the covariance and the full conditional dependence graph.

To explore some dependence structure between two variables X_i and X_j, we propose not to jointly condition on all remaining variables at a time. Instead, we consider sepa-

rately all pairwise partial correlations

$$\omega_{ij|k} = \frac{\rho_{ij} - \rho_{ik}\rho_{jk}}{\sqrt{(1-\rho_{jk}^2)(1-\rho_{ij}^2)}}$$

of X_i and X_j given one of the remaining variable X_k. These partial correlation coefficients are then combined to draw conclusions on some aspect of the dependence between X_i and X_j.

Definition 1 ***(Tri-graph)***
Draw an edge between vertex i and j $(i \neq j)$ if and only if

$$\rho_{ij} \neq 0 \quad \textit{and} \quad \omega_{ij|k} \neq 0 \textit{ for all } k \in \{1,\ldots,p\} \setminus \{i,j\}.$$

The name "tri-graph" refers to the fact that all partial correlations among **tri**ples of random variables are involved. Let $T_{ij} = \{\rho_{ij}, \omega_{ij|k}; k \in \{1,\ldots,p\} \setminus \{i,j\}\}$ be the set of the correlation and partial correlation coefficients for X_i and X_j. As parameter τ_{ij} for an edge between X_i and X_j, we can use the element of T_{ij} with minimum absolute value. We assign an edge if and only if

$$\tau_{ij} = \arg\min_{\tau \in T_{ij}} (|\tau|) > 0$$

A tri-graph still reflects some measure of conditional dependence given each other variable separately, and thus describes a more complex dependence pattern than a covariance graph. In fact, we can show that it can capture the full conditional independence structure well and sometimes even exactly, see Theorem 1. On the other hand, it is still reasonably simple to interpret. An edges between two variables X_i and X_j represents a dependence that cannot be explained by any of the other variables X_k. From a statistical perspective, a tri-graph can be accurately estimated from data even if p is large relative to sample size n, see Section 3.1. Furthermore, estimation of tri-graphs is based on an exhaustive computation, see Section 3, and one does not have to rely on approximate search algorithms as for full conditional dependence graphs.

2.4 Some examples and a rigorous property

We are describing here with some simple examples and a theorem in how far the full independence graph, the tri-graph and the correlation graph relate to each other.
Example 1: Consider 4 random variables $\mathbf{X} = (X_1, X_2, X_3, X_4) \sim N(0, \Sigma)$ with

$$\Sigma = \begin{pmatrix} 1 & -1 & -1 & -1 \\ -1 & 2 & 1 & 1 \\ -1 & 1 & 2 & 1 \\ -1 & 1 & 1 & 2 \end{pmatrix} \quad \text{and} \quad \Sigma^{-1} = \begin{pmatrix} 4 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Based on the inverted covariance matrix Σ^{-1}, we obtain a conditional independence model as shown in Figure 2.2. In such a setting, tri-graph and independence graph are exactly the same whereas the covariance graph is the full graph. A similar situation can be found for the example given in Figure 2.1. Tri-graph and independence graph show the same structural dependence whereas the covariance graph is again full. In general, we have the following theorem.

$$X_2 \perp\!\!\!\perp X_3 \mid (X_1, X_4)$$
$$X_2 \perp\!\!\!\perp X_4 \mid (X_1, X_3)$$
$$X_3 \perp\!\!\!\perp X_4 \mid (X_1, X_2)$$

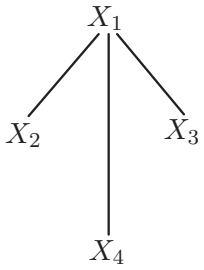

Figure 2.2: A conditional independence model for which tri-graph and full conditional graph are the same.

Theorem 1 *If the full conditional independence graph does not contain any cycles then the tri-graph coincides with the full conditional independence graph.*

A proof is given in the Appendix.

Example 2: Consider 4 random variables $\mathbf{X} = (X_1, X_2, X_3, X_4) \sim N(0, \Sigma)$ with

$$\Sigma = \begin{pmatrix} 4 & -7 & -5 & 6 \\ -7 & 13 & 9 & -11 \\ -5 & 9 & 7 & -8 \\ 6 & -11 & -8 & 10 \end{pmatrix} \quad \text{and} \quad \Sigma^{-1} = \begin{pmatrix} 5 & 2 & 1 & 0 \\ 2 & 2 & 0 & 1 \\ 1 & 0 & 2 & 1 \\ 0 & 1 & 1 & 2 \end{pmatrix}.$$

The full conditional independence graph includes all edges except those between the pairs (X_1, X_4) and (X_2, X_3) as shown in Figure 2.3. From Σ we see that the covariance graph

$$X_2 \perp\!\!\!\perp X_3 \mid (X_1, X_4)$$
$$X_1 \perp\!\!\!\perp X_4 \mid (X_2, X_3)$$

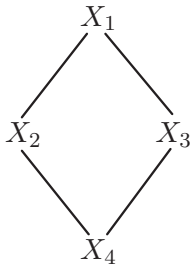

Figure 2.3: A conditional independence model for which the cyclic full conditional independence graph is contained in the tri-graph

includes all edges. The tri-graph also includes all edges since for example, X_2 and X_3 are not conditionally independent on either X_1 or X_4 alone.

2.5 More complex situations

For modeling Gaussian graphs with a large number of nodes, we have to rely on simulations to explore the relationships between the full conditional independence graph, the tri-graph and the covariance graph. Our focus for application is on sparse graphs with a modular topology such as metabolic, genetic regulatory or protein interaction networks. For these

$X_i \perp\!\!\!\perp X_j \mid T$

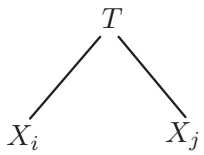

Figure 2.4: Conditional independence model and associated graph for X_i, X_j and the latent variable T.

networks, it has been repeatedly suggested that the connectivity of the vertices follows a power law with exponents γ between 2 and 3 (Jeong *et al.*, 2000; Maslov & Sneppen, 2002). In our simulations, we adopt this modular network structure by sampling the number of edges for each node independently from a power-law distribution $p(k) = \frac{k^{-\gamma}}{\zeta(\gamma)}$ with exponent $\gamma = 2.5$. The normalization constant $\zeta(\gamma)$ is the Riemann zeta function. The graphs that we obtain by this method are very sparse and usually contain fewer edges than the number of nodes (see Table 2.1). In order to simulate graphs with more edges, we also generate graphs with exponent $\gamma = 1.5$ and 0.5.

Edges are then randomly assigned to other nodes (with equal probabilities). This random graph structure is used to define the zeros in the precision matrix: $\Sigma^{-1}_{ij} = 0$ if there is no edge between i and j. In order to model the non-zero elements of Σ^{-1} (and the partial correlation coefficients), we first look at two nodes only. We assume that the conditional dependence between two random variables X_i and X_j is introduced by an underlying latent random variable T. If we only consider the three variables X_i, X_j and T, we could model the effect of T on X_i and X_j in a conditional independence graph (Figure 2.4) with precision matrix

$$\Sigma^{-1}_{X_i,X_j,T} = \begin{pmatrix} 1 & 0 & \frac{-\beta_{ij}}{1+\beta_{ij}^2+\beta_{ji}^2} \\ 0 & 1 & \frac{-\beta_{ji}}{1+\beta_{ij}^2+\beta_{ji}^2} \\ \frac{-\beta_{ij}}{1+\beta_{ij}^2+\beta_{ji}^2} & \frac{-\beta_{ji}}{1+\beta_{ij}^2+\beta_{ji}^2} & 1 \end{pmatrix}.$$

Magnitude and sign of the coefficients β_{ij} and β_{ji} determine how strong the effect of T is on X_i and X_j respectively. After T is integrated out, the precision matrix for the variables X_i and X_j is

$$\Sigma^{-1}_{X_i,X_j} = \begin{pmatrix} 1 & \frac{-\beta_{ij}\beta_{ji}}{\sqrt{1+\beta_{ij}^2}\sqrt{1+\beta_{ji}^2}} \\ \frac{-\beta_{ij}\beta_{ji}}{\sqrt{1+\beta_{ij}^2}\sqrt{1+\beta_{ji}^2}} & 1 \end{pmatrix}.$$

We can therefore write

$$\Sigma^{-1}_{X_i,X_j} = \sqrt{D}(Id + BB^t)\sqrt{D} \tag{2.3}$$

with

$$B = \begin{pmatrix} \beta_{ij} \\ -\beta_{ji} \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} \frac{1}{\sqrt{1+\beta_{ij}^2}} & 0 \\ 0 & \frac{1}{\sqrt{1+\beta_{ji}^2}} \end{pmatrix}.$$

If we model partial correlation coefficients for all variables $X_1, \ldots, X_p$, we also use the scheme as described in (2.3). Let $\{e_{kl}\}$ be the edges in the graph where the indices $k < l$ refer to the indices of the variables X_k and X_l that are connected by e_{kl}. Let further e be the total number of edges and B a $p \times e$ matrix with elements

$$b_{ie_{kl}} = \begin{cases} \beta_{il} & \text{if } i = k \\ -\beta_{ki} & \text{if } i = l \\ 0 & \text{otherwise.} \end{cases}$$

Then we find

$$\begin{aligned}(BB^t)_{ij} &= \sum_{e_{kl}} b_{ie_{kl}} b_{je_{kl}} \\ &= \begin{cases} \sum_{e_{ik}} \beta_{ik}^2 + \sum_{e_{ki}} \beta_{ki}^2 & \text{if } i = j \\ -\beta_{ij}\beta_{ji} & \text{if } i \neq j \text{ and there is an edge between } i \text{ and } j \\ 0 & \text{if } i \neq j \text{ and there is no edge between } i \text{ and } j \end{cases}\end{aligned}$$

and the partial correlation coefficient for two conditionally dependent variables X_i and X_j can be modeled as (Equations (2.1) and (2.3))

$$\omega_{ij} = \frac{\beta_{ij}\beta_{ji}}{\sqrt{1 + \sum_{e_{ik}} \beta_{ik}^2 + \sum_{e_{ki}} \beta_{ki}^2}\sqrt{1 + \sum_{e_{jk}} \beta_{jk}^2 + \sum_{e_{kj}} \beta_{kj}^2}}.$$

Our scheme to generate partial correlation coefficient for a pre-specified independence graph has the advantage that the sampled precision matrix is always positive definite. Therefore, the random graph structure and B define a normal distribution $N(0, \Sigma)$. The magnitude and sign of the coefficients β_{ij} determine the magnitude and sign of the partial correlation coefficients. In our simulations, we sampled the coefficients β_{ij} from three different uniform distributions $U(-\beta_{\max}, \beta_{\max})$ with $\beta_{\max} = 1, 5, 100$.

With this model, we generated 100 graphs and covariance matrices each for graphs with p =5, 10, 20, and 40 vertices and connectivity parameter γ =2.5, 1.5 and 0.5. For each p and each γ, we compared the structure of the independence graph, the covariance graph and the tri-graph. In Table 2.1, the mean and standard deviations for the number of edges per graph is shown. For decreasing γ, the number of edges increases in the full conditional independence graphs. The edges of the conditional independence graph almost always formed a subset of the tri-graph. For graphs with low connectivity ($\gamma = 2.5$), the tri-graph contained only few additional edges. However, for $\gamma = 0.5$, the tri-graphs were considerably larger than the corresponding independence graphs. The covariance graphs always contained many more edges than both the other graphical models.

We also monitored the difference between the correlation and partial correlation coefficients ($\rho_{ij} - \omega_{ij}$) and the difference between tri-graph and partial correlation coefficients ($\tau_{ij} - \omega_{ij}$) for unconnected ($\omega_{ij} = 0$) and connected ($\omega_{ij} \neq 0$) vertices i and j (see Table 2.2 for the root mean squared errors (RMSE) averaged over all $i < j$). Most edges in the tri-graph that are not part of the conditional independence graph have coefficients in the vicinity of 0. In fact, for $\omega_{ij} = 0$ the 5%- and 95%-quantile of the distribution of tri-graph coefficients were located within the interval [-0.05, 0.05] for all simulation settings. For $\omega_{ij} \neq 0$, the 5%-95%-quantile ranges were always larger. This indicates that the tri-graph can capture the conditional independence structure quite well.

number of variables	γ	number of edges in the independence graph	tri-graph	covariance graph
p=5	2.5	3.53(0.70)	3.56(0.81)	6.52(2.98)
	1.5	4.14(1.04)	4.38(1.57)	7.84(2.89)
	0.5	5.59(1.19)	6.47(2.07)	9.82(1.03)
p=10	2.5	7.46(1.51)	7.76(2.31)	20.68(14.38)
	1.5	10.87(2.67)	15.02(7.70)	38.46(11.42)
	0.5	18.86(3.82)	33.02(7.45)	45.00(0.00)
p=20	2.5	15.48(2.91)	16.87(8.08)	56.68(46.09)
	1.5	24.42(4.32)	51.27(23.96)	166.45(43.82)
	0.5	44.97(6.70)	130.00(26.17)	190.00(0.00)
p=40	2.5	30.45(3.80)	31.08(7.40)	115.66(91.62)
	1.5	49.70(6.79)	173.03(85.08)	680.35(162.73)
	0.5	88.34(8.74)	498.33(82.76)	780.00(0.00)

Table 2.1: Mean number of edges (and standard deviation) for the three different graphical models in Section 2.5 as a function of γ and p.

3 Estimation from data

For a pair of edges i, j, we can test all null-hypotheses

$$H_0(i,j|k): \ \omega_{ij|k} = 0 \ \text{ for } k \notin \{i,j\}$$

versus the alternatives $H_1(i,j|k): \ \omega_{ij|k} \neq 0$.

Such hypotheses can be tested with the likelihood ratio test under the Gaussian assumption

$$X_i, X_j, X_k \sim \mathcal{N}_3(\mu, \Sigma).$$

The null hypotheses are $(\Sigma^{-1})_{12} = 0$ (which is equivalent to $\omega_{ij|k} = 0$) and the alternatives are Σ unconstrained. Under the null-hypotheses and the assumption that the data are i.i.d. realizations from a p-dimensional normal distribution, the log-likelihood ratios are asymptotically χ^2-distributed (Lauritzen, 1996) and every likelihood ratio test of $H_0(i,j|k)$ versus $H_1(i,j|k)$ yields a P-value

$$P(i,j|k).$$

Furthermore, the likelihood ratio test of the null hypothesis for the marginal correlation

$$H_0(i,j|\,\emptyset): \ \rho_{ij} = 0 \text{ versus } H_1(i,j|\,\emptyset): \ \rho_{ij} \neq 0$$

yields a P-value $P(i,j|\,\emptyset)$.

number of variables	γ	RMSE covariance graph $\omega_{ij}=0$	covariance graph $\omega_{ij}\neq 0$	tri-graph $\omega_{ij}=0$	tri-graph $\omega_{ij}\neq 0$
p=5	2.5	0.221	0.144	0.002	0.029
	1.5	0.268	0.189	0.009	0.04
	0.5	0.251	0.178	0.02	0.056
p=10	2.5	0.161	0.16	0.004	0.046
	1.5	0.168	0.151	0.01	0.046
	0.5	0.118	0.1	0.018	0.042
p=20	2.5	0.105	0.155	0.001	0.044
	1.5	0.111	0.136	0.007	0.05
	0.5	0.065	0.066	0.011	0.031
p=40	2.5	0.075	0.162	0.001	0.045
	1.5	0.076	0.127	0.005	0.051
	0.5	0.046	0.058	0.006	0.028

Table 2.2: RMSE averaged over all $i<j$ with $\omega_{ij}=0$ and averaged over all $i<j$ with $\omega_{ij}\neq 0$ between correlation coefficients ρ_{ij} and partial correlation coefficient ω_{ij} (right columns) and RMSE between tri-graph coefficients τ_{ij} and partial correlation coefficients ω_{ij} (left columns). $\beta_{\max}=5$.

Recall that an edge in a tri-graph between vertex i and j exists if $H_0(i,j|\,\emptyset)$ is rejected and $H_0(i,j|\,k)$ is rejected for all vertices $k\notin\{i,j\}$. Thus, there is evidence for an edge between vertex i and j if

$$\max_{k\in\{\emptyset,1,2,\ldots,p\}\setminus\{i,j\}} P(i,j|\,k)<\alpha,$$

where α is the significance level. For deciding about a single edge between vertices i,j, it is not necessary to correct for the $p-1$ multiple testing over all conditioning vertices k.

Proposition 1 *Consider the single hypothesis (for some fixed pair (i,j)),*

$$H_0(i,j)\text{: at least one } H_0(i,j|\,k^*) \text{ is true for some } k^*\in\{\emptyset,1,2,\ldots,p\}\setminus\{i,j\}.$$

Assume that for all $k\in\{\emptyset,1,2,\ldots,p\}\setminus\{i,j\}$ the individual test satisfies

$$\mathbb{P}_{\tilde{H}_0(i,j|k)}\left[H_0(i,j|\,k) \text{ rejected }\right]\leq\alpha,$$

where $\tilde{H}_0(i,j|k)=\{H_0(i,j|k) \text{ true }\}\cap\{H_0(i,j|k') \text{ true or false (and compatible with } H_0(i,j|k) \text{ for all } k'\neq k\}$. Then, the type I error

$$\mathbb{P}_{H_0(i,j)}\left[H_0(i,j|\,k) \text{ are rejected for all } k\in\{\emptyset,1,2,\ldots,p\}\setminus\{i,j\}\right]\leq\alpha.$$

A proof is given in the Appendix. Note that the log-likelihood ratio test described above satisfies asymptotically the assumption of Proposition 1. It will be necessary though to correct over the $p(p-1)/2$ multiple tests over all pairs of vertices (i,j). The estimation algorithm is as follows.

Estimation algorithm

1. For all $i, j \in \{1, \ldots, p\},\ i \neq j$ and $k \in \{1, 2, \ldots, p\} \backslash \{i, j\}$, compute P-values $P(i, j|k)$ from the log-likelihood ratio test with respect to the model $X_i, X_j, X_k \sim \mathcal{N}(0, \Sigma)$ with null hypothesis $H_0(i, j|k)$: $\Sigma^{-1}_{ij} = 0$ and alternative $H_1(i, j|k)$: $\Sigma^{-1}_{ij} \neq 0$. Also, compute $P(i, j|\emptyset)$ from the log-likelihood ratio test with null hypothesis $H_0(i, j|\emptyset)$: $\Sigma_{ij} = 0$ and alternative $H_1(i, j|\emptyset)$: $\Sigma_{ij} \neq 0$. Note the symmetry $P(i, j|k) = P(j, i|k)$.

2. For all pairs $(i, j) = (j, i)$ compute the maximum P-values (note the correspondence to Proposition 1)

$$P_{max}(i, j) = \max_{k \in \{\emptyset, 1, 2, \ldots, p\} \backslash \{i, j\}} P(i, j|k).$$

3. Correct the maximum P-values $P_{max}(i, j)$ over the $p(p-1)/2$ multiple tests for all pairs of vertices. For example, use the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995) for controlling the false discovery rate. Denote the corrected maximal P-values by

$$P_{max,corr}(i, j).$$

4. Draw an edge between vertex i and j if and only if

$$P_{max,corr}(i, j) < \alpha,$$

for some pre-specified significance level such as $\alpha = 0.05$.

The corrected maximum P-values $P_{max,corr}(i, j)$ can be used as a measure of significance for an edge between nodes i and j. It is worth pointing out that our estimation for a tri-graph is done in an exhaustive manner. This is in sharp contrast to full conditional dependence graphs where it is often necessary to use non-exhaustive computations in huge graph spaces, e.g. random search methods, greedy stepwise algorithms, or stochastic simulation in the Bayesian framework (Madigan & Raftery, 1994; Giudici & Green, 1999; Dobra *et al.*, 2004).

3.1 Asymptotic consistency for large number of variables

We present here some theory which reflects at least from an asymptotic point of view that tri-graphs can be accurately estimated even if the number p of vertices is large relative to sample size.

Denote the data by $\mathbf{X}_1, \ldots, \mathbf{X}_n$ ($\mathbf{X}_i \in \mathrm{R}^p$) which are assumed to be i.i.d. random vectors. The estimators for the mean $\mu = \mathbb{E}[\mathbf{X}]$, the covariance matrix $\Sigma = \mathrm{Cov}(\mathbf{X})$, the

correlation coefficients ρ_{ij} and the partial correlation coefficients $\omega_{ij|k}$ are as follows:

$$\begin{aligned}
\hat{\mu}(n) &= n^{-1}\sum_{i=1}^{n}\mathbf{X}_i, \\
\hat{\Sigma}(n) &= n^{-1}\sum_{i=1}^{n}(\mathbf{X}_i-\hat{\mu})(\mathbf{X}_i-\hat{\mu})^T \\
\hat{\rho}_{ij} &= \frac{\hat{\Sigma}(n)_{ij}}{\sqrt{\hat{\Sigma}(n)_{ii}\hat{\Sigma}(n)_{jj}}} \\
\hat{\omega}(n)_{ij|k} &= \frac{\hat{\rho}_{ij}-\hat{\rho}_{ik}\hat{\rho}_{jk}}{\sqrt{(1-\hat{\rho}_{ik}^2)(1-\hat{\rho}_{jk}^2)}},\ 1\le i<j\le p,\ k\ne i,j. \qquad (3.4)
\end{aligned}$$

We are giving below some uniform consistency results for these estimators when the dimensionality p is large relative to sample size. The set-up is as follows. We assume that the data are realizations from a triangular array of random vectors of dimension $p=p_n$ where p_n is allowed to grow as sample size $n\to\infty$:

$$\mathbf{X}_{(n),1},\ldots,\mathbf{X}_{(n),n}\ \text{i.i.d.}\ \sim P_{(n)}, \qquad (3.5)$$

where $P_{(n)}$ denotes some probability distribution in R^{p_n}. We denote by $\mu(n)=\mathbb{E}[\mathbf{X}_{(n)}]$ and by $\Sigma(n)=\mathrm{Cov}(\mathbf{X}_{(n)})$; these moments exist by the following assumption.

(A1) $\sup_{n\in\mathrm{N},1\le j\le p_n}\mathbb{E}|(\mathbf{X}_{(n)})_j|^{4s}<\infty$ for some $s\ge 1/2$.

Proposition 2 *The data are as in (3.5), satisfying assumption (A1) for some $s\ge 1/2$. Assume that $p_n=o(n^{s/2})$ $(n\to\infty)$. Then,*

$$\begin{aligned}
\max_{1\le j\le p_n}|\hat{\mu}(n)_j-\mu(n)_j| &= o_P(n^{-3s/2})\ (n\to\infty), \\
\max_{1\le i<j\le p_n}|\hat{\Sigma}(n)_{ij}-\Sigma(n)_{ij}| &= o_P(1)\ (n\to\infty).
\end{aligned}$$

A proof is given in the Appendix. In case where $\mathbf{X}\sim\mathcal{N}_{p_n}(\mu(n),\Sigma(n))$, we could allow of a faster growth rate p_n satisfying $\log(p_n)/n\to 0$.

For uniform consistency of partial correlations, we make an additional assumption:

(A2) $\inf_{n\in\mathrm{N},1\le j\le p_n}\Sigma(n)_{jj}>0$, and
$\sup_{n\in\mathrm{N},1\le i<j\le p_n}|\rho(n)_{ij}|<1$, where $\rho(n)_{ij}=\Sigma(n)_{ij}/\sqrt{\Sigma(n)_{ii}\Sigma(n)_{jj}}$.

The first assumption in (A2) means that none of the variables becomes degenerate as $n\to\infty$, i.e. having a variance tending to zero. The second assumption says that all the variables are linearly identifiable, i.e. none of the variables is an exact linear function of another one.

Theorem 2 *The data are as in (3.5), satisfying assumption (A1) for some $s\ge 1/2$ and (A2). Assume that $p_n=o(n^{s/2})$ $(n\to\infty)$. Then,*

$$\max_{1\le i<j\le p_n;1\le k\le p_n,k\ne i,j}|\hat{\omega}(n)_{ij|k}-\omega(n)_{ij|k}|=o_P(1)\ (n\to\infty).$$

A proof of Theorem 2 is given in the Appendix. Also here, in case where $\mathbf{X} \sim \mathcal{N}_{p_n}(\mu(n), \Sigma(n))$, we could allow of a p_n satisfying $\log(p_n)/n \to 0$. Theorem 2 describes a *uniform* convergence result with respect to all partial correlations: for a small number $\delta > 0$ and with high probability, all estimated partial correlations are within δ-distance from the true partial correlations if the sample size is sufficiently large. This is much stronger than a pointwise result. Since a tri-graph involves all partial correlations, see Definition 1, our uniform consistency result, saying that we can *simultaneously* estimate *all* partial correlations reasonably well, implies that we can estimate a tri-graph reasonably well even if the number of vertices p is much larger than sample size n. In fact, consistent estimation of high-dimensional tri-graphs is possible if true non-zero partial and marginal correlations are bounded away from zero.

It should be stated clearly that the bound in Theorem 2 is generally worse, although still $o_P(1)$, than in Proposition 2 for the covariances. Clearly, the result from Theorem 2 could be generalized to partial correlations $\omega(X_i, X_j|\{X_{k_1}, \ldots, X_{k_m}\})$ $(k_1, \ldots k_m \neq i, j)$ for a *fixed* m with respect to sample size n (although a uniform bound for such partial correlations is expected to become worse as the the value of m increases). If $m = m_n$ would grow with sample size, we would have to further restrict the growth of the dimensionality p_n.

The extreme case is the estimate of $\Sigma(n)^{-1}$ when inverting the estimate $\hat{\Sigma}(n)$ from (3.4). This can only be done if $p_n < n$ and pointwise consistency $|(\hat{\Sigma}(n))^{-1}_{ij} - \Sigma(n)^{-1}_{ij}| = o_P(1)$ $(1 \leq i < j \leq p_n)$ only holds if $p_n = o(n)$ $(n \to \infty)$ (Lauritzen, 1996). Thus, the unconstrained graphical Gaussian model can only be estimated if the dimensionality is "small" relative to the sample size. This is in sharp contrast to tri-graphs, where p_n is allowed to grow much faster than n, as described in Theorem 2. For example, by neglecting the constants in Theorem 2, the following dimensionalities are allowed for $n = 100$ and $4s$ existing moments for the components of $\mathbf{X}$:

$n = 100$	$4s = 8$	$4s = 12$	$4s = 16$	$4s = 20$
$p = o(n^{1/2})$	$o(100)$	$o(1'000)$	$o(10'000)$	$o(100'000)$

When assuming sparseness of the true conditional indendence graph, regularization methods could be used to cope with large p (Dobra *et al.*, 2004; Meinshausen & Bühlmann, 2004). In comparison, consistent tri-graph estimation is not subject to a sparsity assumption.

4 Numerical results for simulated data

4.1 Results with sampled data

From each of the simulated models, we sampled i.i.d. data from $\mathcal{N}(0, \Sigma)$ where Σ is the covariance matrix of the corresponding model parameters as described in Section 2.5. Depending on the size of the graph, we sampled data with few and many observations (see Table 4.3). The effect of the sample sizes on the estimates of the partial correlation coefficients $\hat{\omega}_{ij}$, tri-graph coefficients $\hat{\tau}_{ij}$ and correlation coefficients $\hat{\rho}_{ij}$ can be seen in Figures 4.5-4.8.

Figure 4.5 shows the root mean squared error (RMSE) between true coefficients and the corresponding estimates of the different graphical modeling approaches. Results are

number of variables p	number of observations n
5	10,20,30,50,100,500,1000,5000
10	20,30,50,100,500,1000,5000
20	30,50,100,500,1000,5000
40	50,100,500,1000,5000

Table 4.3: Number of observations n used to sample data from the original graphs with p vertices

shown for $\gamma = 1.5$ and $\beta_{\max} = 5$. It can be seen that for small n, the RMSE of the coefficients $\omega_{ij|k}$ does not differ much from the RMSE of the correlation coefficients $\rho_{ij}$and that both coefficients can be more accurately estimated than the full partial correlation coefficients ω_{ij}. As the number of observations n increases, the RMSEs for all coefficients decrease to 0. In all simulation settings, we found the same underlying pattern as in Figure 4.5. For $\beta_{\max} = 1$, however, the RMSE of the coefficients differed only slightly, even when n was small. Interestingly, estimates of the tri-graph coefficients τ_{ij} are even better than the estimates of the coefficients ρ_{ij} and $\omega_{ij|k}$. This indicates that the minimum of ρ_{ij} and $\omega_{ij|k}$ for $k \in \{1, 2, \ldots, p\} \setminus \{i, j\}$ can be much more reliably estimated than each of the coefficients ρ_{ij} and $\omega_{ij|k}$ for $k \in \{1, 2, \ldots, p\} \setminus \{i, j\}\}$ separately. Theorem 2 can therefore be viewed as providing a conservative upper bound for the estimation accuracy of the tri-graph coefficients.

We also monitored how well the estimates of the full partial correlation coefficients $\hat{\omega}_{ij}$, the tri-graph coefficients $\hat{\tau}_{ij}$ and the correlation coefficients $\hat{\rho}_{ij}$ represent the true full partial correlation coefficients $\omega_{ij}$of the original conditional independence graph. In Figure 4.6, the RSME between the sampled partial correlation coefficients, the sampled tri-graph coefficients, the sampled correlation coefficients and the true partial correlation coefficients are shown. For small to moderate n, the full conditional independence graph is better represented by the estimated tri-graph coefficients than the estimated partial correlation coefficients. Therefore, although being a rather simple estimator of complex dependence patterns, tri-graph coefficients can outperform partial correlation coefficients in detecting conditional dependence/independence.

Figure 4.7 shows the cumulative distribution functions (CDF) of the different coefficients for pairs of vertices with and without edges. Again, one can clearly see that a small to moderate sample size ($n = 50$) leads to rather unreliable estimates $\hat{\omega}_{ij}$ for the conditional independence graph (reflected by a gradual slope of the CDF of $\hat{\omega}_{ij} - \omega_{ij}$ at 0) whereas estimates of the tri-graph coefficients $\hat{\tau}_{ij}$ are much more stable (steeper slope of the CDF of $\hat{\tau}_{ij} - \omega_{ij}$).

In graphs with many nodes, the main purpose of a study may not be to find all connections between nodes but to find some true connections, hopefully the most important ones. In such a procedure, one would only consider gene pairs whose absolute partial correlation coefficient or tri-graph coefficient would be above a certain threshold t. By counting the number of true and false positives, true and false negatives for all values $t \in [0, 1]$, one

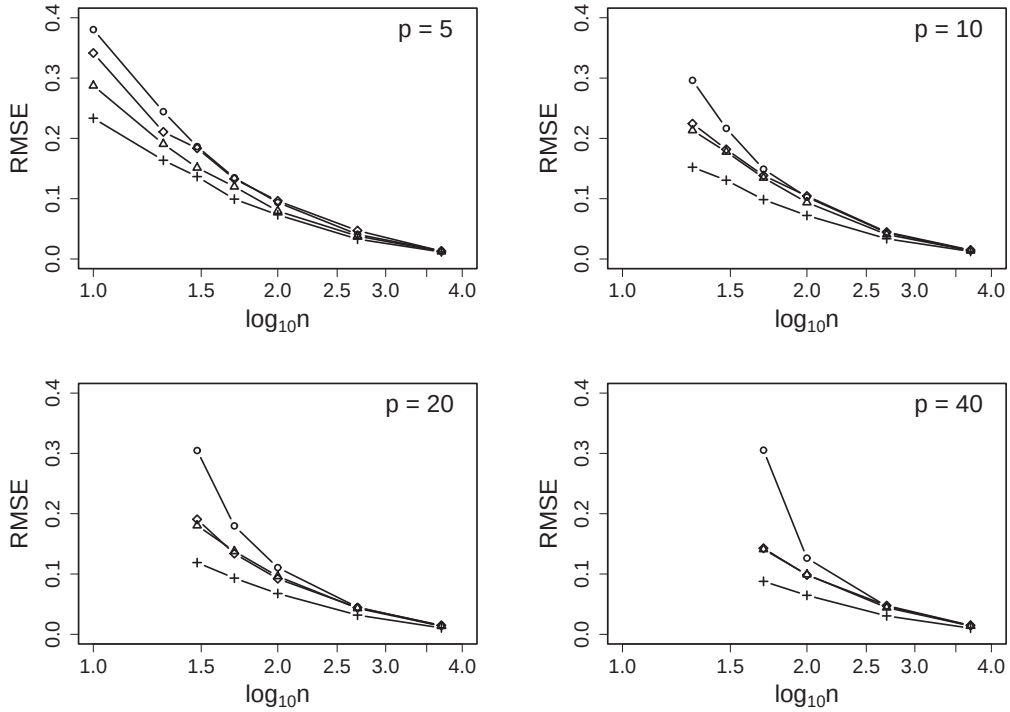

Figure 4.5: Root mean squared error (RMSE) averaged over all $i < j$ between the sampled and true partial correlation coefficients $\hat{\omega}_{ij}$ and ω_{ij} ($\circ$), sampled and true correlation coefficients $\hat{\rho}_{ij}$ and ρ_{ij} ($\triangle$), $\hat{\omega}_{ij|k}$ and $\omega_{ij|k}$ ($\diamond$) and sampled and true tri-graph coefficients $\hat{\tau}_{ij}$ and τ_{ij} ($+$) for different network sizes p and different number of observations n.

obtains the so called ROC curves by plotting the sensitivity (true positive rate) against the complementary specificity (false positive rate) for each t. The upper panel of Figure 4.8 displays the average ROC curves for the conditional independence graph, the covariance graph and the tri-graph for p=40 and $\beta_{\max} = 100$. We also included the ROC curves for learning the full conditional independence graph based on backward selection within the maximum likelihood framework, as implemented in the MIM package (2003). For small complementary specificities, the ROC curve of the tri-graph has a steeper slope than the other ROC curves suggesting the best performance in detecting true positive edges of the full conditional independence graph.

The tri-graph outperforms all the other methods (including the backward selection approach) for a small (n=100) and a large (n=1000) number of observation. For n=1000 observations, however, the ROC curves of the tri-graph, the full conditional independence graph and the backward selection approach differ only marginally. Our findings are further substantiated when we look at the false discovery rate (FDR) as a function of the selected edges. Again, the FDR of the tri-graph is smaller than the ones of the other methods.

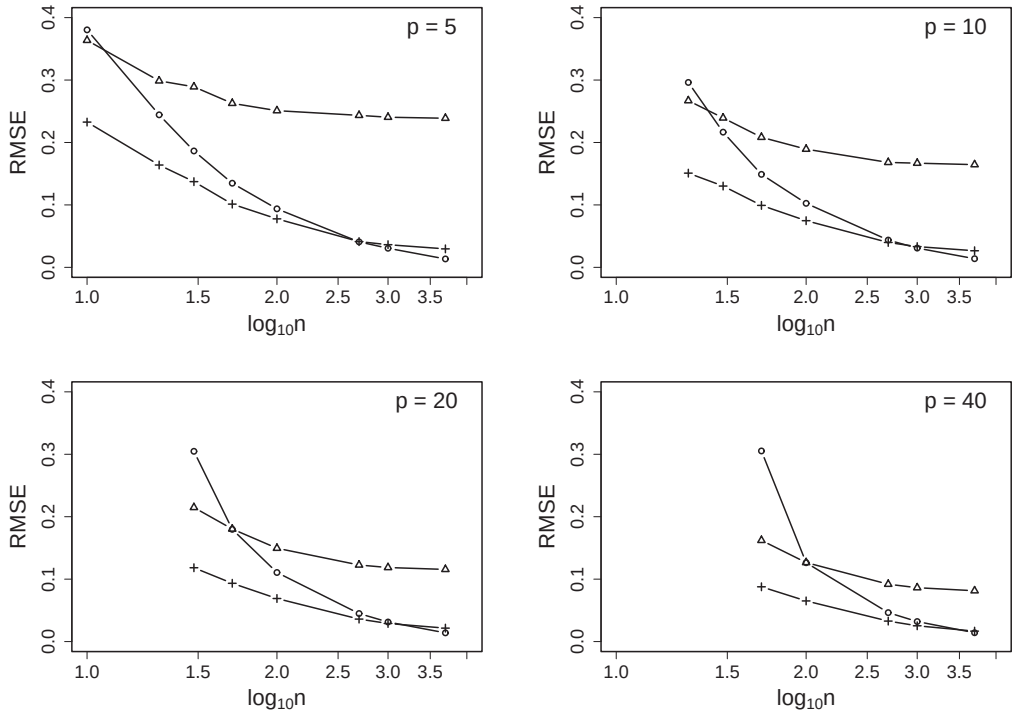

Figure 4.6: Root mean squared error (RMSE) averaged over all $i < j$ between sampled partial correlation coefficients $\hat{\omega}_{ij}$ and true partial correlation coefficients ω_{ij} ($\circ$), sampled correlation coefficients $\hat{\rho}_{ij}$ and ω_{ij} ($\triangle$) and tri-graph coefficients $\hat{\tau}_{ij}$ and ω_{ij} ($+$) for different network sizes p and different number of observations n.

5 An example with gene expression microarray data

In this section, we will demonstrate and further motivate the usefulness of the tri-graph in an application to gene expression profiling. Our aim is to infer aspects of a genetic regulatory network for isoprenoid biosynthesis (left panel of Figure 5.9) in *Arabidopsis thaliana*.

Isoprenoids serve numerous biochemical functions in plants, e.g. as components of membranes (sterols), as photosynthetic pigments (carotenoids and chlorophylls) or as hormones (gibberellins). They are synthesized through condensation of the 5 carbon intermediates isopentenyl diphosphate/dimethylallyl diphosphate (IPP). In higher plants, two distinct pathways for the formation of IPP exist, one in the cytosol (MVA pathway) and the other in the chloroplast (MEP pathway). Although both pathways operate fairly independently under normal conditions, interaction between them has been repeatedly reported (Laule *et al.*, 2003; Rodriguez-Concepcion *et al.*, 2004).

In order to gain better insight into the crosstalk between both pathways on the transcriptional level, gene expression patterns were monitored under various experimental conditions using 118 GeneChip® (Affymetrix) microarrays. For the construction of the genetic regulatory network, we focused on 40 genes, 16 of which were assigned to the cytosolic pathway, 19 to the plastidal pathway and 5 encode proteins located in the mi-

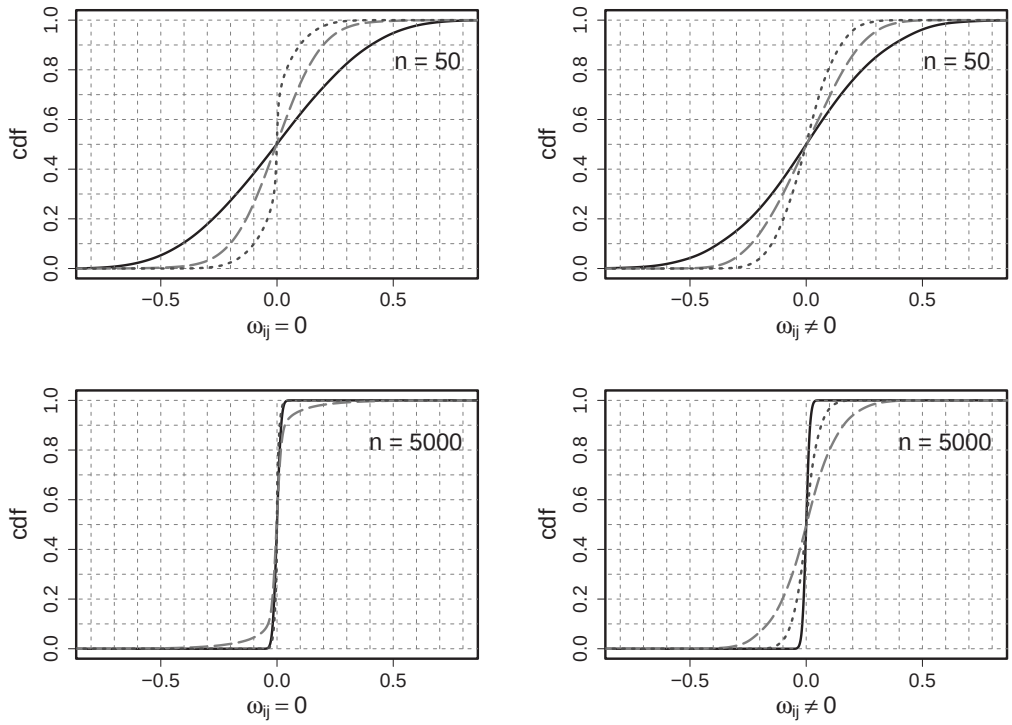

Figure 4.7: Cumulative distribution function (CDF) of the difference between sampled partial correlation coefficient $\hat{\omega}_{ij}$ and true partial correlation coefficients ω_{ij} (black line), between sampled correlation coefficients $\hat{\rho}_{ij}$ and ω_{ij} (dashed pale grey line) and sampled tri-graph coefficients $\hat{\tau}_{ij}$ and ω_{ij} (dotted grey line) for $p = 40$ and $n = 50$ (upper panel) or $n = 5000$ (lower panel) observations.

tochondrion (left panel of Figure 5.9). These 40 genes comprise not only genes of known function but also genes whose encoded proteins exhibited high homology to proteins of known functions. For reference, we adopt the notation from (Lange & Ghassemian, 2003).

The genetic interaction network among these genes was first constructed employing graphical Gaussian modeling with backward selection under the Bayesian Information Criterion (BIC) (Schwarz, 1978). This was carried out with the MIM 3.1 program (MIM, 2003). The obtained network had 178 (out of 780) edges - too many to single out biologically relevant structures. Therefore, bootstrap resampling was applied to determine the statistical confidence of the edges in the model. For the bootstrap edge probabilities, only a cutoff level as high as 0.8 led to a reasonably low number of selected edges (31 edges, right panel of Figure 5.9). However, a comparison between bootstrap edge probabilities and the pairwise correlation coefficients suggested that for such a high cutoff level, many true edges may be missed. For example, the gene AACT2 appears to be completely independent from all genes in the model although it is strongly correlated to MK, MPDC1 and FPPS2.

One conjecture for this finding would be that the simultaneous conditioning on many variables increases both the false positive and the false negative rate. Our tri-graph is expected to improve upon this drawback. Figure 5.9 shows the network model obtained

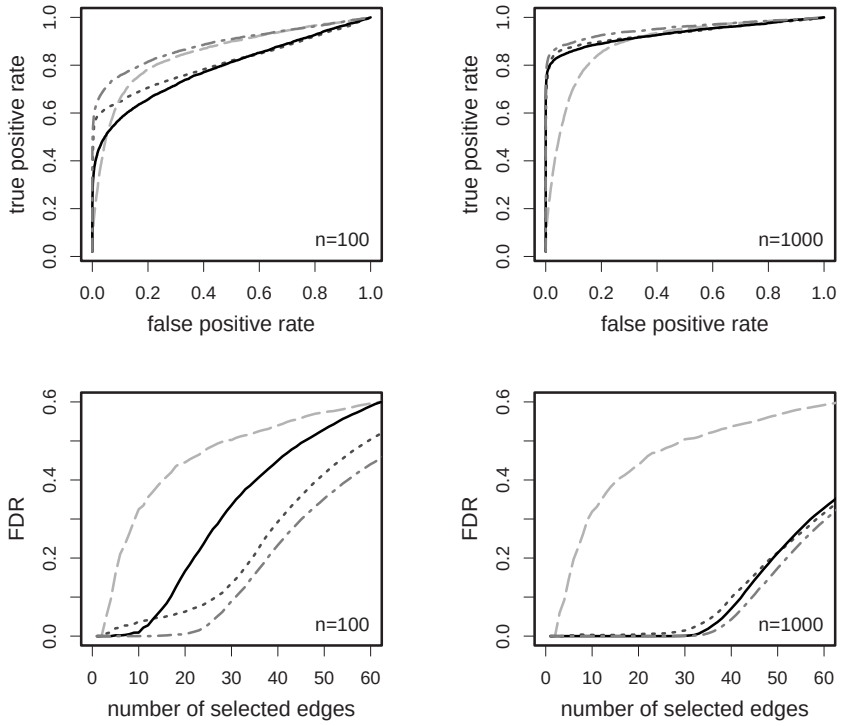

Figure 4.8: ROC curves (upper panel) and the False Discovery Rate (FDR) as a function of the number of selected edges (lower panel) for the covariance graph (dashed pale grey line), the tri-graph (dash-dotted grey line), the full conditional independence graph (black line) and the full conditional independence graph learned under backward selection (dotted dark grey line). Here, $p = 40$.

from the tri-graph. Since we find a module with strongly interconnected genes in each of the two pathways, we split up the graph into two subgraphs each displaying the subnetwork of one module and its neighbors.

In the MEP pathway, the genes DXR, MCT, CMK, and MECPS are nearly fully connected (left panel of Figure 5.9). From this group of genes, there are a few edges to genes in the MVA pathway. Among these genes, AACT1 and HMGR1 form candidates for a crosstalk between the MEP and the MVA pathway because these genes have no further connection to the MVA pathway. Their correlation to DXR, MCT, CMK, MECPS is always negative.

Similarly, the genes AACT2, HMGS, HMGR2, MK, MPDC1, FPPS1 and FPPS2 share many edges in the MVA pathway (right panel of Figure 5.9). The subgroup AACT2, MK, MPDC1, FPPS2 is completely interconnected. From these genes, we find edges to IPPI1 and GGPPS12 in the MEP pathway. Whereas IPPI1 is positively correlated to AACT2, MK, MPDC1 and FPPS2, GGPPS12 displays negative correlation to these four genes.

In the conventional graphical modeling with backward selection, we could only identify the gene module in the MEP pathway. The genes in the MVA pathway did not form a separate regulatory structure, even when the bootstrap cutoff level was lowered and as

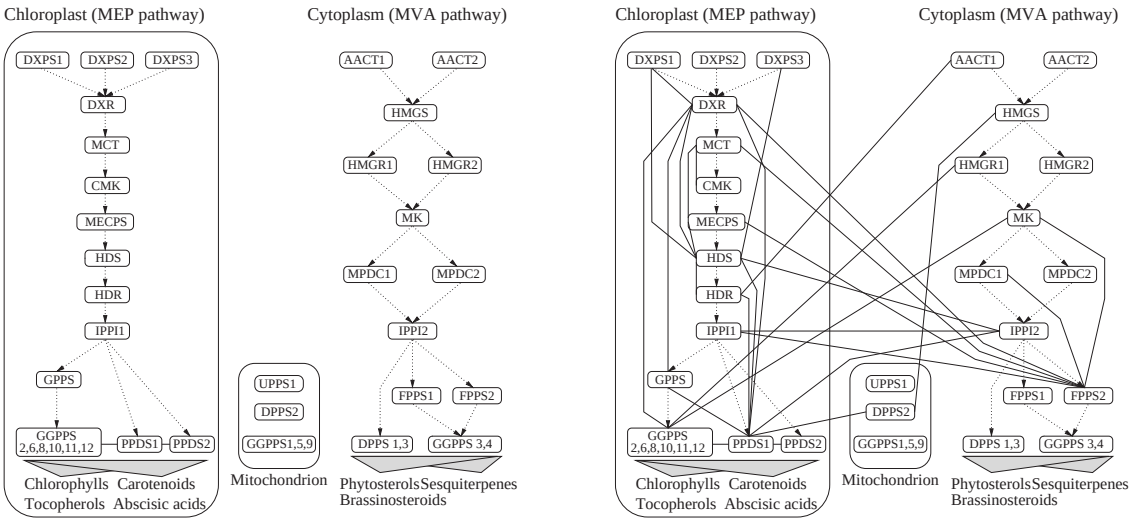

Figure 5.9: Left panel: Isoprenoid (MEP and MVA) pathway, right panel: Bootstrapped graphical model of the isoprenoid pathway (backward selection and BIC were used, bootstrap cutoff 0.8). Dotted directed edges mark the metabolic network and are not part of the model.

many edges as in the tri-graph were included in the model. In the tri-graph, the detection of the additional gene module in the MVA pathway is in good agreement with earlier findings that within a pathway, potentially many consecutive or closely positioned genes are jointly regulated (Ihmels *et al.*, 2004). Also, the aforementioned high level of coexpression between the genes AACT2, MK, MPDC1, FPPS2 suggests a separate regulatory module in the MVA pathway. In view of this and the relatively better performance of the tri-graph with respect to the false positive rate, there is some evidence for the detected structures. The biological relevance of our results will be discussed elsewhere.

6 Conclusions

Graphical Gaussian modeling suffers from unreliable estimates of the full partial correlation coefficients if the number of observations is relatively small in comparison with the number of random variables in the model. In order to still be able to analyze the conditional dependence structure between variables, we introduced a simplified measure based on separate conditioning on variables. The tri-graph coefficients proved to be powerful in two ways: First, we showed theoretically that the tri-graph coefficients show nearly the same good estimation properties as the more simple correlation coefficients. In simulations, we could even demonstrate that estimates of the tri-graph coefficients $\hat{\tau}_{ij}$ can be more accurate than those of the correlation coefficients. Second, for small sample sizes in our simulation framework, the estimated tri-graph coefficients were on average better estimators of the full partial correlation coefficients than the estimated full partial correlation coefficients themselves. In ROC curves, tri-graph estimates often performed better than more conventional graphical modeling. This finding suggests that although full partial correlation coefficients take the effect of many other variables into account, only few of these variables have a large effect on the dependence structure. In fact, Theorem 1 says that for graphs without cycles, the tri-graph and the full conditional independence graph

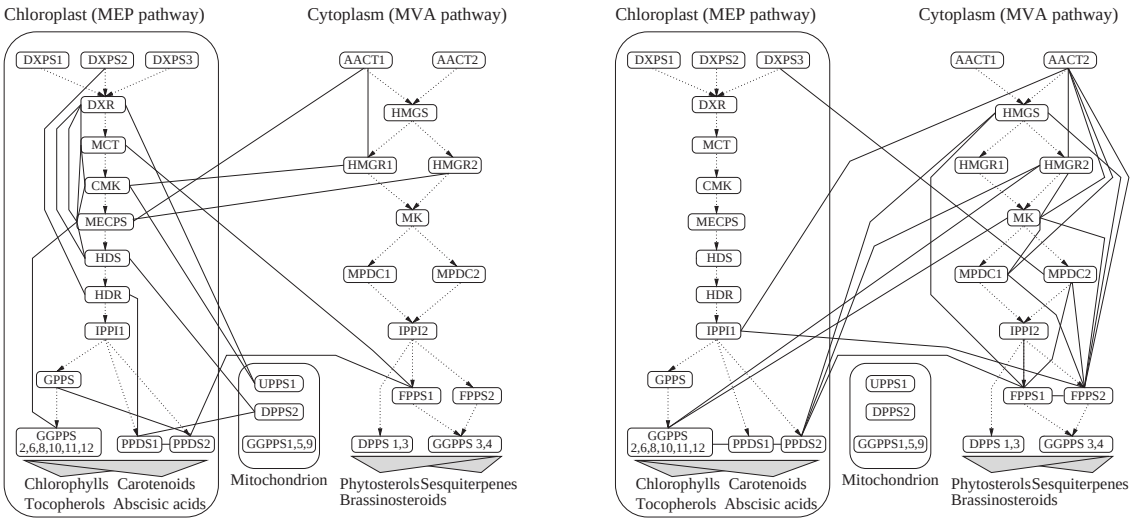

Figure 5.10: Tri-graph of the isoprenoid pathways. Left panel: subgraph of the gene module in the MEP pathway, right panel: subgraph of the gene module in the MVA pathway.

coincide.

Our tri-graph approach carries resemblance to the first two steps in the SGS and PC algorithm (Spirtes *et al.*, 2000). In both algorithms, the conditional dependence between two variables is examined based on all subsets of the remaining $p-2$ variables. In the tri-graph, the modeling is limited to subnetworks with three vertices. By this simplification, we neither have to carry out the statistically unreliable and computationally costly search for conditional independence in large subsets as in the SGS algorithm nor do we have to remove edges in a stepwise fashion as in the PC algorithm. Yet, our approach is powerful in uncovering some conditional dependence pattern between variables and we provide corrected P-values for assigning significance to edges.

By generating the number of edges in a graph according to a power law, we aimed at simulating network topologies found in biological networks. Other examples include computer and social interaction networks (Barabasi & Albert, 1999). With this restriction, only a subclass of sparse conditional independence models is considered. However, the restriction enabled us to consistently study the effect of the sample size, the number of vertices, the level of sparsity and the level of conditional dependencies on the various graphical modeling approaches.

7 Appendix

For the proof of Theorem 1, we will use a lemma that follows directly from the global Markov property for graphical Gaussian models (Lauritzen, 1996).

Lemma 1 *If there is no path between vertices i and j, then*

$$\omega_{ij|\mathbf{k}} = 0$$

for all subsets $\mathbf{k} \subseteq \{\emptyset, 1, \ldots, p\} \setminus \{i, j\}$. Moreover, if the full partial correlation coefficient

satisfies $\omega_{ij} = 0$ *and if there is exactly one path* P_{ij} *between vertices* i *and* j*, then*

$$\omega_{ij|\mathbf{k}\cup m} = 0$$

for every single node $m \in P_{ij} \setminus \{i, j\}$ *and for all subsets* $\mathbf{k} \subseteq \{\emptyset, 1, \ldots, p\} \setminus \{P_{ij}\}$.

Proof of Theorem 1:
1) There is no edges between i and j: $\omega_{ij} = 0$
Then there is either no path between i and j or exactly one path between i and j. Because of Lemma 1, there must be at least one node m such that $\omega_{ij|m} = 0$. Therefore, we have $\tau_{ij} = 0$.

2) There is an edge between i and j: $\omega_{ij} \neq 0$
Let us assume that $\tau_{ij} = 0$, which will lead to a contradiction. For $\tau_{ij} = 0$, we have $\rho_{ij} = 0$ or $\omega_{ij|k} = 0$ for a $k \in \{1, \ldots, p\} \setminus \{i, j\}$. Thus, we have $\omega_{ij|\mathbf{k}} = 0$ for a subset $\mathbf{k} \in \{\emptyset, 1, \ldots, p\} \setminus \{i, j\}$. We will show that $\omega_{ij|\mathbf{k}\cup m} = 0$ for all $m \in \{1, \ldots, p\} \setminus \{i, j, \mathbf{k}\}$.

For every node m, there is at most one path P_{im} and at most one path P_{jm} and we either have $j \in P_{im}$ or $i \in P_{jm}$, since by assumption, the full conditional independence graph contains no loops. Now, it follows directly from Lemma 1 that $\omega_{im|\mathbf{k}\cup j} = 0$ or $\omega_{jm|\mathbf{k}\cup i} = 0$. Without restriction, we assume $\omega_{im|\mathbf{k}\cup j} = 0$. From this and from the recursive formula for the partial correlation coefficients (see (Spirtes *et al.*, 2000))

$$\omega_{im|\mathbf{k}\cup j} = \frac{\omega_{im|\mathbf{k}} - \omega_{ij|\mathbf{k}}\omega_{jm|\mathbf{k}}}{\sqrt{1 - \omega^2_{ij|\mathbf{k}}}\sqrt{1 - \omega^2_{jm|\mathbf{k}}}}, \tag{7.6}$$

we obtain that $\omega_{im|\mathbf{k}} = \omega_{ij|\mathbf{k}}\omega_{jm|\mathbf{k}}$. With this equation, the recursive formula (7.6) applied to $\omega_{ij|\mathbf{k}\cup m}$ transforms to

$$\omega_{ij|\mathbf{k}\cup m} = \frac{\omega_{ij|\mathbf{k}} - \omega_{im|\mathbf{k}}\omega_{jm|\mathbf{k}}}{\sqrt{1 - \omega^2_{im|\mathbf{k}}}\sqrt{1 - \omega^2_{jm|\mathbf{k}}}} = \frac{\omega_{ij|\mathbf{k}}(1 - \omega^2_{jm|\mathbf{k}})}{\sqrt{1 - \omega^2_{im|\mathbf{k}}}\sqrt{1 - \omega^2_{jm|\mathbf{k}}}},$$

which proves that $\omega_{ij|\mathbf{k}\cup m} = 0$ if $\omega_{ij|\mathbf{k}} = 0$. By induction, we can conclude that

$$\omega_{ij} = \omega_{ij|\{1,\ldots,p\}\setminus\{i,j\}} = 0,$$

which is in contradiction to the initial assumption that i and j are connected. □

Proof of Proposition 1: Consider the hypothesis

$$H_0 = H_0(i, j) : \text{ at least one } H_0(i, j|k^*) \text{ is true for some } k^*.$$

The probability for a type I error is

$$\begin{aligned} & \mathbb{P}_{H_0}[H_0(i, j|k) \text{ rejected for all } k] = \mathbb{P}_{H_0}[\cap_k \{H_0(i, j|k) \text{ rejected}\}] \\ \leq \; & \min_k \mathbb{P}_{H_0}[H_0(i, j|k) \text{ rejected}] \leq \mathbb{P}_{H_0}[H_0(i, j|k^*) \text{ rejected}] \leq \alpha, \end{aligned}$$

where the last inequality follows from the assumption in Proposition 1. □

Proof of Proposition 2: We follow the notation from Section 3.1. Consider

$$\hat{\mu}(n)_j = n^{-1}\sum_{i=1}^{n} X_{(n),ij},\ X_{(n),ij} = (\mathbf{X}_{(n),i})_j.$$

By Markov's inequality, for $\gamma > 0$,

$$\mathbb{P}[|\hat{\mu}(n)_j - \mu(n)_j| > \gamma] \leq \gamma^{-4s}\mathbb{E}|n^{-1}\sum_{i=1}^{n} X_{(n),ij} - \mu(n)_j|^{4s},$$

and then by Rosenthal's inequality (cf (Petrov, 1975)) and our assumption (A1),

$$\mathbb{E}|n^{-1}\sum_{i=1}^{n} X_{(n),ij} - \mu(n)_j|^{4s} \leq Cn^{-2s},$$

where $C > 0$ is a constant independent from j and n. Therefore, for $\gamma > 0$,

$$\mathbb{P}[\max_{1\leq j\leq p_n} |\hat{\mu}(n)_j - \mu(n)_j| > \gamma] \leq p_n\gamma^{-4s}Cn^{-2s} = o(n^{-3s/2}),$$

due to our assumption about p_n, which proves the first claim.

For the second assertion, note that

$$\hat{\Sigma}(n)_{ij} = n^{-1}\sum_{r=1}^{n}(X_{(n),ri} - \hat{\mu}(n)_i)(X_{(n),rj} - \hat{\mu}(n)_j)$$

can be asymptotically replaced by

$$\tilde{\Sigma}(n)_{ij} = n^{-1}\sum_{r=1}^{n}(X_{(n),ri} - \mu(n)_i)(X_{(n),rj} - \mu(n)_j),$$

since by the first assertion of Proposition 2, it can be easily shown that

$$\max_{1\leq i<j\leq p_n} |\hat{\Sigma}(n)_{ij} - \tilde{\Sigma}(n)_{ij}| = o_P(1). \tag{7.7}$$

Similarly as for the mean, we get for $\gamma > 0$,

$$\mathbb{P}[|\tilde{\Sigma}(n)_{ij} - \Sigma(n)_{ij}| > \gamma] \leq \gamma^{-2s}\mathbb{E}|n^{-1}\sum_{r=1}^{n} Y_r(i,j)|^{2s},$$
$$Y_r(i,j) = (X_{(n),ri} - \mu(n)_i)(X_{(n),rj} - \mu(n)_j) - \Sigma(n)_{ij},$$

and by Rosenthal's inequality (cf (Petrov, 1975)) and assumption (A1),

$$\mathbb{E}|n^{-1}\sum_{r=1}^{n} Y_r(i,j)|^{2s} \leq Cn^{-s},$$

where $C > 0$ is a constant, independent of j. Note that our assumption (A1) implies that the moments of order $2s$ of the $Y_r(i,j)$ variables are uniformly bounded. Therefore

$$\mathbb{P}[\max_{1\leq i<j\leq p_n} |\tilde{\Sigma}(n)_{ij} - \Sigma(n)_{ij}| > \gamma] \leq p_n^2\gamma^{-2s}Cn^{-s} = o(1),$$

by our assumption about p_n. This, together with (7.7) completes the proof for the second assertion of the Proposition. □

Proof of Theorem 2: The first assumption in (A2) and the uniform convergence from Proposition 2 imply that

$$\max_{1\le i<j\le p_n} |\hat{\rho}(n)_{ij} - \rho(n)_{ij}| = o_P(1) \ (n \to \infty). \tag{7.8}$$

Furthermore, we can use a Taylor expansion for the partial correlations:

$$\hat{\omega}(n)_{ij|k} - \omega(n)_{ij|k} = \frac{x - yz}{uv} - \frac{x_0 - y_0 z_0}{u_0 v_0} = \frac{x - x_0}{u_0 v_0} - \frac{yz - y_0 z_0}{u_0 v_0} - \frac{1}{\tilde{u}^2 \tilde{v}^2}(uv - u_0 v_0)(x - yz),$$

where $|\tilde{u}\tilde{v} - u_0 v_0| \le |uv - u_0 v_0|$, and $x = \hat{\rho}(n)_{ij}, y = \hat{\rho}(n)_{ik}, z = \hat{\rho}(n)_{jk}, u = \sqrt{1 - \hat{\rho}(n)^2_{ik}}, v = \sqrt{1 - \hat{\rho}(n)^2_{jk}}$ and x_0, y_0, z_0, u_0, v_0 the corresponding true population quantities. We now get the assertion of Theorem 2 by the uniform convergence of the correlations in (7.8) and by using the second assumption in (A2) which guarantees that the denominator in $1/(u_0 v_0)$ is bounded and that $\frac{1}{\tilde{u}^2 \tilde{v}^2} = o_P(1)$ uniformly with respect to i, j, k. □

References

Barabasi, A. & Albert, R. (1999) Emergence of scaling in random networks. *Science,* **286** (5439), 509–512.

Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B,* **57**, 289–300.

Cox, D. R. & Wermuth, N. (1993) Linear dependencies represented by chain graphs (with discussion). *Statist Sci,* **8**, 204–218.

Cox, D. R. & Wermuth, N. (1996) *Multivariate dependencies: models analysis and interpretation.* Chapman & Hall, London.

Dobra, A., Hans, C., Jones, B., Nevins, J., Yao, G. & West, M. (2004) Sparse graphical models for exploring gene expression data. *Journal of Mult Analysis,* **90**, 196–212.

Edwards, D. (2000) *Introduction to Graphical Modelling.* Springer Verlag; 2nd edition.

Friedman, N., Linial, M., Nachman, I. & Pe'er, D. (2000) Using bayesian networks to analyze expression data. *J Comput Biol,* **7** (3-4), 601–620.

Giudici, P. & Green, P. (1999) Decomposable graphical gaussian model determination. *Biometrika,* **86**, 785–801.

Hartemink, A. J., Gifford, D. K., Jaakkola, T. S. & Young, R. A. (2001) Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks. In *Pac Symp Biocomput* PSB01 pp. 422–433.

Husmeier, D. (2003) Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic bayesian networks. *Bioinformatics,* **19** (17), 2271–2282.

Ihmels, J., Levy, R. & Barkai, N. (2004) Principles of transcriptional control in the metabolic network of saccharomyces cerevisiae. *Nat Biotechnol,* **22** (1), 86–92.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabasi, A. L. (2000) The large-scale organization of metabolic networks. *Nature,* **407** (6804), 651–654.

Lange, B. & Ghassemian, M. (2003) Genome organization in arabidopsis thaliana: a survey for genes involved in isoprenoid and chlorophyll metabolism. *Plant Mol Biol,* **51** (6), 925–948.

Laule, O., Fürholz, A., Chang, H. S., Zhu, T., Wang, X., Heifetz, P. B., Gruissem, W. & Lange, M. (2003) Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in arabidopsis thaliana. *Proc Natl Acad Sci U S A,* **100** (11), 6866–6871.

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

Madigan, D. & Raftery, A. (1994) Model selection and accounting for model uncertainty in graphical models using occam's window. *J Amer Statist Assoc,* **89**, 1535–1546.

Maslov, S. & Sneppen, K. (2002) Specificity and stability in topology of protein networks. *Science,* **296** (5569), 910–913.

Meinshausen, N. & Bühlmann, P. (2004). Consistent neighbourhood selection for sparse high-dimensional graphs with the lasso. Technical report ETH Zurich.

MIM (2003). Student version 3.1. http://www.hypergraph.dk.

Petrov, V. (1975) *Sums of independent random variables*. Springer, Berlin.

Rodriguez-Concepcion, M., Fores, O., Martinez-Garcia, J. F., Gonzalez, V., Phillips, M., Ferrer, A. & Boronat, A. (2004) Distinct light-mediated pathways regulate the biosynthesis and exchange of isoprenoid precursors during arabidopsis seedling development. *Plant Cell,* **16** (1), 144–156.

Schwarz, G. (1978) Estimating the dimension of a model. *Ann of Statistics,* **6**, 461–464.

Spirtes, P., Glymour, C. & Scheines, R. (2000) *Causation, Prediction, and Search*. 2nd edition,, MIT Press.

Toh, H. & Horimoto, K. (2002) Inference of a genetic network by a combined approach of cluster analysis and graphical gaussian modeling. *Bioinformatics,* **18** (2), 287–297.

Waddell, P. J. & Kishino, H. (2000) Cluster inference methods and graphical models evaluated on nci60 microarray gene expression data. *Genome Informatics,* **11**, 129–140.

Wang, J., Myklebost, O. & Hovig, E. (2003) Mgraph: graphical models for microarray data analysis. *Bioinformatics,* **19** (17), 2210–2211.