Vanna Albieri and Vanessa Didelez*

Comparison of statistical methods for finding network motifs

Abstract: There has been much recent interest in systems biology for investigating the structure of gene regulatory systems. Such networks are often formed of specific patterns, or network motifs, that are interesting from a biological point of view. Our aim in the present paper is to compare statistical methods specifically with regard to the question of how well they can detect such motifs. One popular approach is by network analysis with Gaussian graphical models (GGMs), which are statistical models associated with undirected graphs, where vertices of the graph represent genes and edges indicate regulatory interactions. Gene expression microarray data allow us to observe the amount of mRNA simultaneously for a large number of genes *p* under different experimental conditions *n*, where *p* is usually much larger than *n* prohibiting the use of standard methods. We therefore compare the performance of a number of procedures that have been specifically designed to address this large *p*-small *n* issue: G-Lasso estimation, Neighbourhood selection, Shrinkage estimation using empirical Bayes for model selection, and PC-algorithm. We found that all approaches performed poorly on the benchmark *E. coli* network. Hence we systematically studied their ability to detect specific network motifs, pairs, hubs and cascades, in extensive simulations. We conclude that all methods have difficulty detecting hubs, but the PC-algorithm is most promising.

Keywords: Gaussian graphical models; Lasso; PC-algorithm; Shrinkage.

DOI 10.1515/sagmb-2013-0017

1 Introduction

Recent progress in molecular biology has led to an unprecedented growth in molecular data which has prompted interest in Systems Biology for reconstructing the structure and dynamics of biological processes such as gene regulatory networks (GRNs). In this context, networks are formed by specific patterns, often called network motifs, that are interesting from a biological point of view. These can be interpreted as the functional units which combine to regulate the cellular behaviour as a whole of bacterias and higher organism (Milo et al., 2002; Alon, 2007). Our aim in the present paper is to compare statistical methods specifically with regard to the question of how well they can detect such network structures.

Gaussian Graphical models (GGMs) have become a common tool for structural learning of GRNs (Friedman, 2004), and methods have been developed to deal with the situation where the number of variables *p* (genes) is large compared to the number *n* of observations as is very common for these kinds of data. Here, we specifically consider the Neighbourhood selection (Meinshausen and Bühlmann, 2006), the G-Lasso algorithm (Friedman et al., 2008), the Shrinkage estimator with empirical Bayes approach for model selection (Schäfer and Strimmer, 2005a,b), and the PC-algorithm (Kalisch and Bühlmann, 2007). We start by comparing the methods on the *Escherichia coli* data for which the "true" transcriptional network is known (Gama-Castro et al., 2008) and contains some typical motif, namely two "hubs," i.e., single genes that are highly connected to many other genes. As we find that all approaches performed poorly on the *E. coli* network, we

***Corresponding author: Vanessa Didelez,** School of Mathematics, University of Bristol University Walk, Bristol BS81TW, UK, e-mail: vanessa.didelez@bristol.ac.uk
Vanna Albieri: Statistics, Bioinformatics and Registry, Danish Cancer Society Research Center, Copenhagen, Denmark

then study systematically their ability to detect GRN structures by focusing on three types of motifs, hubs, pairwise structures and chains.

The paper is organised as follows. Section 2 gives a brief overview of the methods compared in our study. Section 3 describes the performance measures used to evaluate the procedures. We present the real and synthetic data as well as the results of our comparative study in Section 4 and Section 5, respectively. Finally, conclusions and outlook are presented in Section 6.

2 Methods

2.1 Gaussian graphical models (GGMs)

Let $\mathscr{G}=(V, E)$ be an undirected graph with a finite set of vertices $V=\{1, \ldots, p\}$ and a set of edges $E\subseteq V\times V$. Its adjacency matrix $A=\{a_{ij}\}$ has $a_{ij}=a_{ji}=1$ if i, j are neighbours, i.e., if $\{i, j\}\in E$, and zero otherwise. Let $\mathbf{X}_V\equiv\mathbf{X}$, with $V=\{1, \ldots, p\}$, be a continuous random vector with joint Normal $N_p(\mu, \Sigma)$ distribution, mean vector $\mu=(\mu_1, \ldots, \mu_p)^T$ and a positive definite covariance matrix $\Sigma=\{\sigma_{ij}\}$, $1\leq i, j\leq p$. A Gaussian graphical model (GGM) with graph $\mathscr{G}$ is the family of normal distributions for $\mathbf{X}$ that satisfy the undirected Markov property with respect to $\mathscr{G}$ (Lauritzen, 1996), which means that

$$\{i, j\}\notin E \quad\Rightarrow\quad X_i \perp\!\!\!\perp X_j \,|\, X_{V\setminus\{i,j\}}.$$

Here $A\perp\!\!\!\perp B|C$ means that A is conditionally independent of B given C (Dawid, 1979). Since $X\sim N_p(\mu, \Sigma)$, it holds that

$$X_i \perp\!\!\!\perp X_j \,|\, X_{V\setminus\{i,j\}} \Leftrightarrow \rho_{ij\cdot V\setminus\{i,j\}}=0,$$

where $\rho_{ij\cdot V/\{i, j\}}$ is the partial correlation coefficient, i.e., the correlation between X_i and X_j after removing the linear relations with all remaining variables $X_{V\setminus\{i, j\}}$. The partial correlation can be expressed in terms of the concentration matrix where $\Sigma^{-1}=\Omega=\{\omega_{ij}\}$ as

$$\rho_{ij\cdot V\setminus\{i,j\}}=-\frac{\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}} \tag{1}$$

(see Lauritzen, 1996, p. 130).

Given a random sample of n observations of $\mathbf{X}$, a standard procedure for learning a GGM is as follows. We estimate Σ by the sample covariance matrix $S=\hat{\Sigma}$, compute its inverse S^{-1} as an estimate of $\hat{\Omega}$, and obtain the sample partial correlations using (1). The graph edges are determined by those partial correlations found to be "large enough." The decision could for example be based on the $p(p-1)/2$ statistical tests for

$$H_0: \rho_{ij\cdot V\setminus\{i,j\}}=0 \quad \text{vs} \quad H_1: \rho_{ij\cdot V\setminus\{i,j\}}\neq 0.$$

Here, we use the test statistic given by

$$t=\sqrt{n-p}\,\frac{\hat{\rho}_{ij\cdot V\setminus\{i,j\}}}{\sqrt{1-(\hat{\rho}_{ij\cdot V\setminus\{i,j\}})^2}},$$

which has a Student's t distribution with $n-p$ degrees of freedom under the null hypothesis that $\rho_{ij\cdot V\setminus\{i, j\}}=0$ (Lauritzen, 1996, Section 5.3.3). Due to the large number of tests, a multiple testing adjustment will usually be applied; here we use the false discovery rate correction (FDR) (Benjamini and Hochberg, 1995).

The above procedure cannot be used when $n<p$ as S will not be invertible; this has lead to the development of alternative methods as follows.

2.2 Lasso penalisation

A general approach to the *large p – small n* problem is to penalise model complexity. We consider two approaches inspired by the Lasso method for regressions (Tibshirani, 1996): Neighbourhood selection and the Graphical Lasso.

2.2.1 Neighbourhood selection

The approach of Meinshausen and Bühlmann (2006) does not aim to estimate Ω itself. Instead it reconstructs the network using the fact that in an undirected graphical model, each variable (node) is conditionally independent of all other variables that are not its graph neighbours, given these neighbours. For a GGM this means that if we linearly regress a variable on all other variables, the only ones that have non-zero coefficients are its neighbours. When $n<p$, the Lasso approach uses an ℓ_1 penalisation on the regression coefficients. More formally, let X_a be the vector of observations on variable a, and let $\mathbf{X}_{-a}$ be the matrix of observations on all other variables, then the Lasso estimate $\hat{\theta}^a$ of the regression coefficients is given by

$$\arg\min n^{-1} || X_a - \mathbf{X}_{-a}\theta ||_2^2 + \lambda || \theta ||_1,$$

where λ is a tuning parameter. The larger λ the sparser the solution, i.e., more zero coefficients in $\hat{\theta}^a$. Meinshausen and Bühlmann (2006) show that choosing λ based on prediction optimality (e.g., cross-validation) leads to an inconsistent estimation of the neighbourhoods, and recommend to choose it larger.

The network structure is determined by the zero pattern of the coefficients. In practice (for finite samples) it is possible that the regression coefficient for X_a on X_b is zero while for X_b on X_a it is non-zero. This can be resolved in two obvious ways; here we use the AND rule, i.e., add an $\{a, b\}$–edge only if both are non-zero. Under the assumptions stated by Meinshausen and Bühlmann (2006) there will be no difference between the two rules for large n, and for appropriate choices of λ the correct neighbourhoods will be selected.

2.2.2 G-Lasso

As proposed by Friedman et al. (2008), it is also possible to estimate the concentration matrix $\Omega=\Sigma^{-1}$ itself using an ℓ_1 penalisation on its entries. The approach simultaneously performs (sparse) parameter estimation and model selection as the entries estimated to be zero can immediately be translated into the absence of edges in the network.

As before, inference is based on imposing a penalty term λ and using the ℓ_1–norm to estimate $\hat{\Omega}$, i.e., we maximize the penalised log-likelihood

$$\arg\max \log\det\Omega - \mathrm{tr}(S\Omega) - \lambda||\Omega||_1.$$

To solve the optimisation problem the blockwise coordinate descent algorithm introduced by Banerjee et al. (2008) is used.

The authors do not make any suggestion for the selection of λ, but they make two important remarks. First, setting $\lambda=0$ the algorithm computes the maximum likelihood estimator S^{-1} (if it exists) using a linear regression at each stage. Second, the penalty term can be either a scalar or a matrix, the latter allows us to penalize each inverse covariance element by a different amount.

The G-Lasso appears to be better targeted at the problem than the above Neighbourhood selection. However, as explained in Meinshausen (2008), it is in fact not consistent for some graphs, regardless of the choice of λ. We therefore include both methods in our comparison. The size of λ across the two Lasso methods is not comparable as it penalises different types of quantities, regression parameters in the Neighbourhood selection case, and elements of the inverse covariance in the G-Lasso case.

2.3 Shrinkage estimation and empirical Bayes approach

An alternative approach is to find an estimator of $\sum$ that can be inverted even when $p>n$. Schäfer and Strimmer (2005a,b) propose to shrink the empirical (unbiased) covariance estimator S towards an invertible (but possibly biased) estimator T. The weighted shrinkage estimator $S^\star$ for the covariance matrix is given by

$$S^\star=\lambda T+(1-\lambda)S,$$

where the shrinkage parameter λ is obtained by minimizing a risk function, e.g., the MSE, and depends on the covariance target T (see Schäfer and Strimmer, 2005b, Table 2, for different choices of T and the relative values of λ).

The model selection procedure, i.e., the decision of which partial correlations $\rho_{ij\cdot V\setminus\{i,j\}}$ are "close" to zero, is based on an empirical Bayes approach. Their assumed distribution across edges which is taken as the mixture

$$f(\rho)=\eta_0 f_0(\rho;k)+(1-\eta_0)f_A(\rho),$$

where ρ is a place-holder for the partial correlations. Here, f_0 is the null distribution [see Hotelling (1953)], while $f_A \sim \mathscr{U}(-1,1)$ is assumed to be the distribution of observed partial correlations, k is the degrees of freedom, and η_0 is the (unknown) proportion of missing edges. Fitting this mixture distribution to the observed partial correlation coefficients allows us to infer the parameters $\hat{\eta}_0$ and $\hat{k}$. It is then straightforward to compute two-sided p-values for all edges using the null distribution f_0 with $\hat{k}$ as plug-in estimate. In our comparison, we will use the shrinkage estimator with "diagonal-unequal variance" target matrix T, i.e., $t_{ij}=s_{ii}$ if $i=j$, and $t_{ij}=0$ if $i\neq j$, and then we apply the above empirical Bayes approach to determine the graph.

For the tests involved in the shrinkage approach (as well as MLE where applicable), we use the false discovery rate correction (FDR) (Benjamini and Hochberg, 1995) at overall level α to correct for the multiple testing problem. When we present a single result we use FDR $\alpha=0.05$. The FDR decision rule requires also specification of the fraction of true non-edges η_0. For the *E. coli* data, we set η_0 equal to the number of non-edges derived from the benchmark transcriptional network. For the simulated data, we set η_0 equal to the true number of non-edges for each synthetic network. This provides a slight advantage as the true values would not be known in practice.

2.4 PC-algorithm

In contrast to the previous methods, the PC-algorithm aims to find a directed acyclic graph (DAGs) rather than an undirected graph (Spirtes and Glymour, 1991). The PC-algorithm starts with a complete undirected graph and successively deletes edges based on conditional independence decisions, resulting in an undirected graph (the skeleton) which can then be partially oriented and extended to an equivalence class of DAGs. Among others, this procedure assumes that the true distribution is "faithful" to some DAG, which means that all conditional independencies correspond to separations found in this DAG. Violations could occur if for instance positive and negative correlations via different paths cancel each other out which is generally regarded as unlikely in practice.

All statistical inference is carried out when the algorithm establishes the skeleton by testing marginal independencies, then conditional independencies given a set of size one, then the size of the conditioning sets is increased in each step. One can show that for the conditioning set it is sufficient to consider the set of

Table 1 (A) Table for classification of edges. (B) Performance measures for the evaluation of the methods.

	True graph	
	Edges	**Non-edges**
Estimated edges	Tp	Fp
Estimated non-edges	Fn	Tn
	P=(Tp+Fn)	N=(Fp+Tn)

(A)

Precision rate	Prec=$Tp/(Tp+Fp)$.
True Positive rate (recall)	Tpr=Tp/P.
Accuracy	Acc=$(Tp+Tn)/(P+N)$.
Error rate	Err=$(Fp+Fn)/(P+N)$.
False Positive rate	Fpr=Fp/N.
False Negative rate	Fnr=Fn/P.
True Negative rate	Tnr=Tn/N.

(B)

current adjacent nodes for any given node. The rationale of using the PC-algorithm when $p>n$ is that if the true graph is sparse, then the separating sets between any two nodes are small so that the algorithm can stop early without considering conditioning sets of arbitrary size.

Due to the changing conditioning sets, it is useful that the partial correlations are recursively related as follows, where we assume that the partial correlations without variable h are already known,

$$\rho_{ij\cdot\mathbf{k}}=\frac{\rho_{ij\cdot\mathbf{k}\backslash h}-\rho_{ih\cdot\mathbf{k}\backslash h}\,\rho_{jh\cdot\mathbf{k}\backslash h}}{\sqrt{(1-\rho^2_{ih\cdot\mathbf{k}\backslash h})(1-\rho^2_{jh\cdot\mathbf{k}\backslash h})}},$$

for some $h\in\mathbf{k}$, with $\mathbf{k}\subseteq V\backslash\{i, j\}$. Kalisch and Bühlmann (2007) apply Fisher's z-transform

$$Z(ij\cdot\mathbf{k})=\frac{1}{2}\log\left(\frac{1+\hat{\rho}_{ij\cdot\mathbf{k}}}{1-\hat{\rho}_{ij\cdot\mathbf{k}}}\right)$$

to the estimated partial correlation coefficients obtaining a suitable test statistic. Using the *nominal* significance level α, the null hypothesis $H_0(ij\cdot\mathbf{k})$:$\rho_{ij\cdot\mathbf{k}}=0$ against the two-sided alternative $H_1(ij\cdot\mathbf{k})$:$\rho_{ij\cdot\mathbf{k}}\neq 0$ is rejected if $\sqrt{n-|\mathbf{k}|-3}\,|Z(ij\cdot\mathbf{k})|>\Phi^{-1}(1-\alpha/2)$. Due to the indirect way the PC-algorithm produces the final graph, it is difficult to adjust the nominal significance level α so as to obtain a given overall error probability.

In our comparative study, we consider three different undirected graphical structures (motifs), all of which are equivalent to some DAG. However, the estimated DAG may not be equivalent to an undirected graph, so that, to ensure a fair comparison, we moralise the estimated DAG by (i) adding an undirected edge between every pair of non-adjacent vertices that have a common child and (ii) turning all directed edges into undirected edges (Lauritzen, 1996, p. 7). In practice the true graph is unknown and the appropriate undirected graph to be compared with the other methods is then the moralised output of the PC-algorithm. In general one does not know the true structure so that a DAG itself may or may not be more appropriate than an undirected graph.

3 Performance measures

In order to evaluate the above approaches, we classify the edges for each method as: True positive (Tp), True negative (Tn), False positive (Fp), and False negative (Fn) (Table 1A). Based on the resulting frequencies we calculate the quantities in Table 1B. For our simulation study we repeat this 2000 times for p=20 and 100 times for p=100, 200. As we specifically want to investigate the ability of the methods to detect different motifs, we must take into account not so much whether the Fpr is low, but whether the false positive edges are systematically false as this can hide a motif. Therefore we focus here on adjacency plots and Precision-Recall (PR) curves as summaries of the above measures.

An adjacency matrix can easily be visualised with black points indicating edges; Figures 1 and 5 show these for the *E. coli* data and the true simulation settings, respectively. For the actual simulations, we use the gray intensity to represent the proportion of times a given method has found an edge over all the replications (black=100%, white=0%).

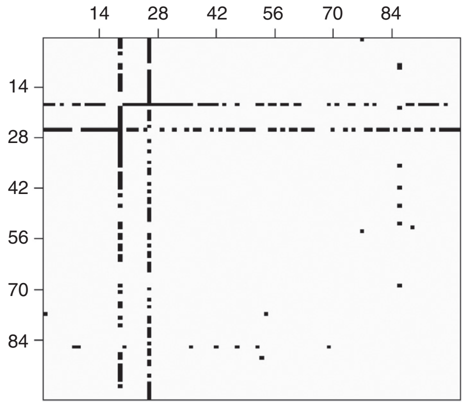

Figure 1 The benchmark adjacency matrix for *E. coli* data with 100 genes.

For the interpretation of PR curves one can regard the recall as an estimate of the probability that a true edge is indeed detected by a given method, while its precision is the probability that an estimated edge is indeed true. The ideal PR curve has a precision of one for recall between zero and one and then drops steeply to the proportion of true edges $P/(P+N)$ when recall tends towards one. To determine such curve for our methods we obtain different recall values by varying the tuning parameters, i.e., the levels α (for MLE, Shrinkage, and PC-algorithm), and the penalty parameters λ (for Neighbourhood selection and G-Lasso). Further, for the Lasso methods and the PC-algorithm, similar recall values are grouped together and the average precision is plotted. This is necessary, because we need to re-estimate the concentration matrices (or the graph) for different values of the tuning parameters, whereas with MLE and Shrinkage estimation the threshold can be varied using a single estimate of this matrix. We use PR curves instead of the well-known ROC curves since the latter are based on the true positive rate (sensitivity/recall) and false positive rate (1-specificity). With sparse graphs it is not of much interest to consider settings where the false positive rate gets large because the denominator N is very large and it requires unreasonable values of the tuning parameters to make the numerator large (but see supplementary material, Appendices B and G).

The implementation of the methods we compare uses the following R packages: “ggm” for the MLE, “glasso” for Neighbourhood Selection and G-Lasso, “GeneNet” and “fdrtool” for Shrinkage estimation and FDR adjustment, and finally “pcalg” for the PC-algorithm. The synthetic data was created with “mvtnorm”. PR-curves were computed using the R package “ROCR” (Sing et al., 2005).

Table 2 Performance measures for *E. coli* data for selected values of tuning parameters.

λ	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
5e-05	0.6953	0.0334	0.4719	0.5281	0.5340	0.3046	0.4659
0.005	0.1796	0.0594	0.9052	0.0947	0.0755	0.8203	0.9245
0.2	0.0234	0.1034	0.9695	0.0305	0.0054	0.9765	0.9946
0.4	0.0234	0.1667	0.9717	0.0283	0.0031	0.9766	0.9969
0.5	0.0078	0.0667	0.9715	0.0285	0.0029	0.9922	0.9971
0.8	0.0000	0.0000	0.9721	0.0279	0.0021	1.0000	0.9979

(A) Neighbourhood selection

λ	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
5e-05	0.8828	0.0319	0.3050	0.6949	0.7102	0.1171	0.2897
0.005	0.6406	0.0486	0.6667	0.3333	0.3326	0.3593	0.6673
0.2	0.1640	0.0245	0.8099	0.1901	0.1729	0.8359	0.8270
0.4	0.1250	0.0216	0.8309	0.1691	0.1504	0.8750	0.8497
0.5	0.1250	0.0234	0.8430	0.1569	0.1379	0.8750	0.8620
0.8	0.0468	0.0124	0.8789	0.1210	0.0989	0.9531	0.9010

(B) G–Lasso

α	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
1e-08	0.0078	0.0322	0.9682	0.0317	0.0062	0.9921	0.9937
1e-05	0.0156	0.0465	0.9662	0.0337	0.0085	0.9843	0.9914
0.001	0.0312	0.0655	0.9634	0.0365	0.0118	0.9687	0.9881
0.05	0.0390	0.0454	0.9539	0.0460	0.0217	0.9609	0.9782
0.1	0.0468	0.0441	0.9490	0.0509	0.0269	0.9531	0.9730

(C) PC–Algorithm

Tpr	0.0625
Precision	0.1142
Accuracy	0.9632
Error Rate	0.0367
Fpr	0.0128
Fnr	0.9375
Tnr	0.9871

(D) Shrinkage

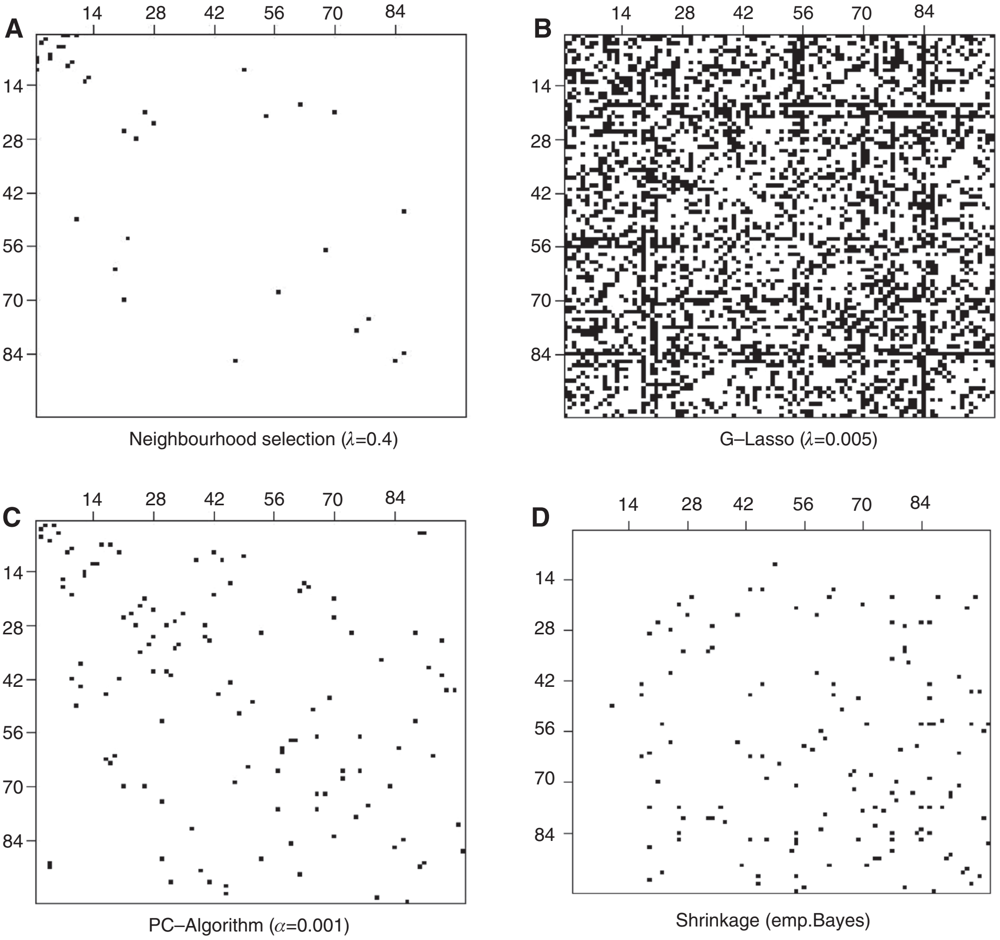

Figure 2 Plot of adjacency matrices for *E. coli* data. Tuning parameters were chosen to maximise the precision of each method.

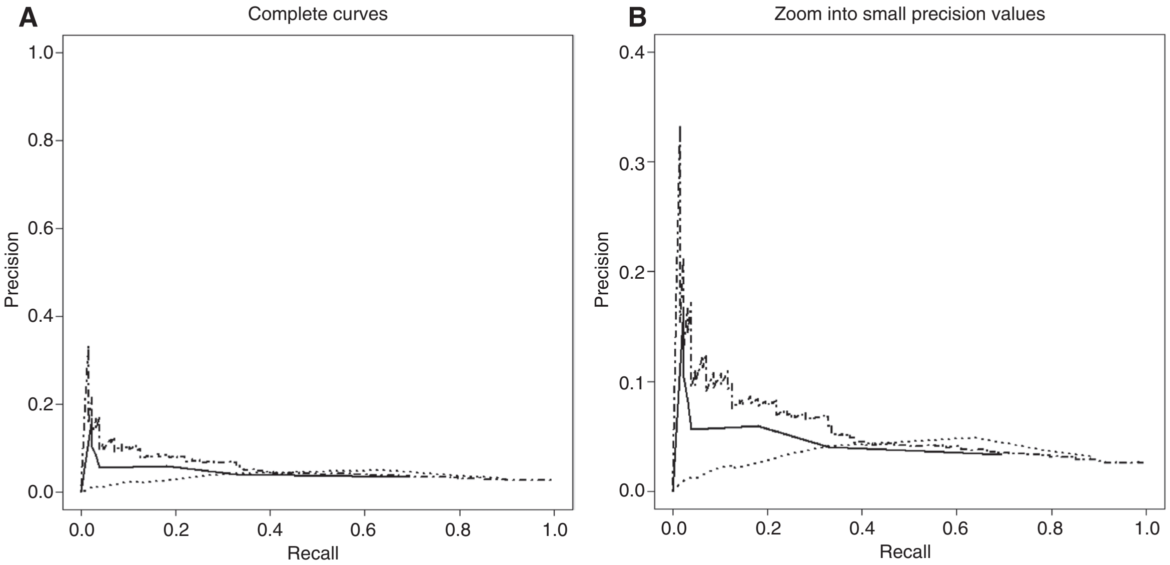

Figure 3 Precision-Recall curves for *E. coli* data: dotted for G-Lasso, solid for Neighbourhood selection, and dotdash for Shrinkage estimator. Note that no curve for the PC-algorithm is given as the range of its recall values is too small.

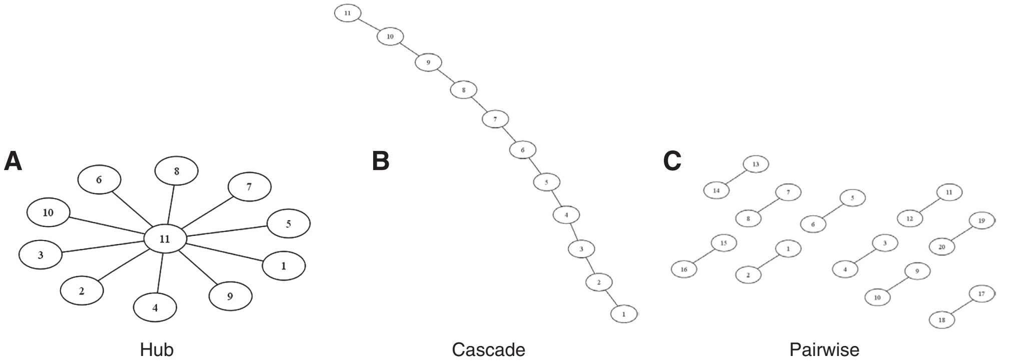

Figure 4 Network motifs under study.

4 Comparison using *Escherichia coli* data

4.1 *Escherichia coli* data

We consider the microarray data of *Escherichia coli* (*E. coli*) provided by the EcoliOxygen data file in the R package "qpgraph" (Castelo and Roverato, 2009). The data are from n=43 experiments of various mutants under oxygen deprivation (Covert et al., 2004) [downloaded from the Gene Expression Omnibus (Barrett et al., 2007) with accession GDS680]. The mutants were designed to monitor the response from *E. coli* during an oxygen shift in order to target the *a priori* most relevant part of the transcriptional network. In addition, the EcoliOxygen data file contains the *E. coli* transcriptional network from RegulonDB (Gama-Castro et al., 2008). This is a database that collects the available experimental data on regulatory interactions between transcription factor (TF) and their target genes (TG) in *E. coli*.

We obtained our "gold-standard network" by following the instructions of Castelo and Roverato (2009): filter the expression profile data in EcoliOxygen retaining only those genes forming part of the RegulonDB regulatory modules of the five knocked-out transcription factors. This results in a reduction to p=378 genes involved in only 681 interactions out of 71,253 interactions in the complete network. For simplicity, we further reduced the data set by retaining only those 100 genes with largest variability, in terms of expression profile, measured by the interquartile range. Hence, our final *E. coli* data set has p=100 and n=43, with 128 interactions (edges) out of a possible 4950 interactions in the complete network, i.e., $P/(P+N)$=2.6%. Figure 1 shows the resulting "gold-standard" adjacency matrix of the network.

4.2 Results of the analysis with *E. coli* data

From Table 2, Figures 2 and 3 it is clear that none of the methods recover interesting aspects of the benchmark network correctly (further plots are given in supplementary material, Appendix A). Based on the PR-curves one could say that the Shrinkage approach performs best, but it is not much better than random guessing.

The true network is mainly formed by two hub motifs, i.e., two genes that have a high number of interactions with others genes. This particular structure could be an explanation for the poor performance of the four methods. Indeed, this structure violates the definition of sparseness used in almost all structural learning algorithms where every gene is expected to have only few interactions with others genes. However, one has to keep in mind as alternative explanations for the disappointing performances that what we use as the benchmark network may not in fact be the true network, or that the assumptions of a GGM are not suitable in this application.

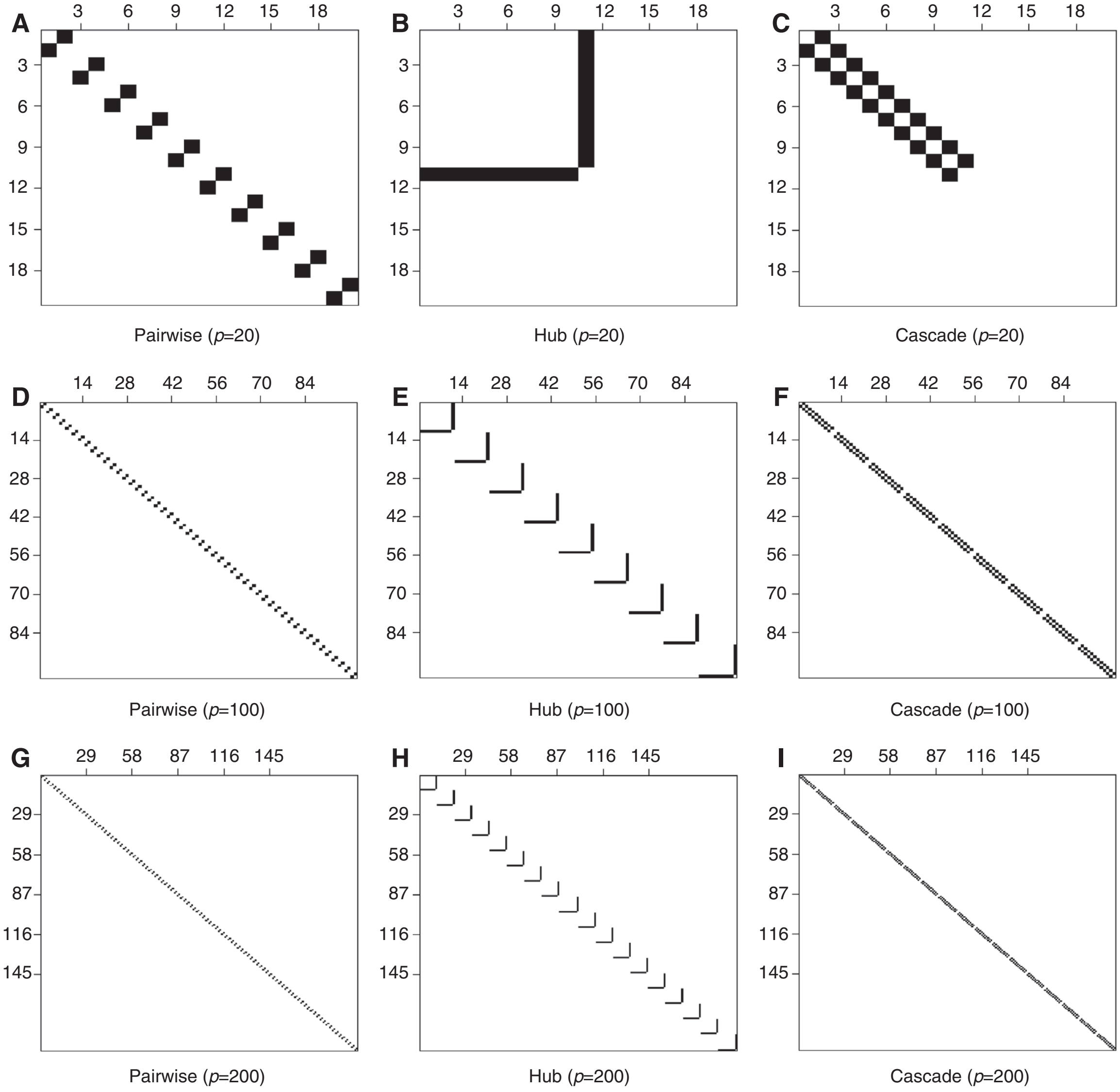

Figure 5 Plot of the true adjacency matrix for the synthetic data.

5 Comparative study with synthetic data

5.1 Synthetic data

For the generation of the synthetic data, we take into consideration both the information on the network structure obtained previously from the *E. coli* data and other network motifs that have been found of interest (Milo et al., 2002; Alon, 2007). Hence, we generated synthetic data based on three types of motifs: hub structure, cascade structure, and pairwise structure (Figure 4). The hub (a) is a common type of network motif in a gene regulatory network, but it is also one of the most difficult structures to be discovered by structural learning algorithms. Indeed, an upper bound on the number of neighbours of any vertex is commonly assumed which excludes the presence of hubs from the network. The cascade (b) is represented by a sequence of interactions between genes, in which every gene has at least one and at most two connections. Finally, the pairwise structure (c) refers to the simple case where only pairs of genes are connected. For interpretation, it is important to underline a statistical particularity of this last structure: the marginal independencies coincide with the con-

Table 3 Simulation results for Neighbourhood selection (n=150 throughout).

λ	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
5e-04	1.0000	0.0532	0.0626	0.9374	0.9894	0.0000	0.0106
0.005	1.0000	0.0584	0.1514	0.8486	0.8958	0.0000	0.1042
0.05	1.0000	0.1718	0.7433	0.2567	0.2710	0.0000	0.7290
0.1	1.0000	0.4829	0.9411	0.0589	0.0622	0.0000	0.9378
0.2	1.0000	0.9786	0.9987	0.0013	0.0013	0.0000	0.9987
0.3	0.9993	0.9998	1.0000	0.0000	0.0000	0.0007	1.0000
0.4	0.9769	1.0000	0.9988	0.0012	0.0000	0.0231	1.0000
0.5	0.8052	1.0000	0.9897	0.0103	0.0000	0.1948	1.0000
0.6	0.4122	0.9945	0.9691	0.0309	0.0000	0.5878	1.0000
0.7	0.1096	0.6860	0.9531	0.0469	0.0000	0.8904	1.0000
0.8	0.0139	0.1290	0.9481	0.0519	0.0000	0.9862	1.0000
0.9	0.0010	0.0105	0.9474	0.0526	0.0000	0.9990	1.0000
1	0.0001	0.0010	0.9474	0.0526	0.0000	0.9999	1.0000
1.1	0.0000	NaN	0.9474	0.0526	0.0000	1.0000	1.0000

(A) Pairwise (p=20)

λ	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
5e-04	0.9996	0.0664	0.2591	0.7409	0.7820	0.0004	0.2180
0.005	0.9996	0.1257	0.6327	0.3673	0.3877	0.0004	0.6123
0.05	0.8417	0.2195	0.8314	0.1686	0.1692	0.1583	0.8308
0.1	0.5879	0.2829	0.8951	0.1049	0.0879	0.4121	0.9121
0.2	0.2624	0.3140	0.9273	0.0727	0.0358	0.7377	0.9642
0.3	0.1222	0.2423	0.9318	0.0682	0.0232	0.8778	0.9768
0.4	0.0658	0.1830	0.9348	0.0652	0.0170	0.9342	0.9830
0.5	0.0390	0.1399	0.9370	0.0630	0.0131	0.9610	0.9869
0.6	0.0243	0.1106	0.9386	0.0614	0.0106	0.9757	0.9894
0.7	0.0163	0.0898	0.9400	0.0600	0.0086	0.9837	0.9914
0.8	0.0118	0.0753	0.9412	0.0588	0.0072	0.9882	0.9928
0.9	0.0079	0.0592	0.9428	0.0572	0.0052	0.9921	0.9948
1	0.0040	0.0349	0.9447	0.0553	0.0030	0.9960	0.9970
1.1	0.0017	0.0141	0.9463	0.0537	0.0012	0.9984	0.9988

(B) Hub (p=20)

λ	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
5e-04	1.0000	0.0761	0.3599	0.6401	0.6757	0.0000	0.3243
0.005	1.0000	0.1801	0.7593	0.2407	0.2540	0.0000	0.7460
0.05	1.0000	0.3321	0.8923	0.1077	0.1136	0.0000	0.8864
0.1	0.9781	0.5359	0.9524	0.0476	0.0490	0.0219	0.9510
0.2	0.7691	0.8461	0.9799	0.0201	0.0084	0.2309	0.9916
0.3	0.5576	0.8928	0.9727	0.0273	0.0043	0.4424	0.9957
0.4	0.4163	0.8992	0.9663	0.0337	0.0032	0.5837	0.9968
0.5	0.3173	0.9032	0.9618	0.0382	0.0024	0.6826	0.9976
0.6	0.2511	0.9024	0.9587	0.0413	0.0020	0.7488	0.9980
0.7	0.2017	0.8979	0.9565	0.0435	0.0016	0.7983	0.9984
0.8	0.1639	0.8687	0.9548	0.0452	0.0012	0.8361	0.9988
0.9	0.1182	0.7377	0.9528	0.0472	0.0009	0.8818	0.9991
1	0.0656	0.4430	0.9503	0.0497	0.0005	0.9344	0.9995
1.1	0.0250	0.1853	0.9485	0.0515	0.0002	0.9751	0.9998

(C) Cascade (p=20)

λ	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
5e-04	1.0000	0.0103	0.0289	0.9711	0.9811	0.0000	0.0189
0.005	1.0000	0.0122	0.1826	0.8174	0.8258	0.0000	0.1742
0.05	1.0000	0.0467	0.7936	0.2064	0.2085	0.0000	0.7915
0.1	1.0000	0.1640	0.9483	0.0517	0.0522	0.0000	0.9478
0.2	1.0000	0.8914	0.9987	0.0013	0.0013	0.0000	0.9987
0.3	0.9998	0.9990	1.0000	0.0000	0.0000	0.0002	1.0000
0.4	0.9804	1.0000	0.9998	0.0002	0.0000	0.0196	1.0000
0.5	0.7976	1.0000	0.9980	0.0020	0.0000	0.2024	1.0000
0.6	0.4072	1.0000	0.9940	0.0060	0.0000	0.5928	1.0000
0.7	0.1092	1.0000	0.9910	0.0090	0.0000	0.8908	1.0000
0.8	0.0138	0.4500	0.9900	0.0100	0.0000	0.9862	1.0000
0.9	0.0016	0.0800	0.9899	0.0101	0.0000	0.9984	1.0000
1	0.0002	0.0100	0.9899	0.0101	0.0000	0.9998	1.0000
1.1	0.0000	NaN	0.9899	0.0101	0.0000	1.0000	1.0000

(D) Pairwise (p=100)

λ	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
5e-04	1.0000	0.2270	0.9373	0.0627	0.0639	0.0000	0.9361
0.005	0.9997	0.3133	0.9597	0.0403	0.0411	0.0003	0.9589
0.05	0.8377	0.4212	0.9757	0.0243	0.0217	0.1623	0.9783
0.1	0.5910	0.4190	0.9772	0.0228	0.0156	0.4090	0.9844
0.2	0.2664	0.3132	0.9756	0.0244	0.0111	0.7336	0.9889
0.3	0.1234	0.2241	0.9759	0.0241	0.0081	0.8766	0.9919
0.4	0.0626	0.1628	0.9769	0.0231	0.0060	0.9374	0.9940
0.5	0.0352	0.1243	0.9777	0.0223	0.0046	0.9648	0.9954
0.6	0.0226	0.1014	0.9784	0.0216	0.0037	0.9774	0.9963
0.7	0.0145	0.0822	0.9789	0.0211	0.0030	0.9855	0.9970
0.8	0.0090	0.0648	0.9794	0.0206	0.0024	0.9910	0.9976
0.9	0.0064	0.0640	0.9800	0.0200	0.0018	0.9936	0.9982
1	0.0030	0.0494	0.9807	0.0193	0.0010	0.9970	0.9990
1.1	0.0008	0.0239	0.9813	0.0187	0.0004	0.9992	0.9996

(E) Hub (p=100)

λ	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
5e-04	1.0000	0.4140	0.9739	0.0261	0.0266	0.0000	0.9734
0.005	1.0000	0.6517	0.9901	0.0099	0.0101	0.0000	0.9899
0.05	0.9999	0.8400	0.9965	0.0035	0.0036	0.0001	0.9964
0.1	0.9766	0.8821	0.9972	0.0028	0.0025	0.0234	0.9975
0.2	0.7652	0.8854	0.9938	0.0062	0.0019	0.2348	0.9981
0.3	0.5625	0.8789	0.9905	0.0095	0.0015	0.4375	0.9985
0.4	0.4208	0.8785	0.9883	0.0117	0.0011	0.5792	0.9989
0.5	0.3246	0.8826	0.9868	0.0132	0.0008	0.6754	0.9992
0.6	0.2565	0.8820	0.9857	0.0143	0.0007	0.7435	0.9993
0.7	0.2046	0.8789	0.9849	0.0151	0.0005	0.7954	0.9995
0.8	0.1642	0.8761	0.9842	0.0158	0.0004	0.8358	0.9996
0.9	0.1213	0.8679	0.9835	0.0165	0.0003	0.8787	0.9997
1	0.0705	0.8888	0.9827	0.0173	0.0002	0.9295	0.9998
1.1	0.0257	0.8480	0.9820	0.0180	0.0001	0.9743	0.9999

(F) Cascade (p=100)

λ	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
5e-04	0.9995	0.0082	0.3902	0.6098	0.6129	0.0005	0.3871
0.005	1.0000	0.0097	0.4881	0.5119	0.5145	0.0000	0.4855
0.05	1.0000	0.0323	0.8496	0.1504	0.1512	0.0000	0.8488
0.1	1.0000	0.1045	0.9569	0.0431	0.0433	0.0000	0.9567
0.2	1.0000	0.8002	0.9987	0.0013	0.0013	0.0000	0.9987
0.3	0.9995	0.9975	1.0000	0.0000	0.0000	0.0005	1.0000
0.4	0.9761	0.9999	0.9999	0.0001	0.0000	0.0239	1.0000
0.5	0.7969	1.0000	0.9990	0.0010	0.0000	0.2031	1.0000
0.6	0.4052	1.0000	0.9970	0.0030	0.0000	0.5948	1.0000
0.7	0.1047	1.0000	0.9955	0.0045	0.0000	0.8953	1.0000
0.8	0.0125	0.7300	0.9950	0.0050	0.0000	0.9875	1.0000
0.9	0.0010	0.1000	0.9950	0.0050	0.0000	0.9990	1.0000
1	0.0000	NaN	0.9950	0.0050	0.0000	1.0000	1.0000
1.1	0.0000	NaN	0.9950	0.0050	0.0000	1.0000	1.0000

(G) Pairwise (p=200)

λ	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
5e-04	0.9996	0.1775	0.9576	0.0424	0.0428	0.0004	0.9572
0.005	0.9997	0.3064	0.9793	0.0207	0.0209	0.0003	0.9791
0.05	0.8367	0.4184	0.9878	0.0122	0.0108	0.1633	0.9892
0.1	0.5796	0.4079	0.9884	0.0116	0.0078	0.4204	0.9922
0.2	0.2591	0.3085	0.9879	0.0121	0.0054	0.7409	0.9946
0.3	0.1261	0.2316	0.9882	0.0118	0.0039	0.8739	0.9961
0.4	0.0675	0.1780	0.9886	0.0114	0.0029	0.9325	0.9971
0.5	0.0409	0.1455	0.9891	0.0109	0.0022	0.9591	0.9978
0.6	0.0262	0.1204	0.9894	0.0106	0.0017	0.9738	0.9983
0.7	0.0185	0.1059	0.9896	0.0104	0.0014	0.9815	0.9986
0.8	0.0128	0.0911	0.9898	0.0102	0.0012	0.9872	0.9988
0.9	0.0078	0.0747	0.9901	0.0099	0.0009	0.9922	0.9991
1	0.0042	0.0717	0.9904	0.0096	0.0005	0.9958	0.9995
1.1	0.0015	0.0853	0.9907	0.0093	0.0002	0.9985	0.9998

(H) Hub (p=200)

λ	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
5e-04	1.0000	0.3071	0.9793	0.0207	0.0208	0.0000	0.9792
0.005	1.0000	0.6404	0.9948	0.0052	0.0052	0.0000	0.9948
0.05	0.9997	0.8364	0.9982	0.0018	0.0018	0.0003	0.9982
0.1	0.9758	0.8821	0.9986	0.0014	0.0012	0.0242	0.9988
0.2	0.7696	0.8874	0.9970	0.0030	0.0009	0.2304	0.9991
0.3	0.5589	0.8855	0.9953	0.0047	0.0007	0.4411	0.9993
0.4	0.4132	0.8832	0.9941	0.0059	0.0005	0.5868	0.9995
0.5	0.3182	0.8818	0.9934	0.0066	0.0004	0.6818	0.9996
0.6	0.2499	0.8786	0.9928	0.0072	0.0003	0.7501	0.9997
0.7	0.1998	0.8771	0.9924	0.0076	0.0003	0.8002	0.9997
0.8	0.1592	0.8733	0.9921	0.0079	0.0002	0.8408	0.9998
0.9	0.1142	0.8657	0.9917	0.0083	0.0002	0.8858	0.9998
1	0.0650	0.8705	0.9914	0.0086	0.0001	0.9350	0.9999
1.1	0.0243	0.8694	0.9910	0.0090	0.0000	0.9757	1.0000

(I) Cascade (p=200)

Table 4 Simulation results for G-Lasso algorithm (n=150 throughout).

λ	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
0.05	1.0000	0.1146	0.5910	0.4090	0.4317	0.0000	0.5683
0.1	1.0000	0.2411	0.8307	0.1693	0.1787	0.0000	0.8213
0.2	1.0000	0.8146	0.9868	0.0132	0.0139	0.0000	0.9861
0.3	0.9993	0.9951	0.9997	0.0003	0.0003	0.0007	0.9997
0.4	0.9769	1.0000	0.9988	0.0012	0.0000	0.0231	1.0000
0.5	0.8052	1.0000	0.9897	0.0103	0.0000	0.1948	1.0000
0.6	0.4122	0.9945	0.9691	0.0309	0.0000	0.5878	1.0000
0.7	0.1096	0.6860	0.9531	0.0469	0.0000	0.8904	1.0000
0.8	0.0139	0.1290	0.9481	0.0519	0.0000	0.9862	1.0000
0.9	0.0010	0.0105	0.9474	0.0526	0.0000	0.9990	1.0000
1	0.0001	0.0010	0.9474	0.0526	0.0000	0.9999	1.0000
1.1	0.0000	NaN	0.9474	0.0526	0.0000	1.0000	1.0000
1.2	0.0000	NaN	0.9474	0.0526	0.0000	1.0000	1.0000
1.3	0.0000	NaN	0.9474	0.0526	0.0000	1.0000	1.0000

(A) Pairwise (p=20)

λ	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
0.05	1.0000	0.1130	0.5844	0.4156	0.4387	0.0000	0.5613
0.1	1.0000	0.1422	0.6797	0.3203	0.3381	0.0000	0.6619
0.2	1.0000	0.1781	0.7568	0.2432	0.2567	0.0000	0.7433
0.3	1.0000	0.1817	0.7630	0.2370	0.2502	0.0000	0.7498
0.4	1.0000	0.1818	0.7632	0.2368	0.2500	0.0000	0.7500
0.5	1.0000	0.1818	0.7632	0.2368	0.2500	0.0000	0.7500
0.6	1.0000	0.1818	0.7632	0.2368	0.2500	0.0000	0.7500
0.7	0.9995	0.1817	0.7633	0.2367	0.2499	0.0005	0.7501
0.8	0.9650	0.1765	0.7699	0.2301	0.2410	0.0349	0.7590
0.9	0.7887	0.1478	0.8032	0.1969	0.1960	0.2114	0.8040
1	0.4734	0.0909	0.8611	0.1389	0.1174	0.5266	0.8826
1.1	0.1808	0.0373	0.9148	0.0852	0.0444	0.8192	0.9556
1.2	0.0530	0.0113	0.9379	0.0621	0.0129	0.9470	0.9871
1.3	0.0094	0.0020	0.9457	0.0543	0.0023	0.9905	0.9977

(B) Hub (p=20)

λ	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
0.05	1.0000	0.1134	0.5864	0.4136	0.4366	0.0000	0.5634
0.1	1.0000	0.1429	0.6820	0.3180	0.3356	0.0000	0.6644
0.2	1.0000	0.1782	0.7570	0.2430	0.2565	0.0000	0.7435
0.3	1.0000	0.1817	0.7630	0.2370	0.2502	0.0000	0.7498
0.4	1.0000	0.1818	0.7632	0.2368	0.2500	0.0000	0.7500
0.5	1.0000	0.1818	0.7632	0.2368	0.2500	0.0000	0.7500
0.6	0.9994	0.1820	0.7633	0.2367	0.2498	0.0006	0.7502
0.7	0.9972	0.1818	0.7639	0.2361	0.2491	0.0028	0.7509
0.8	0.9582	0.1797	0.7717	0.2283	0.2387	0.0418	0.7613
0.9	0.7965	0.1660	0.8033	0.1967	0.1964	0.2036	0.8036
1	0.4675	0.1072	0.8646	0.1354	0.1134	0.5325	0.8866
1.1	0.1815	0.0535	0.9165	0.0835	0.0427	0.8184	0.9573
1.2	0.0512	0.0172	0.9386	0.0614	0.0121	0.9488	0.9879
1.3	0.0092	0.0036	0.9458	0.0543	0.0022	0.9908	0.9978

(C) Cascade (p=20)

λ	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
0.05	1.0000	0.0246	0.5994	0.4006	0.4047	0.0000	0.5953
0.1	1.0000	0.0562	0.8303	0.1697	0.1715	0.0000	0.8285
0.2	1.0000	0.4260	0.9862	0.0138	0.0139	0.0000	0.9861
0.3	0.9998	0.9742	0.9997	0.0003	0.0003	0.0002	0.9997
0.4	0.9804	0.9996	0.9998	0.0002	0.0000	0.0196	1.0000
0.5	0.7976	1.0000	0.9980	0.0020	0.0000	0.2024	1.0000
0.6	0.4072	1.0000	0.9940	0.0060	0.0000	0.5928	1.0000
0.7	0.1092	1.0000	0.9910	0.0090	0.0000	0.8908	1.0000
0.8	0.0138	0.4500	0.9900	0.0100	0.0000	0.9862	1.0000
0.9	0.0016	0.0800	0.9899	0.0101	0.0000	0.9984	1.0000
1	0.0002	0.0100	0.9899	0.0101	0.0000	0.9998	1.0000
1.1	0.0000	NaN	0.9899	0.0101	0.0000	1.0000	1.0000
1.2	0.0000	NaN	0.9899	0.0101	0.0000	1.0000	1.0000
1.3	0.0000	NaN	0.9899	0.0101	0.0000	1.0000	1.0000

(D) Pairwise (p=100)

λ	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
0.05	1.0000	0.1439	0.8904	0.1096	0.1116	0.0000	0.8884
0.1	1.0000	0.1527	0.8975	0.1025	0.1044	0.0000	0.8956
0.2	1.0000	0.1770	0.9144	0.0856	0.0872	0.0000	0.9128
0.3	1.0000	0.1798	0.9162	0.0838	0.0854	0.0000	0.9146
0.4	1.0000	0.1798	0.9162	0.0838	0.0854	0.0000	0.9146
0.5	1.0000	0.1798	0.9162	0.0838	0.0854	0.0000	0.9146
0.6	1.0000	0.1798	0.9162	0.0838	0.0854	0.0000	0.9146
0.7	1.0000	0.1799	0.9162	0.0838	0.0854	0.0000	0.9146
0.8	0.9605	0.1800	0.9188	0.0812	0.0820	0.0395	0.9180
0.9	0.7819	0.1805	0.9307	0.0693	0.0665	0.2181	0.9335
1	0.4688	0.1801	0.9515	0.0485	0.0395	0.5312	0.9605
1.1	0.1666	0.1426	0.9711	0.0289	0.0138	0.8334	0.9862
1.2	0.0412	0.0583	0.9789	0.0211	0.0036	0.9588	0.9964
1.3	0.0080	0.0115	0.9811	0.0189	0.0007	0.9920	0.9993

(E) Hub (p=100)

λ	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
0.05	1.0000	0.1426	0.8893	0.1107	0.1127	0.0000	0.8873
0.1	1.0000	0.1525	0.8975	0.1025	0.1044	0.0000	0.8956
0.2	1.0000	0.1765	0.9141	0.0859	0.0875	0.0000	0.9125
0.3	1.0000	0.1798	0.9162	0.0838	0.0854	0.0000	0.9146
0.4	1.0000	0.1798	0.9162	0.0838	0.0854	0.0000	0.9146
0.5	1.0000	0.1798	0.9162	0.0838	0.0854	0.0000	0.9146
0.6	1.0000	0.1798	0.9162	0.0838	0.0854	0.0000	0.9146
0.7	0.9967	0.1799	0.9164	0.0836	0.0851	0.0033	0.9149
0.8	0.9724	0.1802	0.9181	0.0819	0.0829	0.0276	0.9171
0.9	0.8268	0.1822	0.9285	0.0715	0.0695	0.1732	0.9305
1	0.4882	0.1875	0.9513	0.0487	0.0400	0.5118	0.9600
1.1	0.1827	0.2016	0.9704	0.0296	0.0148	0.8173	0.9852
1.2	0.0384	0.1102	0.9794	0.0206	0.0029	0.9616	0.9971
1.3	0.0047	0.0391	0.9814	0.0186	0.0003	0.9953	0.9997

(F) Cascade (p=100)

λ	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
0.05	1.0000	0.0133	0.6276	0.3724	0.3743	0.0000	0.6257
0.1	1.0000	0.0299	0.8369	0.1631	0.1639	0.0000	0.8361
0.2	1.0000	0.2682	0.9862	0.0138	0.0138	0.0000	0.9862
0.3	0.9995	0.9427	0.9997	0.0003	0.0003	0.0005	0.9997
0.4	0.9761	0.9997	0.9999	0.0001	0.0000	0.0239	1.0000
0.5	0.7969	1.0000	0.9990	0.0010	0.0000	0.2031	1.0000
0.6	0.4052	1.0000	0.9970	0.0030	0.0000	0.5948	1.0000
0.7	0.1047	1.0000	0.9955	0.0045	0.0000	0.8953	1.0000
0.8	0.0125	0.7300	0.9950	0.0050	0.0000	0.9875	1.0000
0.9	0.0010	0.1000	0.9950	0.0050	0.0000	0.9990	1.0000
1	0.0000	NaN	0.9950	0.0050	0.0000	1.0000	1.0000
1.1	0.0000	NaN	0.9950	0.0050	0.0000	1.0000	1.0000
1.2	0.0000	NaN	0.9950	0.0050	0.0000	1.0000	1.0000
1.3	0.0000	NaN	0.9950	0.0050	0.0000	1.0000	1.0000

(G) Pairwise (p=200)

λ	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
0.05	1.0000	0.1158	0.9300	0.0700	0.0706	0.0000	0.9294
0.1	1.0000	0.1267	0.9367	0.0633	0.0639	0.0000	0.9361
0.2	1.0000	0.1722	0.9560	0.0440	0.0444	0.0000	0.9556
0.3	1.0000	0.1798	0.9583	0.0417	0.0421	0.0000	0.9579
0.4	1.0000	0.1798	0.9583	0.0417	0.0421	0.0000	0.9579
0.5	1.0000	0.1798	0.9583	0.0417	0.0421	0.0000	0.9579
0.6	1.0000	0.1798	0.9583	0.0417	0.0421	0.0000	0.9579
0.7	0.9984	0.1798	0.9583	0.0417	0.0420	0.0016	0.9580
0.8	0.9735	0.1800	0.9592	0.0408	0.0409	0.0265	0.9591
0.9	0.8077	0.1808	0.9648	0.0352	0.0338	0.1923	0.9662
1	0.4910	0.1811	0.9750	0.0250	0.0205	0.5090	0.9795
1.1	0.1805	0.1816	0.9851	0.0149	0.0075	0.8195	0.9925
1.2	0.0427	0.1173	0.9895	0.0105	0.0018	0.9573	0.9982
1.3	0.0073	0.0241	0.9906	0.0094	0.0003	0.9927	0.9997

(H) Hub (p=200)

λ	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
0.05	1.0000	0.1157	0.9300	0.0700	0.0706	0.0000	0.9294
0.1	1.0000	0.1302	0.9387	0.0613	0.0618	0.0000	0.9382
0.2	1.0000	0.1723	0.9560	0.0440	0.0444	0.0000	0.9556
0.3	1.0000	0.1796	0.9582	0.0418	0.0422	0.0000	0.9578
0.4	1.0000	0.1798	0.9583	0.0417	0.0421	0.0000	0.9579
0.5	1.0000	0.1798	0.9583	0.0417	0.0421	0.0000	0.9579
0.6	1.0000	0.1798	0.9583	0.0417	0.0421	0.0000	0.9579
0.7	0.9976	0.1799	0.9584	0.0416	0.0420	0.0024	0.9580
0.8	0.9666	0.1803	0.9595	0.0405	0.0406	0.0334	0.9594
0.9	0.8000	0.1822	0.9653	0.0347	0.0332	0.2000	0.9668
1	0.4788	0.1857	0.9760	0.0240	0.0194	0.5212	0.9806
1.1	0.1785	0.2105	0.9855	0.0145	0.0071	0.8215	0.9929
1.2	0.0441	0.1597	0.9896	0.0104	0.0017	0.9559	0.9983
1.3	0.0040	0.0171	0.9907	0.0093	0.0002	0.9960	0.9998

(I) Cascade (p=200)

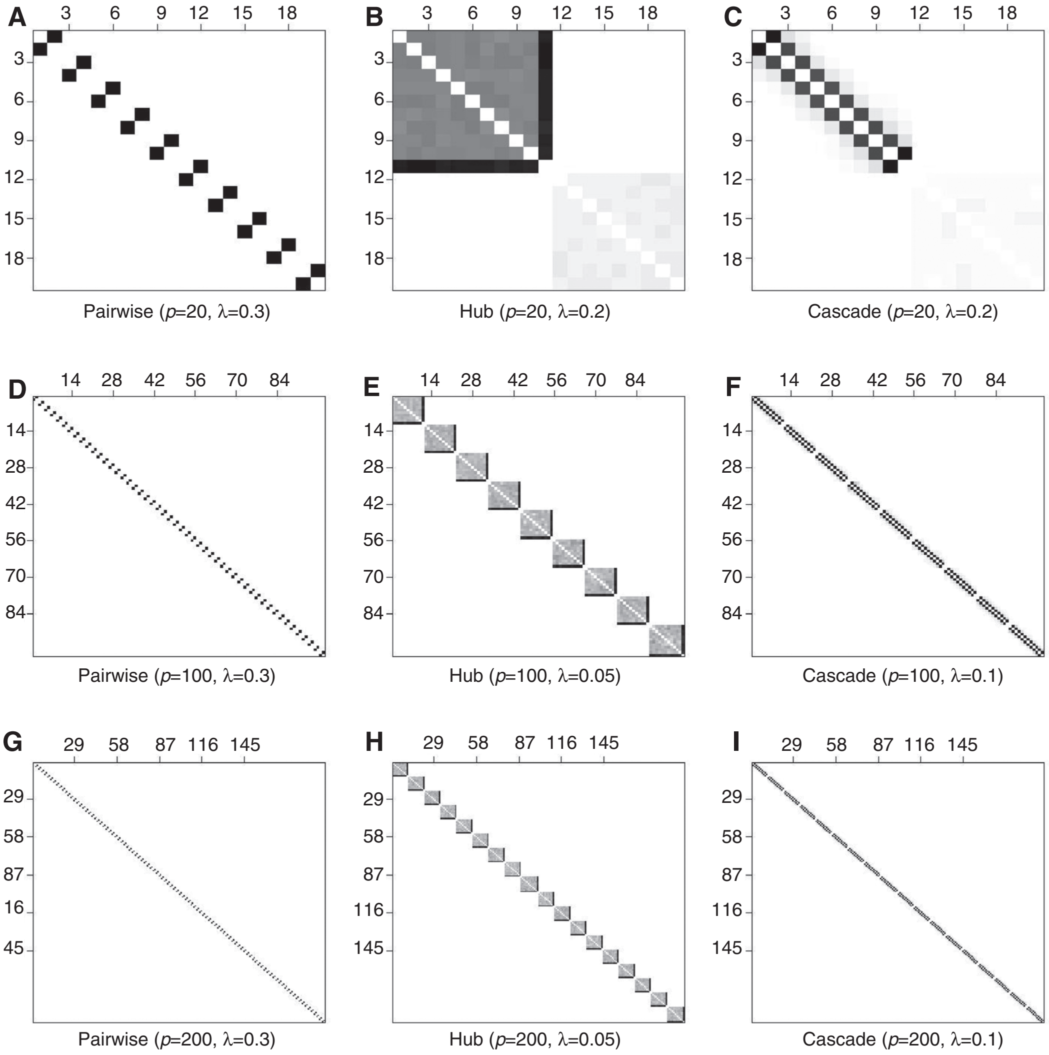

Figure 6 Plots of the averaged estimated adjacent matrices with Neighbourhood selection.

ditional independencies, i.e., zeros in Σ and Ω are the same, hence we expect that model selection methods should rarely get "confused" by indirect associations. Also, the pairwise structure is sparse in two ways: connectivity is very low and each node has only a single neighbour which should make it easy to learn.

For each motif, we construct three networks with an increasing number of genes (variables/vertices) $p=\{20, 100, 200\}$. The true adjacency matrices are shown in Figure 5.

A Gaussian sample of size $n=150$, with mean zero and covariance matrix according to the given network is then simulated several times (i.e., replications). More details on the simulation design are given in the supplementary material (Appendix C).

5.2 Results of the analysis with synthetic data

5.2.1 Neighbourhood selection and G-Lasso

Tables 3 and 4 as well as Figures 6 and 7 give the results for Neighbourhood selection and G-Lasso. Note that the λ values shown were chosen with knowledge of the true structure. The complete set of adjacency matrix plots is given in supplementary material (Appendix D and E).

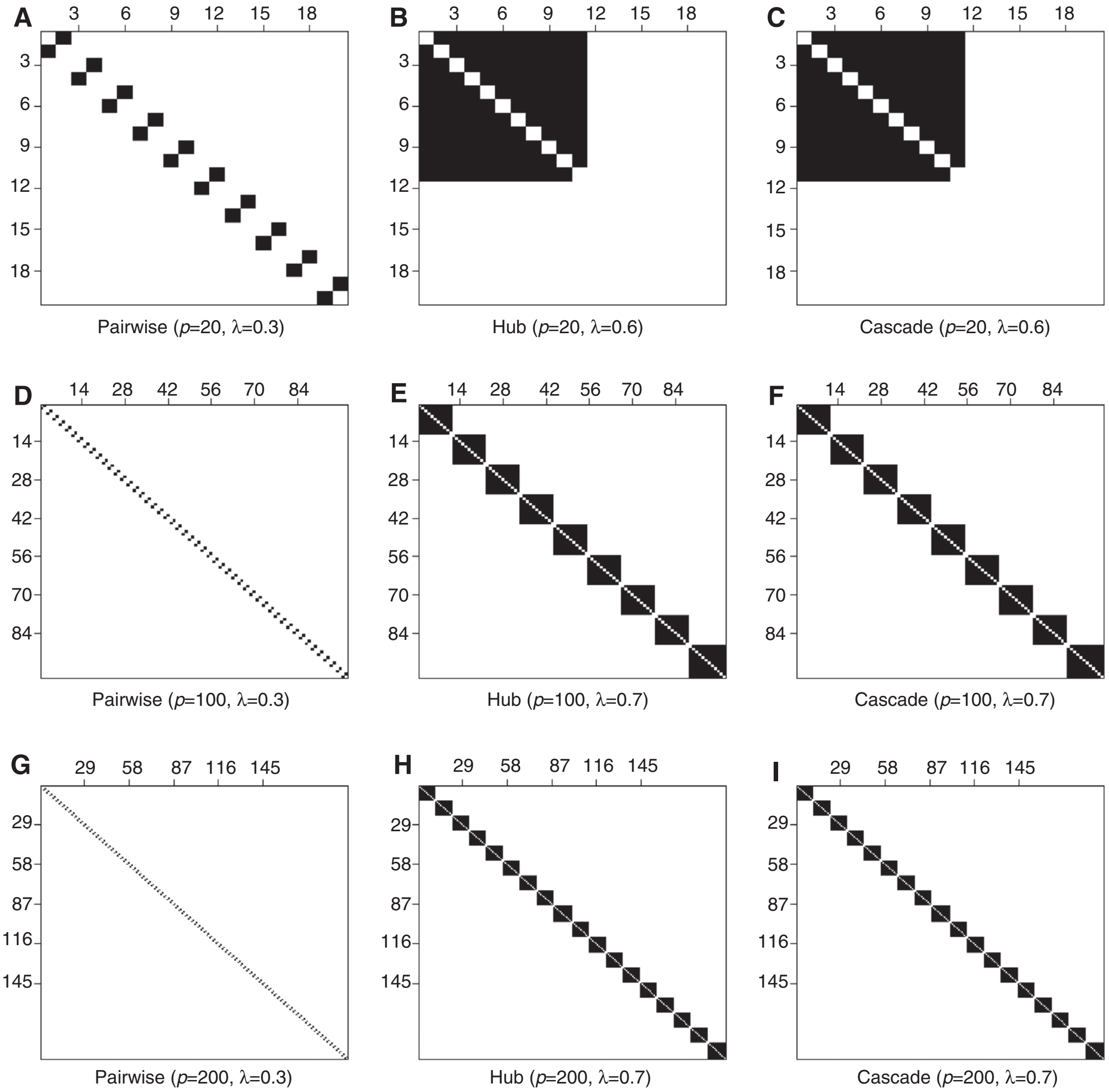

Figure 7 Plots of the averaged estimated adjacent matrices with G-Lasso.

In summary we find that Neighbourhood selection performs generally well for the pairwise motif structure, and still acceptable for the cascades where either the Tpr or the precision are rather low. It does not perform well for the hub structure. In some cases it finds a counter-cluster among those variables that should be mutually independent. In others there is a systematic pattern of false positives: it finds a clique among those variables that are attached to the hub and cannot distinguish the direct from the indirect associations.

G-Lasso performs similarly well for the pairwise structure as Neigbourhood selection, and just as badly for the hub structure (but with worse precision). For the cascade structure G-Lasso exhibits a clear systematic pattern of false positives.

5.2.2 Shrinkage estimator

For the Shrinkage estimator we find that, surprisingly, the *t*-test approach almost always selects only very few if any edges (Table 5B,D). A plausible explanation could be that the false discovery rate imposes too low a significance level compared to the estimated *p*-values.

Table 5 Simulation results for Shrinkage estimator (n=150 throughout).

	PAIRWISE	HUB	CASCADE
Tpr	0.9909	0.0467	0.4345
Precision	0.9873	0.0136	0.6593
Accuracy	0.9991	0.9264	0.9556
Error Rate	0.0009	0.0736	0.0444
Fpr	0.0005	0.0248	0.0154
Fnr	0.0091	0.9533	0.5655
Tnr	0.9995	0.9752	0.9846

(A) Empirical bayes approach (p=20)

	PAIRWISE	HUB	CASCADE
Tpr	0.9822	0.0000	0.0000
Precision	1.0000	0.0000	0.0000
Accuracy	0.9991	0.9473	0.9474
Error Rate	0.0009	0.0527	0.0526
Fpr	0.0000	0.0000	0.0000
Fnr	0.0179	1.0000	1.0000
Tnr	1.0000	0.9999	0.9999

(B) t-test approach (p=20)

	PAIRWISE	HUB	CASCADE
Tpr	0.9998	1.0000	1.0000
Precision	0.9994	0.1798	0.1962
Accuracy	0.9999	0.9162	0.9240
Error Rate	0.0000	0.0838	0.0759
Fpr	0.0000	0.0854	0.0774
Fnr	0.0002	0.0000	0.0000
Tnr	0.9999	0.9146	0.9226

(C) Empirical bayes approach (p=100)

	PAIRWISE	HUB	CASCADE
Tpr	0.0000	0.0000	0.0000
Precision	NaN	NaN	NaN
Accuracy	0.9899	0.9816	0.9816
Error Rate	0.0101	0.0184	0.0184
Fpr	0.0000	0.0000	0.0000
Fnr	1.0000	1.0000	1.0000
Tnr	1.0000	1.0000	1.0000

(D) t-test approach (p=100)

	PAIRWISE	HUB	CASCADE
Tpr	0.9996	1.0000	1.0000
Precision	0.9997	0.1795	0.1795
Accuracy	0.9999	0.9582	0.9582
Error Rate	0.0000	0.0418	0.0418
Fpr	0.0000	0.0422	0.0422
Fnr	0.0004	0.0000	0.0000
Tnr	0.9999	0.9578	0.9578

(E) Empirical bayes approach (p=200)

From Table 5 (A,C,E) and Figure 8 we see that, using empirical Bayes, the results for the pairwise structure are again almost ideal. For the cascade structure we find that the precision is low, and when p=100, 200, false positives appear more likely to be found for nodes that are "close" together on the cascade, but too many are found where there is only indirect association. For the hub structure there is still low precision and the same systematic pattern of false positives as for Neighbourhood selection and G-Lasso.

5.2.3 Maximum likelihood estimator

As can be seen from Table 6, both the empirical Bayes approach and the t-test approach yield almost identical results; we only present the adjacency matrices for the former in Figure 9. The results are almost perfect with difficulties only occurring when the sample size is small compared to the number of nodes, and with low Tpr for the hub structure. However, it is noticeable that the MLE approach does not have a systematic pattern of false positives like the previous methods.

5.2.4 PC-algorithm

Table 7 shows the performance measures for the PC-algorithm with nominal significance levels below 10%. Figure 10 shows the adjacency plots for selected α values (further plots see supplementary material, Appendix F). For all the structures, the algorithm seems to do quite well learning the true edges, but there is not a generally best choice of α. For the hub structure, as with the other methods, but to a lesser extent, there are systematic false positives, namely edges connecting the nodes that should only connect to the hub.

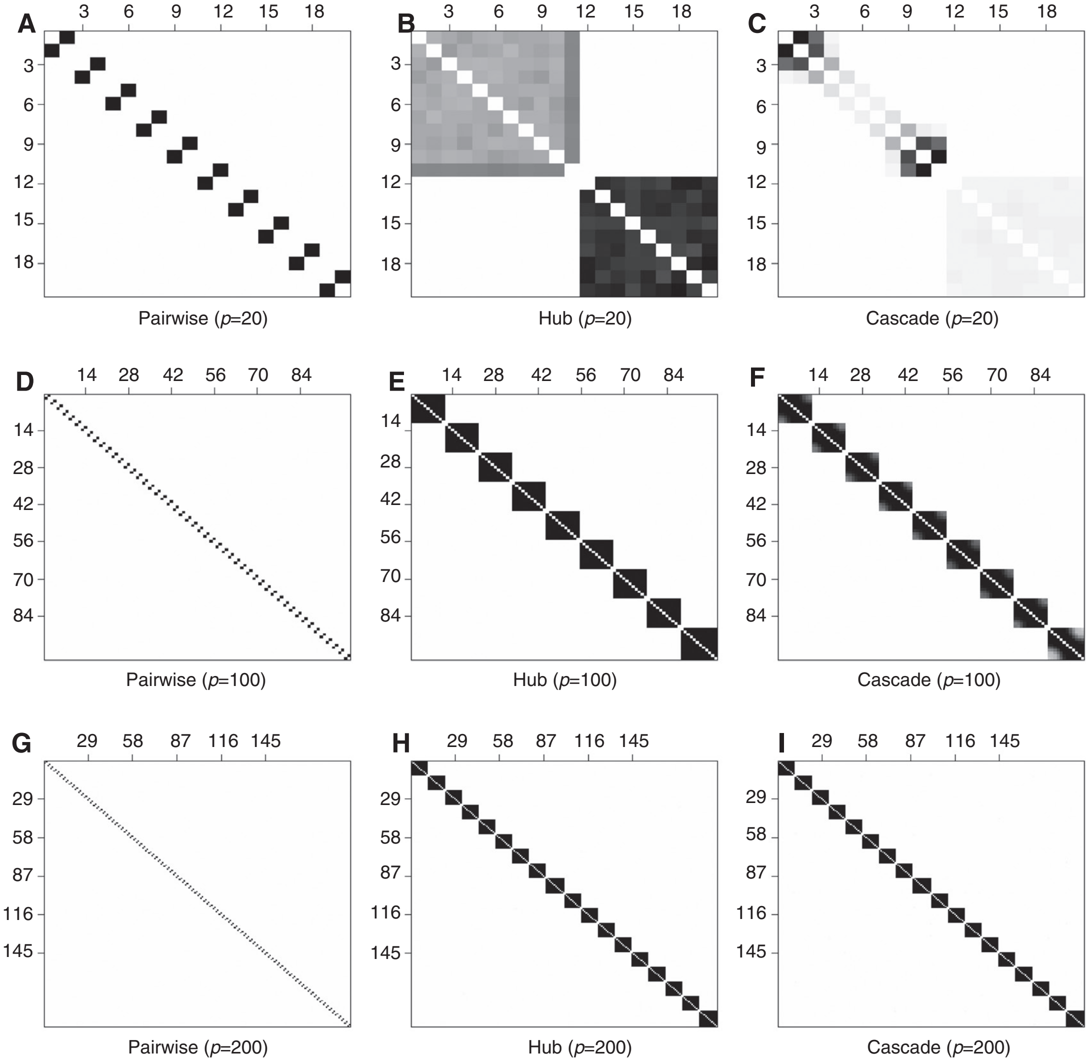

Figure 8 Plots of the averaged estimated adjacent matrices with Shrinkage and empirical Bayes approach.

Table 6 Simulation results for Maximum Likelihood (n=150 throughout).

	PAIRWISE	HUB	CASCADE
Tpr	0.9818	0.5039	0.9553
Precision	0.9859	0.9654	0.9882
Accuracy	0.9985	0.9734	0.9972
Error Rate	0.0015	0.0266	0.0028
Fpr	0.0006	0.0006	0.0005
Fnr	0.0183	0.4961	0.0447
Tnr	0.9994	0.9994	0.9995

(A) Empirical bayes approach (p=20)

	PAIRWISE	HUB	CASCADE
Tpr	0.9999	0.5277	0.9965
Precision	0.9957	0.9917	0.9959
Accuracy	0.9997	0.9749	0.9996
Error Rate	0.0003	0.0251	0.0004
Fpr	0.0003	0.0002	0.0002
Fnr	0.0001	0.4722	0.0035
Tnr	0.9997	0.9998	0.9997

(B) t-test approach (p=20)

	PAIRWISE	HUB	CASCADE
Tpr	0.5850	0.0133	0.3876
Precision	0.9990	0.5862	1.0000
Accuracy	0.9958	0.9819	0.9887
Error Rate	0.0042	0.0181	0.0113
Fpr	0.0000	0.0000	0.0000
Fnr	0.4150	0.9867	0.6124
Tnr	0.9999	0.9999	1.0000

(C) Empirical bayes approach (p=100)

	PAIRWISE	HUB	CASCADE
Tpr	0.5844	0.0119	0.3859
Precision	0.9986	0.6500	0.9987
Accuracy	0.9958	0.9818	0.9887
Error Rate	0.0042	0.0182	0.0113
Fpr	0.0000	0.0000	0.0000
Fnr	0.4156	0.9881	0.6141
Tnr	0.9999	0.9999	0.9999

(D) t-test approach (p=100)

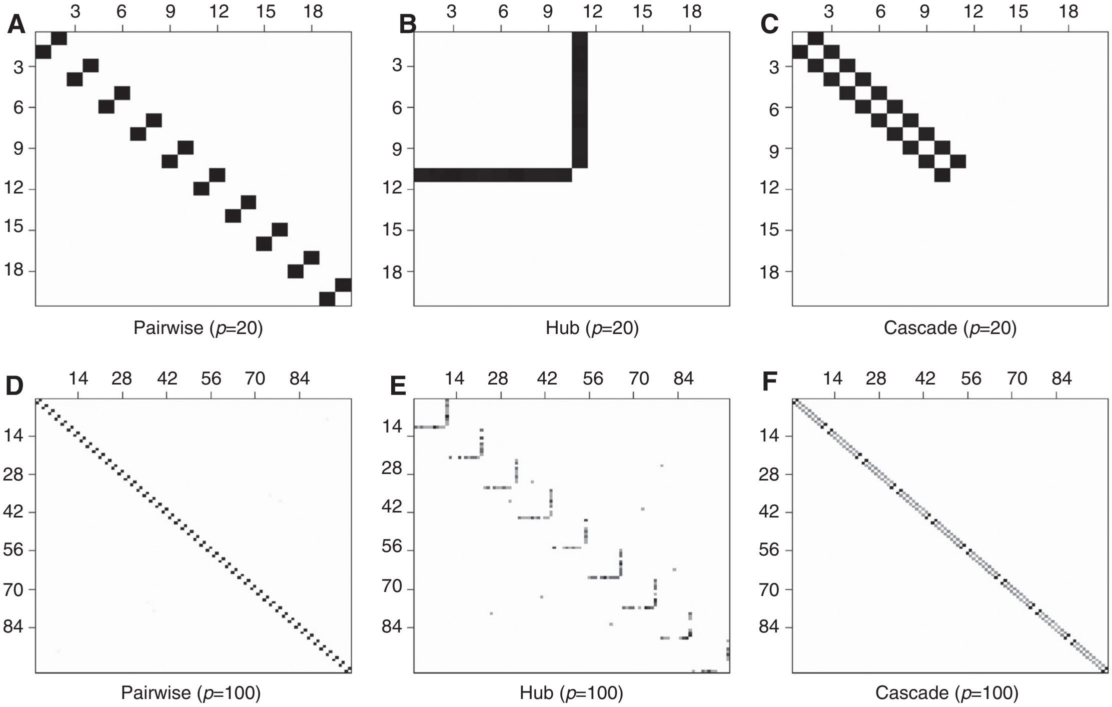

Figure 9 Plots of the averaged estimated adjacent matrices with MLE and empirical Bayes approach.

5.2.5 PR curves

It turns out that we can plot only partial PR curves for the PC-algorithm, and only for the hub motif, as this method does not yield a sufficiently wide range of recall values otherwise. Further, remember that the MLE cannot be used in the p=200 case. All PR-curves are shown in Figure 11.

Not surprisingly, the results for all methods are near perfect in case of the pairwise structure. For the hub structure, the MLE only struggles with the larger node set. Neighbourhood Selection, G-Lasso and Shrinkage exhibit qualitatively similar behaviour; the precision remains below 40% most of the time, in fact G-Lasso has a precision of 18% almost throughout, which is exactly the proportion of false positives in the corresponding networks that have additional edges between all nodes involved in the same hub. The PC-algorithm appears to outperform all other methods for the hub.

A clear ranking of the methods seems possible for the cascade structure: MLE (when possible) is best, then Neighbourhood selection, then Shrinkage, while G-Lasso is the weakest. The difference between Neighbourhood Selection and G-Lasso, two methods that are superficially quite similar, is especially striking.

6 Conclusion and discussion

We have compared a number of popular methods for learning genetic regulatory networks. In summary, the results suggest that the PC-algorithm seems to be the most promising approach if one desires to detect specific network motifs when $n<p$, but the choice of α deserves further investigation. In addition, if one considers the result before moralisation, one may even obtain different information about the network structure from the directions of the edges. Under certain assumptions, a (partially) directed network can inform about causal structures, see recent work of Maathuis et al. (2009) and Colombo et al. (2012).

Neighbourhood selection also appears to be an acceptable method, only somewhat disappointing with the hub structure. Meinshausen and Bühlmann (2006) give some advice on how to choose λ so as to limit

Table 7 Simulation results for PC-algorithm algorithm (n=150 throughout).

α	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
1e-08	0.9895	1.0000	0.9994	0.0006	0.0000	0.0105	1.0000
1e-07	0.9970	1.0000	0.9998	0.0002	0.0000	0.0030	1.0000
1e-06	0.9992	1.0000	1.0000	0.0000	0.0000	0.0008	1.0000
1e-05	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	1.0000
1e-04	1.0000	0.9999	1.0000	0.0000	0.0000	0.0000	1.0000
0.001	1.0000	0.9967	0.9998	0.0002	0.0002	0.0000	0.9998
0.01	1.0000	0.9554	0.9971	0.0029	0.0031	0.0000	0.9969
0.05	1.0000	0.7479	0.9800	0.0200	0.0212	0.0000	0.9789
0.1	1.0000	0.5525	0.9535	0.0465	0.0491	0.0000	0.9509

(A) Pairwise (p=20)

α	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
1e-08	0.4061	0.7690	0.9594	0.0406	0.0099	0.5939	0.9901
1e-07	0.4540	0.8392	0.9644	0.0356	0.0072	0.5460	0.9928
1e-06	0.5104	0.9086	0.9702	0.0298	0.0043	0.4896	0.9957
1e-05	0.5745	0.9607	0.9758	0.0242	0.0019	0.4255	0.9981
1e-04	0.6542	0.9880	0.9812	0.0188	0.0006	0.3458	0.9994
0.001	0.7531	0.9892	0.9865	0.0135	0.0005	0.2468	0.9995
0.01	0.8699	0.9191	0.9887	0.0113	0.0047	0.1301	0.9953
0.05	0.9522	0.6859	0.9715	0.0285	0.0274	0.0478	0.9726
0.1	0.9785	0.5045	0.9418	0.0582	0.0603	0.0215	0.9397

(B) Hub (p=20)

α	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
1e-08	1.0000	0.9899	0.9994	0.0006	0.0006	0.0000	0.9994
1e-07	1.0000	0.9969	0.9998	0.0002	0.0002	0.0000	0.9998
1e-06	1.0000	0.9991	0.9999	0.0001	0.0001	0.0000	0.9999
1e-05	1.0000	0.9999	1.0000	0.0000	0.0000	0.0000	1.0000
1e-04	1.0000	0.9997	1.0000	0.0000	0.0000	0.0000	1.0000
0.001	1.0000	0.9967	0.9998	0.0002	0.0002	0.0000	0.9998
0.01	1.0000	0.9641	0.9978	0.0022	0.0023	0.0000	0.9977
0.05	1.0000	0.8329	0.9882	0.0118	0.0125	0.0000	0.9875
0.1	1.0000	0.6854	0.9734	0.0266	0.0281	0.0000	0.9719

(C) Cascade (p=20)

α	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
1e-08	0.9912	1.0000	0.9999	0.0001	0.0000	0.0088	1.0000
1e-07	0.9982	1.0000	1.0000	0.0000	0.0000	0.0018	1.0000
1e-06	0.9998	1.0000	1.0000	0.0000	0.0000	0.0002	1.0000
1e-05	0.9998	0.9996	1.0000	0.0000	0.0000	0.0002	1.0000
1e-04	0.9998	0.9988	1.0000	0.0000	0.0000	0.0002	1.0000
0.001	1.0000	0.9839	0.9998	0.0002	0.0002	0.0000	0.9998
0.01	1.0000	0.8052	0.9975	0.0025	0.0026	0.0000	0.9974
0.05	1.0000	0.3929	0.9842	0.0158	0.0160	0.0000	0.9840
0.1	1.0000	0.2245	0.9649	0.0351	0.0354	0.0000	0.9646

(D) Pairwise (p=100)

α	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
1e-08	0.4027	0.7025	0.9857	0.0143	0.0034	0.5973	0.9966
1e-07	0.4508	0.7945	0.9876	0.0124	0.0023	0.5492	0.9977
1e-06	0.5058	0.8739	0.9895	0.0105	0.0014	0.4942	0.9986
1e-05	0.5703	0.9533	0.9916	0.0084	0.0005	0.4297	0.9995
1e-04	0.6525	0.9879	0.9935	0.0065	0.0002	0.3475	0.9998
0.001	0.7475	0.9926	0.9953	0.0047	0.0001	0.2525	0.9999
0.01	0.8660	0.9485	0.9967	0.0033	0.0009	0.1340	0.9991
0.05	0.9496	0.7221	0.9901	0.0099	0.0091	0.0504	0.9909
0.1	0.9786	0.4807	0.9762	0.0238	0.0238	0.0214	0.9762

(E) Hub (p=100)

α	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
1e-08	1.0000	0.9905	0.9998	0.0002	0.0002	0.0000	0.9998
1e-07	1.0000	0.9966	0.9999	0.0001	0.0001	0.0000	0.9999
1e-06	1.0000	0.9988	1.0000	0.0000	0.0000	0.0000	1.0000
1e-05	1.0000	0.9999	1.0000	0.0000	0.0000	0.0000	1.0000
1e-04	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	1.0000
0.001	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	1.0000
0.01	1.0000	0.9986	1.0000	0.0000	0.0000	0.0000	1.0000
0.05	1.0000	0.9728	0.9994	0.0006	0.0006	0.0000	0.9994
0.1	1.0000	0.9197	0.9982	0.0018	0.0018	0.0000	0.9982

(F) Cascade (p=100)

α	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
1e-08	0.9894	1.0000	0.9999	0.0001	0.0000	0.0106	1.0000
1e-07	0.9967	1.0000	1.0000	0.0000	0.0000	0.0033	1.0000
1e-06	0.9992	0.9998	1.0000	0.0000	0.0000	0.0008	1.0000
1e-05	0.9998	0.9996	1.0000	0.0000	0.0000	0.0002	1.0000
1e-04	1.0000	0.9975	1.0000	0.0000	0.0000	0.0000	1.0000
0.001	1.0000	0.9644	0.9998	0.0002	0.0002	0.0000	0.9998
0.01	1.0000	0.6763	0.9976	0.0024	0.0025	0.0000	0.9975
0.05	1.0000	0.2724	0.9865	0.0135	0.0136	0.0000	0.9864
0.1	1.0000	0.1485	0.9711	0.0289	0.0290	0.0000	0.9710

(G) Pairwise (p=200)

α	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
1e-08	0.4029	0.6888	0.9928	0.0072	0.0017	0.5971	0.9983
1e-07	0.4503	0.7742	0.9937	0.0063	0.0012	0.5497	0.9988
1e-06	0.5039	0.8649	0.9947	0.0053	0.0008	0.4961	0.9992
1e-05	0.5694	0.9505	0.9958	0.0042	0.0003	0.4306	0.9997
1e-04	0.6482	0.9858	0.9967	0.0033	0.0001	0.3518	0.9999
0.001	0.7459	0.9919	0.9976	0.0024	0.0001	0.2541	0.9999
0.01	0.8643	0.9456	0.9983	0.0017	0.0005	0.1357	0.9995
0.05	0.9523	0.5826	0.9911	0.0089	0.0085	0.0477	0.9915
0.1	0.9803	0.4014	0.9853	0.0147	0.0147	0.0197	0.9853

(H) Hub (p=200)

α	Tpr	Prec.	Acc.	Err.	Fpr	Fnr	Tnr
1e-08	1.0000	0.9891	0.9999	0.0001	0.0001	0.0000	0.9999
1e-07	1.0000	0.9962	1.0000	0.0000	0.0000	0.0000	1.0000
1e-06	1.0000	0.9989	1.0000	0.0000	0.0000	0.0000	1.0000
1e-05	1.0000	0.9997	1.0000	0.0000	0.0000	0.0000	1.0000
1e-04	1.0000	0.9999	1.0000	0.0000	0.0000	0.0000	1.0000
0.001	1.0000	0.9997	1.0000	0.0000	0.0000	0.0000	1.0000
0.01	1.0000	0.9986	1.0000	0.0000	0.0000	0.0000	1.0000
0.05	1.0000	0.9633	0.9996	0.0004	0.0004	0.0000	0.9996
0.1	1.0000	0.9148	0.9991	0.0009	0.0009	0.0000	0.9991

(I) Cascade (p=200)

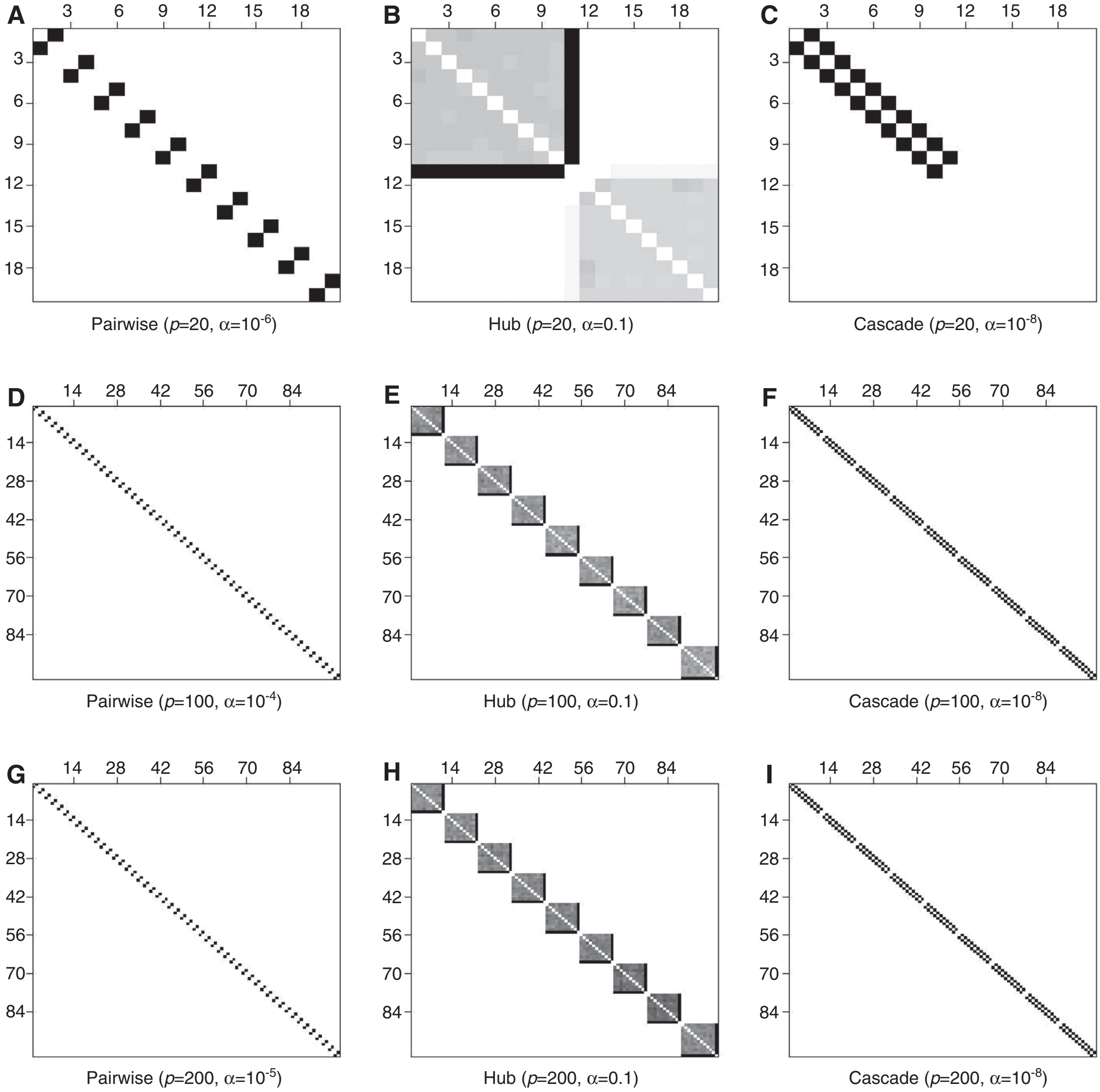

Figure 10 Plots of the averaged estimated adjacent matrices with the PC-algorithm.

the probability of falsely connecting two unconnected components of the graph. In a more recent paper Meinshausen and Bühlmann (2010) propose a different approach to this question based on the stability of selected edges over ranges of λ values.

Neither the Shrinkage nor the G-Lasso approach outperform the other methods and they exhibit systematic false positives in particular cases. For the latter, one could additionally consider using a set of penalty parameters, instead of a single one, to improve performance. For instance, if there is prior knowledge that certain nodes might be hubs then a smaller penalty could be chosen for the corresponding entries in the concentration matrix. The R package "SIMoNe" implements Neighbourhood selection, G-Lasso as well as further methods that allow for time-course data as well as for latent clustering, see Chiquet (2009) and references therein. Longitudinal data can also be analysed using a functional data approach implemented in "GeneNet", see Opgen-Rhein and Strimmer (2006).

None of the methods designed for the $n<p$ case are of course as good as a MLE approach when $n>p$. Uhler (2012) has results on the existence of the MLE for $n<p$ which may be useful in model search but we are not aware of any concrete proposals so far.

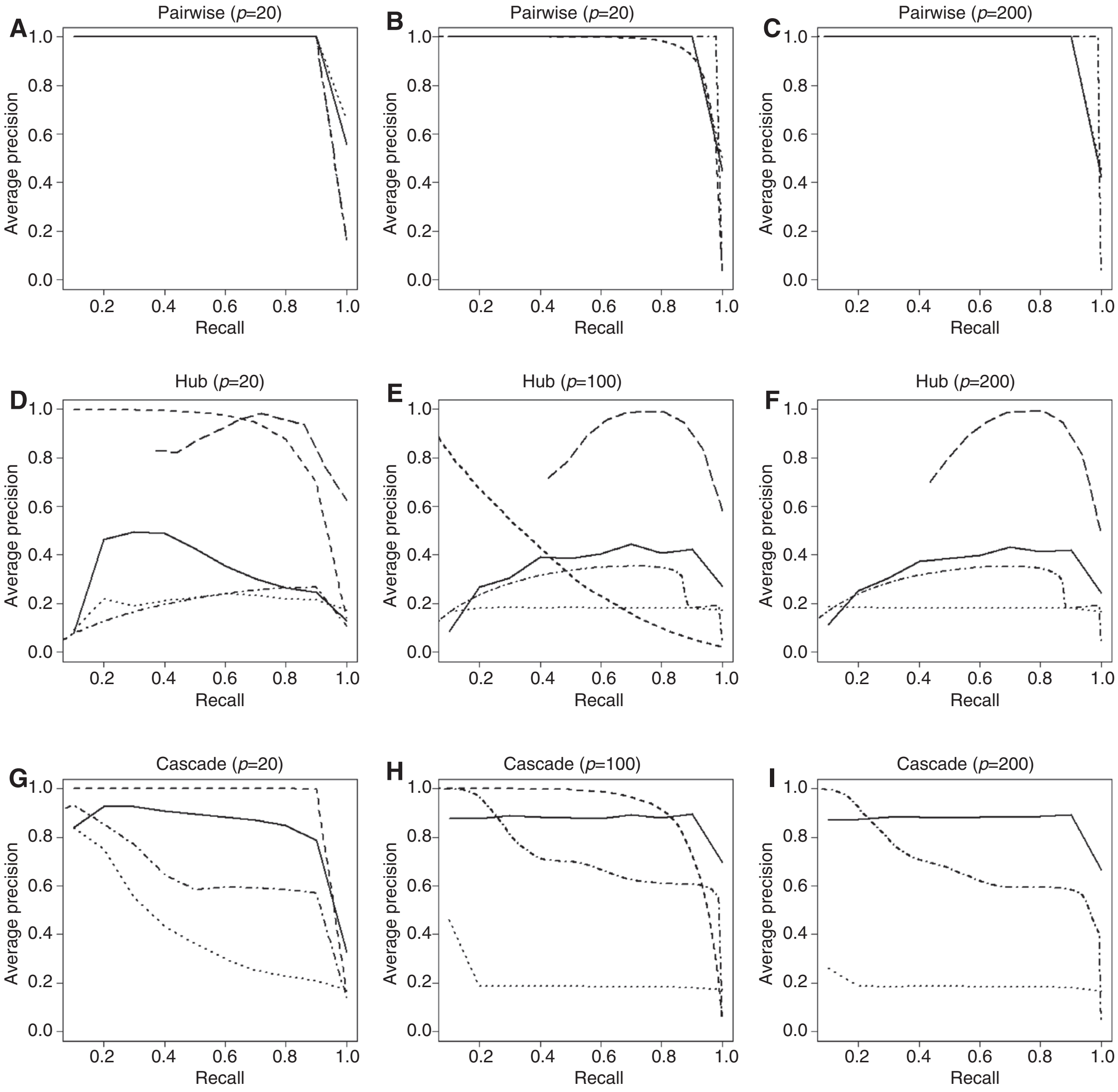

Figure 11 Precision-Recall curves for ML estimator (dashed), G-Lasso estimator (dotted), Neighbourhood selection (solid), Shrinkage estimator (dotdash), and PC-algorithm (only for hub structures, longdash).

We have not considered any Bayesian approach to structure learning as it would go beyond the scope of this paper. It may be interesting to consider whether prior information can be formulated that encourages hubs so as to further improve on the above methods.

Acknowledgments: We would like to thank Alberto Roverato, Sofia Massa, Arnoldo Frigessi, and Edmund Jones for helpful comments, and Markus Kalisch for help with the "pcAlgo" package. Financial support from the Leverhulme Trust (Research Fellowship RF-2011-320) is gratefully acknowledged.

References

Alon, U. (2007): "Network motifs: theory and experimental approaches," Nat. Rev. Genet., 8, 450–461.

Banerjee, O., L. E. Ghaoui and A. d'Aspremont (2008): "Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data," J. Mach. Learn. Res., 9, 485–516.

Barrett, T., T. O. Suzek, D. B. Troup, S. E. Wilhite, W. C. Ngau, P. Ledoux, D. Rudnev, A. E. Lash, W. Fujibuchi and R. Edgar (2007): "Ncbi geo: mining millions of expression profilesdatabase and tools," Nucl. Acids Res., 35, D562–D566.

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

Castelo, R. and A. Roverato (2009): "Reverse engineering molecular regulatory networks from microarray data with qp-graphs," J. Comput. Biol., 16, 2621–2650.

Chiquet, S. A. G. G. M. C. A. C., J. (2009): "SIMoNe: statistical inference for MOdular NEtworks," Bioinformatics, 25, 417–418.

Colombo, D., M. Maathuis, M. Kalisch and T. Richardson (2012): "Learning high-dimensional directed acyclic graphs with latent and selection variables," Ann. Stat., 40, 294–321.

Covert, M. W., E. M. Knight, J. L. Reed, M. J. Herrgard and B. O. Palsson (2004): "Integrating high-throughput and computational data elucidates bacterial networks," Nature, 429, 92–96.

Dawid, A. (1979): "Conditional independence in statistical theory," J. R. Stat. Soc. Ser. B (Methodol.), 41, 1–31.

Friedman, N. (2004): "Inferring cellular network using probabilistic graphical models," Science, 303, 799–805.

Friedman, J., T. Hastie and R. Tibshirani (2008): "Sparse inverse covariance estimation with the graphical lasso," Biostatistics, 9, 432–441.

Gama-Castro, S., V. Jimenez-Jacinto, M. Peralta-Gil, A. Santos-Zavaleta, M. I. Pealoza-Spinola, B. Contreras-Moreira, J. Segura-Salazar, L. Muniz-Rascado, I. Martnez-Flores, H. Salgado, C. Bonavides-Martnez, C. Abreu-Goodger, C. Rodrguez-Penagos, J. Miranda-Ros, E. Morett, E. Merino, A. M. Huerta, L. Trevino-Quintanilla and J. Collado-Vides (2008): "Regulondb (version 6.0): gene regulation model of escherichia coli k-12 beyond transcription, active (experimental) annotated promoters and textpresso navigation," Nucl. Acids Res., 36, D120–D124.

Hotelling, H. (1953): "New light on the correlation coefficient and its transforms," J. R. Statist. Soc. B, 15, 193–232.

Kalisch, M. and P. Bühlmann (2007): "Estimating high-dimensional directed acyclic graphs with the pc-algorithm," J. Mach. Learn. Res., 8, 613–636.

Lauritzen, S. (1996): Graphical models, Oxford: Oxford University Press.

Maathuis, M., M. Kalisch and P. Bühlmann (2009): "Estimating high-dimensional intervention effects from observational data," Ann. Stat., 37, 3133–3164.

Meinshausen, N. (2008): "A note on the lasso for graphical gaussian model selection," Statist. Probab. Lett., 78, 880–884.

Meinshausen, N. and P. Bühlmann (2006): "High-dimensional graphs and variable selection with the lasso," Ann. Statist., 34, 1436–1462.

Meinshausen, N. and P. Bühlmann (2010): "Stability selection," J. R. Stat. Soc. Ser. B, 72, 417–473.

Milo, R., S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii and U. Alon (2002): "Network motifs: simple building blocks of complex networks," Science, 298, 824–827.

Opgen-Rhein, R. and K. Strimmer (2006): "Inferring gene dependency networks from genomic longitudinal data: a functional data approach," REVSTAT, 4, 53–65.

Schäfer, J. and K. Strimmer (2005a): "An empirical bayes approach to inferring large-scale gene association networks," Bioinformatics, 21, 754–764.

Schäfer, J. and K. Strimmer (2005b): "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics," Stat. Appl. Genet. Mol. Biol., 4, 1–32.

Sing, T., O. Sander, N. Beerenwinkel and T. Lengauer (2005): "Rocr: visualizing classifier performance in r," Bioinformatics, 21, 3940–3941.

Spirtes, P. and C. Glymour (1991): "An algorithm for fast recovery of sparse causal graphs," Soc. Sci. Comput. Rev., 9, 62–72.

Tibshirani, R. (1996): "Regression shrinkage and selection via the lasso," J. R. Statist. Soc. B, 50, 267–288.

Uhler, C. (2012): "Geometry of maximum likelihood estimation in gaussian graphical models," The Annals of Statistics, 40, 238–261.

Supplemental Material: The online version of this article (DOI 10.1515/sagmb-2013-0017) offers supplementary material, available to authorized users.