BMC Bioinformatics

Research article

Regularized estimation of large-scale gene association networks using graphical Gaussian models

Nicole Krämer*[1], Juliane Schäfer[2,3] and Anne-Laure Boulesteix[4,5]

Addresses: [1]Machine Learning/Intelligent Data Analysis Group, Berlin Institute of Technology, Franklinstr 28/29, D-10587 Berlin, Germany, [2]Seminar für Statistik, ETH Zurich, CH-8092 Zurich, Switzerland, [3]Basel Institute for Clinical Epidemiology and Biostatistics, University Hospital Basel, CH-4031 Basel, Switzerland, [4]Department of Statistics, University of Munich, Ludwigstr 33, D-80539 Munich, Germany and [5]Computational Molecular Medicine Research Group, Department of Medical Informatics, Biometry and Epidemiology, University of Munich, Marchioninstr 15, 81377 Munich, Germany

E-mail: Nicole Krämer* - nkraemer@cs.tu-berlin.de; Juliane Schäfer - jschaefer@uhbs.ch; Anne-Laure Boulesteix - boulesteix@ibe.med.uni-muenchen.de
*Corresponding author

This article is available from: http://www.biomedcentral.com/1471-2105/10/384

Abstract

Background: Graphical Gaussian models are popular tools for the estimation of (undirected) gene association networks from microarray data. A key issue when the number of variables greatly exceeds the number of samples is the estimation of the matrix of partial correlations. Since the (Moore-Penrose) inverse of the sample covariance matrix leads to poor estimates in this scenario, standard methods are inappropriate and adequate regularization techniques are needed. Popular approaches include biased estimates of the covariance matrix and high-dimensional regression schemes, such as the Lasso and Partial Least Squares.

Results: In this article, we investigate a general framework for combining regularized regression methods with the estimation of Graphical Gaussian models. This framework includes various existing methods as well as two new approaches based on ridge regression and adaptive lasso, respectively. These methods are extensively compared both qualitatively and quantitatively within a simulation study and through an application to six diverse real data sets. In addition, all proposed algorithms are implemented in the R package "parcor", available from the R repository CRAN.

Conclusion: In our simulation studies, the investigated non-sparse regression methods, i.e. Ridge Regression and Partial Least Squares, exhibit rather conservative behavior when combined with (local) false discovery rate multiple testing in order to decide whether or not an edge is present in the network. For networks with higher densities, the difference in performance of the methods decreases. For sparse networks, we confirm the Lasso's well known tendency towards selecting too many edges, whereas the two-stage adaptive Lasso is an interesting alternative that provides sparser solutions. In our simulations, both sparse and non-sparse methods are able to reconstruct networks with cluster structures. On six real data sets, we also clearly distinguish the results obtained using the non-sparse methods and those obtained using the sparse methods where specification of the regularization parameter automatically means model selection. In five out of six data sets, Partial Least Squares selects very dense networks. Furthermore, for data that violate the assumption of uncorrelated observations (due to replications), the Lasso and the adaptive Lasso yield very complex structures, indicating that they might not be suited under these conditions. The shrinkage approach is more stable than the regression based approaches when using subsampling.

Background

Besides Bayesian networks [1], auto-regressive models [2], and state-space models [3], graphical Gaussian models (GGMs) are a popular method for modeling genetic networks based on microarray transcriptome data. In the GGM methodology [4], which is considered in the present article, networks are represented as undirected graphs. Each vertex represents a gene, and an edge connects two genes if they are partially correlated. In contrast to correlation, which measures both direct and indirect interactions between pairs of variables, partial correlation measures the strength of direct interaction only. Since investigators are primarily interested in direct gene interactions, the GGM framework is attractive for modeling of regulatory networks: several recent methodological articles report successful applications of GGMs to the estimation of genetic networks from microarray data [5-10]. These approaches are used in numerous applied studies, e.g., for estimating *Arabidopsis* gene networks [11] or for the study of genetically mediated cortical networks [12].

Nonetheless, reconstructing GGMs from high-dimensional microarray data remains a difficult task. The standard estimation of partial correlations involves either the inversion of the sample covariance matrix, or the estimation of p least squares regression problems, where p is the number of genes. If the number n of observations (arrays) is much smaller than the number p of variables (genes), these approaches are inappropriate. Suitable alternatives are based either on regularized estimation of the (inverse) covariance matrix, or on regularized high-dimensional regression. The present paper focuses on the latter approach, and presents a comparative study on the use of various approaches to high-dimensional regression for covariance selection. The chosen methods are extensively compared in simulations and real data studies. Since for real data the ground truth (i.e. the true underlying network) is unknown, our performance analysis focuses on the similarities and differences between the investigated methods. In particular, we examine the connectivity and size of the resulting graphs, as the differences between the estimated networks. Moreover, we compare the stability of the methods with respect to subsampling and with respect to violations of i.i.d. assumptions.

In the remainder of this section, we give a brief overview of graphical Gaussian modeling in the classical setting with $n > p$. Subsequently, we discuss the case of high-dimensional data in the "Methods" section.

Gene Regulatory Networks and Graphical Gaussian Models

Graphical Gaussian models (GGMs) [4] are fundamental tools in order to represent direct covariate interactions.

Formally, a GGM is an undirected graph whose nodes represent variables, and whose edges represent conditional dependency relations. An edge between two nodes is missing if and only if they are conditionally independent given all other nodes. Assuming a joint normal distribution, the conditional dependence can be quantified in terms of partial correlations. For a random variable X and a finite set of random variables $\mathsf{Z} = \{Z_1,...,Z_k\}$, the orthogonal complement of X with respect to Z is

$$X_{\setminus \mathcal{Z}} = X - \mathcal{P}_{\mathcal{Z}} X,$$

where the projection P_{Z} is defined with respect to the inner product $\langle X_1, X_2 \rangle = E[X_1 X_2]$ between two random variables X_1 and X_2. Here, we tacitly assume that all involved moments exist. The partial correlation $\rho_{\mathsf{Z}}(X_1, X_2)$ between X_1 and X_2 with respect to Z is the correlation of the orthogonal complements of X_1 and X_2 with respect to Z:

$$\rho_{\mathcal{Z}}(X_1, X_2) = \mathrm{cor}(X_{1 \setminus \mathcal{Z}}, X_{2 \setminus \mathcal{Z}}). \tag{1}$$

In the context of gene regulatory networks, each of the p genes is represented by a random variable X_i ($i = 1,..., p$). For each pair of genes (i, j), we are interested in their partial correlation ρ_{ij} with respect to all other genes, i.e. with respect to the set of random variables $\mathsf{Z}_{\setminus ij} = \{X_1,...,X_p\} \setminus \{X_i, X_j\}$.

Given n observations (arrays) $x_1,..., x_n \in \mathbb{R}^p$ of the set of p genes, the standard unbiased plug-in estimate for the partial correlation coefficients ρ_{ij} in the case $n > p$ can be formulated in two equivalent ways [4], as outlined below.

Notations
In the rest of this article,

$$X = (x_1,..., x_n)^\top \in \mathbb{R}^{n \times p} \tag{2}$$

denotes the $n \times p$ column-centered data matrix with rows corresponding to observations (arrays) and columns corresponding to variables (genes). The standard unbiased estimate of the $p \times p$ covariance matrix Σ is then given as

$$\hat{\Sigma} = \frac{1}{n-1} X^\top X.$$

Formulation 1: Inversion of the Covariance Matrix
If the estimate $\hat{\Sigma}$ is invertible, an unbiased estimate of the partial correlation between genes i and j is obtained as

$$\hat{\rho}_{ij} = -\frac{\hat{\omega}_{ij}}{\sqrt{\hat{\omega}_{ii} \hat{\omega}_{jj}}}. \tag{3}$$

with $\hat{\Omega}$ denoting the inverse of the estimated covariance matrix:

$$\hat{\Omega} = (\hat{\omega}_{ij}) = \hat{\Sigma}^{-1}.$$

Formulation 2: Least Squares Regression
Let us consider the p linear regression models

$$X_i = \sum_{j \neq i} \beta_j^{(i)} X_j + \varepsilon, \quad \text{for} \quad i = 1, ..., p, \tag{4}$$

where ε stands for i.i.d. noise. Note that we do not include an intercept in the model because the variables are centered. For $i = 1, ..., p$, the least squares estimate $\hat{\boldsymbol{\beta}}^{(i)} = (\beta_1^{(i)}, ..., \beta_{i-1}^{(i)}, \beta_{i+1}^{(i)}, ..., \beta_p^{(i)})^{\top}$ of the vector of regression coefficients is the solution of the optimization problem

$$\hat{\boldsymbol{\beta}}^{(i)} = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^{p-1}} \left\| X^{(i)} - X^{(\backslash i)} \boldsymbol{\beta} \right\|^2 \tag{5}$$

$$= \left(X^{(\backslash i)\top} X^{(\backslash i)} \right)^{-1} X^{(\backslash i)\top} X^{(i)}, \tag{6}$$

where $X^{(i)} \in \mathbb{R}^n$ is the ith column of X and $X^{(\backslash i)} \in \mathbb{R}^{n \times (p-1)}$ is the matrix obtained from X by deleting the ith column. The partial correlation between genes i and j is then estimated as

$$\hat{\rho}_{ij} = \text{sign}\left(\hat{\beta}_j^{(i)} \right) \sqrt{\hat{\beta}_j^{(i)} \hat{\beta}_i^{(j)}}. \tag{7}$$

In the $n > p$ setting, the two regression coefficients $\hat{\beta}_j^{(i)}$ and $\hat{\beta}_i^{(j)}$ always have the same sign. Hence, $\sqrt{\hat{\beta}_j^{(i)} \hat{\beta}_i^{(j)}}$ is well-defined. Moreover, it can be shown that both formulations 1 and 2 are equivalent [4] in the sense that they always yield the same estimate. In the $n \geq p$ setting, a test of the null hypothesis $\rho_{ij} = 0$ is available using results on the distribution of $\hat{\rho}_{ij}$.

In microarray data, the number n of samples is typically very small as compared to the number p of considered genes. Hence, the above framework is inappropriate for two reasons. Firstly, the standard estimate of the partial correlation matrix given by Eqs. (3) and (7) is not appropriate when $n < p$: in formulation 1, the estimated covariance matrix $\hat{\Sigma}$ is typically ill-conditioned or even singular, and its generalized (Moore-Penrose) inverse has large mean squared error [6]. In formulation 2, the least squares criterion (5) is ill-posed and leads to overfitting. Hence, an alternative regularized estimate of the partial correlation matrix has to be used in the context of GGMs with high-dimensional data. The two formulations 1 and 2 lead to two different strategies for

the regularized estimation of the partial correlations in the $p \gg n$ setting, which are reviewed in the Methods section.

Secondly, the testing approach mentioned above breaks down in the $p \gg n$ setting, since the sampling distribution of estimates $\hat{\rho}_{ij}$ under the null hypothesis of zero partial correlation is unknown. Two alternatives have been proposed in order to assess statistical significance: (i) methods based on sparse estimates of the partial correlation matrix that do not require separate testing, and (ii) methods based on empirical null modeling and (local) false discovery rate multiple testing [7,13,14].

Methods
This section reviews the available strategies for estimating GGMs in the $p \gg n$ setting: biased large-scale covariance estimation and regularized regression including our two novel variants (Ridge Regression and Adaptive Lasso).

Regularized Estimation of the (Inverse) Covariance Matrix
This approach is derived from formulation 1. The general approach is to plug a regularized estimate of the inverse of the sample covariance matrix into Eq. (3). Schäfer & Strimmer [6] adopt this approach and propose a shrinkage estimator of the covariance matrix. This shrinkage estimator is constructed as a convex combination of the unrestricted sample covariance matrix $\hat{\Sigma}$ and an estimator $\hat{T}$ of a specified low-dimensional sub-model T:

$$\hat{\Sigma}_{\lambda} = \lambda \hat{T} + (1 - \lambda)\hat{\Sigma},$$

where the factor $\lambda \in [0, 1]$ controls the shrinkage intensity. Let us assume a parametrization of covariances in terms of correlations and variances, whereas shrinkage is applied to the correlations and diagonal entries are left intact, i.e. the estimator does not shrink the variances. For correlation shrinkage, we consider the identity matrix as the most commonly employed shrinkage target. Notice that the optimal shrinkage intensity λ can be determined analytically and be estimated from the data. Thus, the resulting correlation shrinkage estimator is positive definite, and favorable properties carry over to derived quantities, such as sample partial correlations. Subsequently, model selection of the gene association network can be achieved using empirical null modeling and (local) false discovery rate multiple testing [7,13,14].

Estimates of the inverse covariance matrix can also be obtained using bootstrap aggregating (bagging) as a technique for variance reduction [15]. In some implicit

way, the bootstrap procedure presumably helps to regularize the problem. However, bagging schemes are inferior to the shrinkage estimator [6], and computationally much more expensive. A recent extension using the augmented bootstrap [16] is in fact closely related to the shrinkage estimator [17,18] and is expected to perform similarly.

In this paper, we use the correlation shrinkage based approach as a reference method in comparison with the regression based approaches to covariance selection.

Finally, recent novel approaches are to be noted that are based on, ℓ_1 regularized maximum likelihood estimation in graphical Gaussian models [9,19-21]. Corresponding inverse covariance estimates exploit the sparsity in the graphical structure and conduct parameter estimation and model selection simultaneously. However, despite recent advances in semidefinite programming computation remains challenging in practice due to the high-dimensionality and positive definiteness constraint [22].

Regularized Regression
Here, the strategy is to replace the least squares estimator in (6) by some regularized estimator of the regression coefficients that can be used in formula (7) to obtain estimators of the partial correlations. More formally, we define the following class of estimates of the partial correlations.

Definition 1. *For any regression method* reg*that yields (regularized) estimates* $\hat{\boldsymbol{\beta}}_{\text{reg}}^{(i)}$ *of the linear regression model (4), we define the corresponding estimate of the partial correlations as*

$$\hat{\rho}_{ij,\text{reg}} = \text{sign}\left(\hat{\beta}_{j,\text{reg}}^{(i)} \right) \min\left\{ 1, \sqrt{\hat{\beta}_{j,\text{reg}}^{(i)} \hat{\beta}_{i,\text{reg}}^{(j)}} \right\}$$

$$\text{if } \text{sign}\left(\hat{\beta}_{j,\text{reg}}^{(i)} \right) = \text{sign}\left(\hat{\beta}_{i,\text{reg}}^{(j)} \right)$$

and 0 *otherwise.*

This definition ensures that the estimated partial correlation coefficients are always well-defined and that they lie in the interval [-1, 1]. Again, we can roughly distinguish between regression methods that require testing to construct the undirected graphs, and sparse regression methods.

In the rest of this subsection, we discuss two regularized regression methods (PLS and the Lasso) that have been proposed for the estimation of large-scale GGMs in the literature. Furthermore, we propose two additional attractive methods (ridge regression and the adaptive Lasso).

Partial Least Squares
Tenenhaus et al. [23] suggest Partial Least Squares (PLS) regression [24,25] as a plug-in for Def. 1. PLS is a method for supervised dimensionality reduction. It has its seed in the chemometrics community, but its success has lead to applications in various other scientific fields, e.g. in chemo- and bioinformatics [26,27].

The main idea of PLS is to build a few orthogonal components from the original data $X^{(\backslash i)}$ and to use them as predictors in a least squares fit. A PLS component $t = X^{(\backslash i)} w$ is a linear combination of the original predictors that have maximal covariance with the response vector $X^{(i)}$, under the additional assumption that the components are mutually orthogonal. Formally, the k-th PLS component is defined by

$$w_k = \arg\max_{\|w\|=1} \text{cov}\left(X^{(\backslash i)}w, X^{(i)} \right)^2$$

$$\text{s.t.} \quad w^\top X^{(\backslash i)\top} X^{(\backslash i)} w_l = 0 \text{ for } l < k.$$

Hence, PLS regularizes the regression problem by compressing the p variables into a small number m of orthogonal components $T = (t_1,..., t_m)$ and regressing the response variable onto these components. After rescaling the weight vectors w_k ($k = 1,..., m$) such that t_k has length 1, this leads to the regression coefficients

$$\hat{\boldsymbol{\beta}}_{\text{pls}}^{(i)} = (w_1,..., w_m)T^\top X^{(i)}.$$

While the original formulation of PLS scales with the number p of variables, it is also possible to represent the algorithm in a way that it only scales with the number n of observations [28,29]. This leads to a dramatic decrease in computation time for $p \gg n$. Note that the number of PLS components is a model parameter that has to be optimized for each of the p regression models (4). The standard model selection techniques are cross-validation or information criteria based on degrees of freedom [30]. In the context of gene regulatory networks, Tenenhaus et al. [23] propose to use the same number of components m for all p regression models. They observe empirically that the partial correlation coefficients (Def. 1) obtained from PLS regression reach a plateau when the number of PLS components m increases, and suggest a heuristic procedure to choose the smallest m for which the plateau is reached. However, in our experiments, we use the theoretically well-funded and popular cross-validation technique with k folds.

As the PLS coefficients are not sparse, the obtained partial correlations are in general non-zero. Thus, a statistical testing procedure has to be used to determine which edges are significant. (Alternatively, one might

also use a sparsification of PLS as proposed by Chun & Keles [31].) In the present article, we use large-scale simultaneous hypothesis testing with local false discovery rate (fdr) level 0.2, in order to identify unusual outliers among the estimated partial correlations.

For the sake of completeness, let us mention in this section a variant of the PLS approach described above, which was recently suggested by Pihur et al. [10]. Instead of estimating the partial correlation using Eq. (7), they propose an alternative measure of correlation strength which is very similar to the PLS-based partial correlation coefficient except that, roughly speaking, the square root of the product of $\hat{\beta}_{j,pls}^{(i)}$ and $\hat{\beta}_{i,pls}^{(j)}$ is replaced by their sum. We remark that Pihur et al. do not optimize the number of PLS components m and recommend to use $m \approx 3$.

Ridge Regression
Ridge regression (see e.g. [32]) is probably the most popular and most straightforward regularized regression technique. Regularization is performed by adding a penalty term $P(\boldsymbol{\beta})$ to the least squares criterion (5). Ridge regression is based on an, ℓ_2 penalty term of the form

$$P(\boldsymbol{\beta}) = \lambda \|\boldsymbol{\beta}\|_2^2 = \lambda \sum_i \beta_i^2, \qquad (8)$$

where $\lambda > 0$ denotes the penalty parameter. This leads to a reduction of variance and thus avoids overfitting. The solution obtained by ridge regression depends on the penalty parameter λ. In our paper, we use standard k-fold cross-validation to select the optimal amount of penalization λ. As ridge regression does not lead to sparse solutions, we use large-scale false discovery rate multiple testing [14] to test for significant edges, as described above in the subsection on PLS. Again, we adopt a level of 0.2.

The Lasso
Meinshausen and Bühlmann [33] propose to estimate the regression coefficients in Def. 1 with the Lasso [34] and study under which conditions model selection consistency applies, hinging on the choice of the penalty. Similarly to Ridge Regression, the estimated regression coefficients are chosen to minimize a penalized least squares criterion. Lasso regression is based on a ℓ_1-penalty of the form

$$P(\boldsymbol{\beta}) = \lambda \|\boldsymbol{\beta}\|_1 = \lambda \sum_i |\beta_i|, \qquad (9)$$

where $\lambda > 0$ is the regularization parameter. With the ℓ_1-penalty, many estimated regression coefficients will be equal to 0. As a result, with variable selection in mind,

the Lasso has a major advantage: a sparse estimator of the matrix of partial correlations is yielded and a graph can be obtained by assigning an edge between two genes if and only if $\hat{\rho}_{ij,\text{lasso}} \neq 0$. The choice of the penalty λ has to be determined for each of the p high-dimensional regressions successively. Again, this can be done using some cross-validation scheme or information criteria. Meinshausen & Bühlmann [33] motivate a choice of the penalty parameter that aims at controlling the probability of falsely connecting two nodes in the graph, i.e. that is a choice tailored to the graph structure. However, experiments [6] indicate that this approach leads to graphs that are too dense, i.e. too many edges are selected. Therefore, in this paper, we use the oracle penalty for optimal prediction that is determined using k-fold cross-validation.

The two-stage adaptive Lasso
The Lasso is only asymptotically consistent for covariance selection when requiring certain necessary conditions among the variables in the GGM. Zhou et al. [35] show that the two-stage adaptive Lasso procedure [36] is consistent for high-dimensional model selection in graphical Gaussian models under rather general and less restrictive conditions. The adaptive Lasso [36] considers the Lasso with penalty weights as

$$P(\boldsymbol{\beta}) = \lambda \sum_i \hat{\omega}_i |\beta_i|, \qquad (10)$$

where the weights $\hat{\omega}_i$ are chosen in a data-dependent manner. Specifically, the adaptive Lasso is defined as follows. Suppose $\hat{\boldsymbol{\beta}}$ is a $\sqrt{n}$ consistent initial estimator of $\boldsymbol{\beta}$. For example, we can use the least squares estimator $\hat{\boldsymbol{\beta}}_{\text{ols}}$. Pick a $\gamma > 0$, and define the weights $\hat{\omega}_i = 1/\left|\hat{\beta}_{i,\text{ols}}\right|^\gamma$. The most common choice is $\gamma = 1$. Here, we use the Lasso estimator $\hat{\boldsymbol{\beta}}_{\text{lasso}}$ as initial estimator, and define the weights

$$\hat{\omega}_i = 1/\left|\hat{\beta}_{i,\text{lasso}}\right|.$$

Note that the amount of penalization in both the initial stage Lasso and the second stage Lasso with penalty weights is determined via k-fold cross-validation. The adaptive Lasso will be at least as sparse as the Lasso. For graphical Gaussian modeling, the adaptive Lasso estimates are used in Def. 1, and two genes are connected if and only if the partial correlation coefficient $\hat{\rho}_{ij,\text{adaptive lasso}} \neq 0$. We remark that for model selection, the optimal weights have to be determined in each of the k cross-validation splits. As the optimal weights themselves are determined via k-fold cross-validation, this implies that a lasso fit has to be computed k^2 times! This leads to high computational costs.

Results

In this section, we perform extensive experiments to compare regression-based methods for reconstructing gene regulatory networks. We consider the recently proposed techniques PLS regression and Lasso regression, and the two additional methods, ridge regression and adaptive Lasso regression, that have not been applied in practice for this purpose before.

As a reference method, we use the shrinkage approach to covariance estimation, followed by matrix inversion. An overview of the five considered methods and their respective parameters and characteristic features is given in Table 1. All methods are implemented in the R package "parcor" [37], available from the R repository CRAN.

Simulations

The performance of the proposed methods is assessed in a simulation study with a set-up similar to [6]. The number of variables is fixed at $p = 100$, and various sample sizes ranging from 25 to 200 in steps of 25 are investigated. We consider two different scenarios. First, we simulate networks with varying degree of density and no network topology, and second, we investigate sparse networks with different network topologies (see additional file 1 for an illustration and below for a detailed explanation). These scenarios correspond to particular choices of the partial correlation matrix P (see below). For all experiments, a total of 20 replications are performed for each sample size to average out variability due to random sampling. For each replication, the data are drawn randomly from a multivariate normal distribution with correlation structure derived from P.

Varying degree of density

Partial correlation matrices P of size $p \times p$ with a proportion of

$$d = 5\%; 10\%; 15\%; 20\%; 25\%$$

non-zero entries are constructed by first drawing the non-zero entries from a uniform distribution on [-1, 1] and then rescaling the non-diagonal entries to ensure that we obtain a feasible partial correlation matrix (for more details, see the R-package GeneNet[38]). Hence,

the range of the non-zero partial correlations depend on the density of the network. If the network is rather dense, the absolute values of the non-zero partial correlation coefficients are very small compared to a sparse network. This is illustrated in the additional file 2. Here, we plot the histogram of the non-zero partial correlations for a random matrix P of density d. It is important to note that due to the small values, the reconstruction of the network becomes more delicate for a higher degree of density: it is more difficult to select the correct non-zero entries if their true vales are close to zero. We remark that this effect cannot be entirely eliminated by a more clever simulation design, and that the simulation of partial correlation matrices is far from trivial [39].

For each generated data set, P is then estimated based on PLS regression, ridge regression, the Lasso, the adaptive Lasso and the shrinkage covariance estimator, successively. For all regression-based methods, $k = 5$-fold cross-validation is used to optimize the model parameters, i.e. the number of components m for PLS and the penalty λ for ridge regression, the Lasso and the two-stage adaptive Lasso, respectively. For the Lasso and the adaptive Lasso, we follow the parametrization implemented in the lars package [40], based on the ratio of the ℓ_1-norm of the Lasso and the ℓ_1-norm of the least squares estimates. Specifically, the regularization parameter is chosen from an equidistant sequence between 0 and 1 of length 1000.

Furthermore, we normalize this parameter to avoid the peaking phenomenon at $n = p$ (see [41] for details). For ridge regression, we consider a logarithmically spaced sequence $l_1,..., l_{1000}$ ranging from 10^{-10} to 10^{-1}. The candidate penalty parameters are then defined as $\lambda_s = l_s n p$ (with $s = 1,..., 1000$). Finally, the range of the number of PLS components is from 1 to 15.

We evaluate the accuracy of the resulting estimators in two respects: (i) the estimation error of the partial correlation matrix itself, and (ii) the recovery of the underlying networked topology. The difference between the estimated and true matrix of partial correlations is measured in terms of the mean squared error (MSE). In the upper left panel of Figures 1, 2, 3, 4 and 5, the MSE is displayed as a function of the sample size n.

Table 1: Overview of the methods

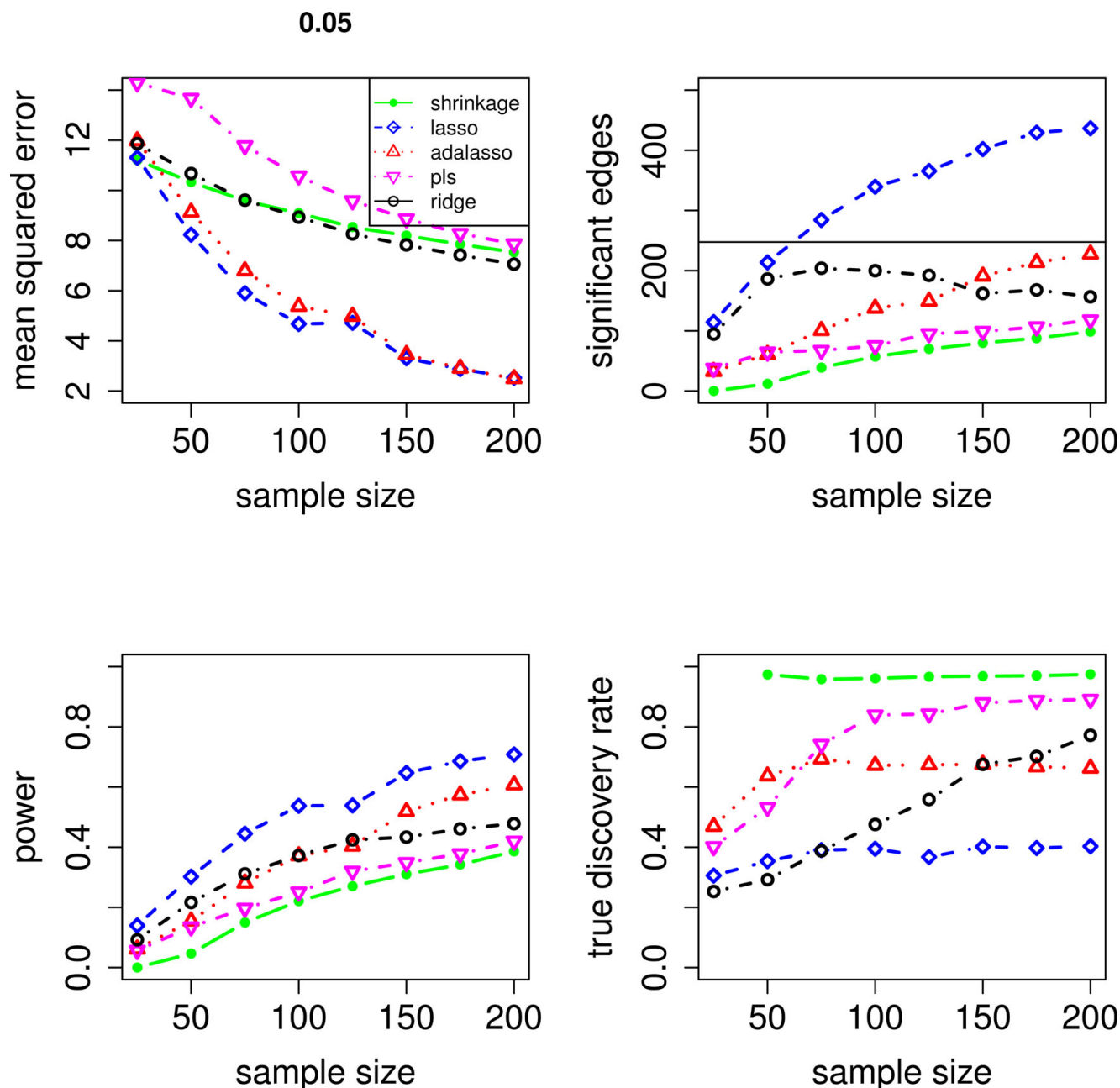| Method	Type	parameter(s)	choice	edge if
shrink	regularized estimation of the covariance	shrinkage intensity λ	analytic	fdr < 0.2
pls	regression	number of components m	CV	fdr < 0.2
ridge	regression	penalty λ	CV	fdr < 0.2
lasso	regression	penalty λ	CV	$\hat{\rho}_{ij} \neq 0$
adalasso	regression	penalty λ (×2)	nested CV (×2)	$\hat{\rho}_{ij} \neq 0$

Figure 1
MSE, number of edges, power and true discovery rate for a density of 0.05.

For sparse networks, the two sparse estimates based on the Lasso and the adaptive Lasso, respectively, yield a lower MSE compared to the three other methods that are not sparse and are likely to contain many non-zero but non-significant (small) entries, which ultimately lead to a higher MSE. This effect vanishes for higher degrees of density. A notable exception is PLS. For denser networks, the MSE becomes larger. These networks correspond to small absolute values of the entries in P. Therefore, we

conjecture that PLS is not able to shrink the regression coefficients enough, as the regularization parameter m (number of components) is discrete. This is in contrast to the four other methods. Note however that for the reconstruction of the underlying networked topology the MSE is only of secondary interest.

For each investigated sample size, the resulting number of selected edges is displayed in the upper right panel
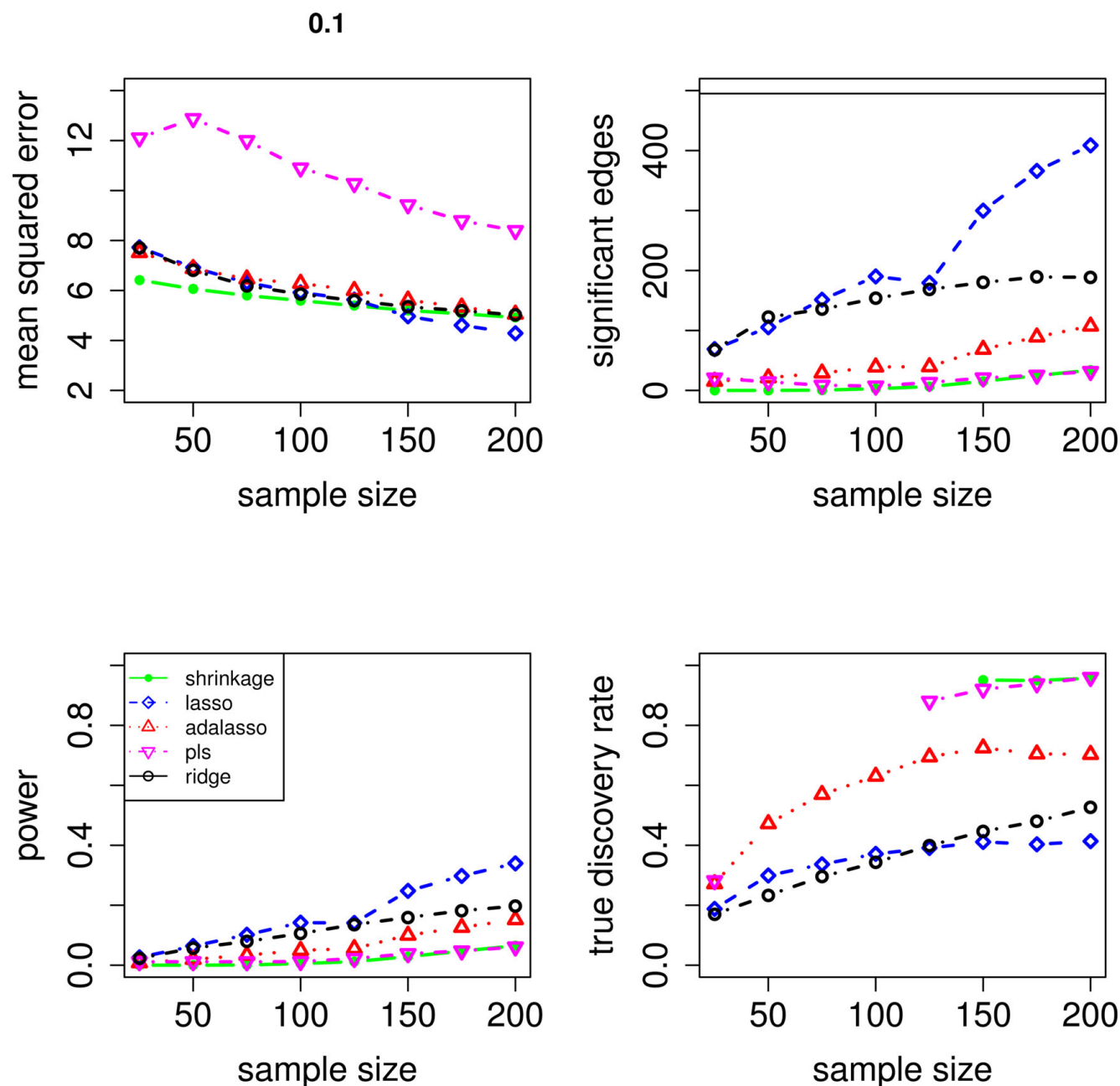
Figure 2
MSE, number of edges, power and true discovery rate for a density of 0.10.

of Figures 1, 2, 3, 4 and 5, while the horizontal line is the number of true edges. For sparse networks, the Lasso with its regularization parameter chosen to be prediction optimal tends to select too many edges. PLS, ridge regression and the approach based on shrinkage covariance estimation are in contrast far more conservative and rather select too few edges, even in the $n > p$ case. The adaptive Lasso is less conservative and appears to be a promising alternative. Again, these differences vanish for higher degrees of densities. As remarked above, the reconstruction task becomes more difficult for higher degrees of density. This explains the low number of selected edges for higher degrees of density.

The two lower panels in Figures 1, 2, 3, 4 and 5 correspond to the power (left) and the true discovery rate (tdr, right) which are defined as
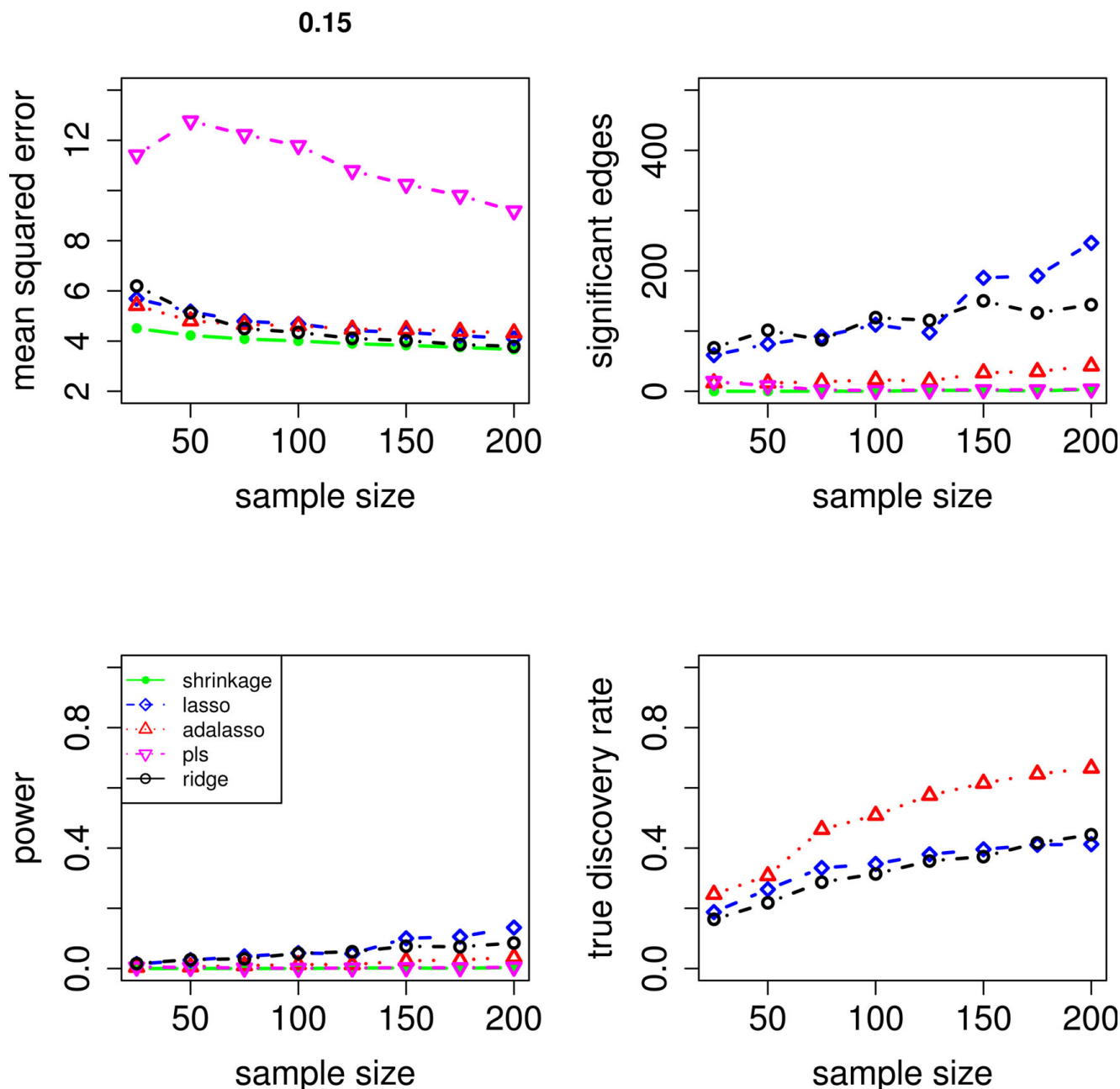
Figure 3
MSE, number of edges, power and true discovery rate for a density of 0.15.

$$\text{power} = \frac{\#\{\text{true edges that are selected}\}}{\#\{\text{true edges}\}} \quad \text{and}$$

$$\text{tdr} = \frac{\#\{\text{true edges that are selected}\}}{\#\{\text{selected edges}\}},$$

respectively. The panels illustrate that for sparse networks, the Lasso's comparatively high power comes at the prize of rather low true discovery rate. Again, the

power decreases with the increase in density of the network. In many practical applications, we argue that it might be more valuable to report more stable results with fewer false positives.

However, it is to be noted that the non-sparse methods using fdr-based procedures for edge selection involve an arbitrary parameter: the fdr threshold (here 0.2). These methods can thus be made more or less sparse by changing
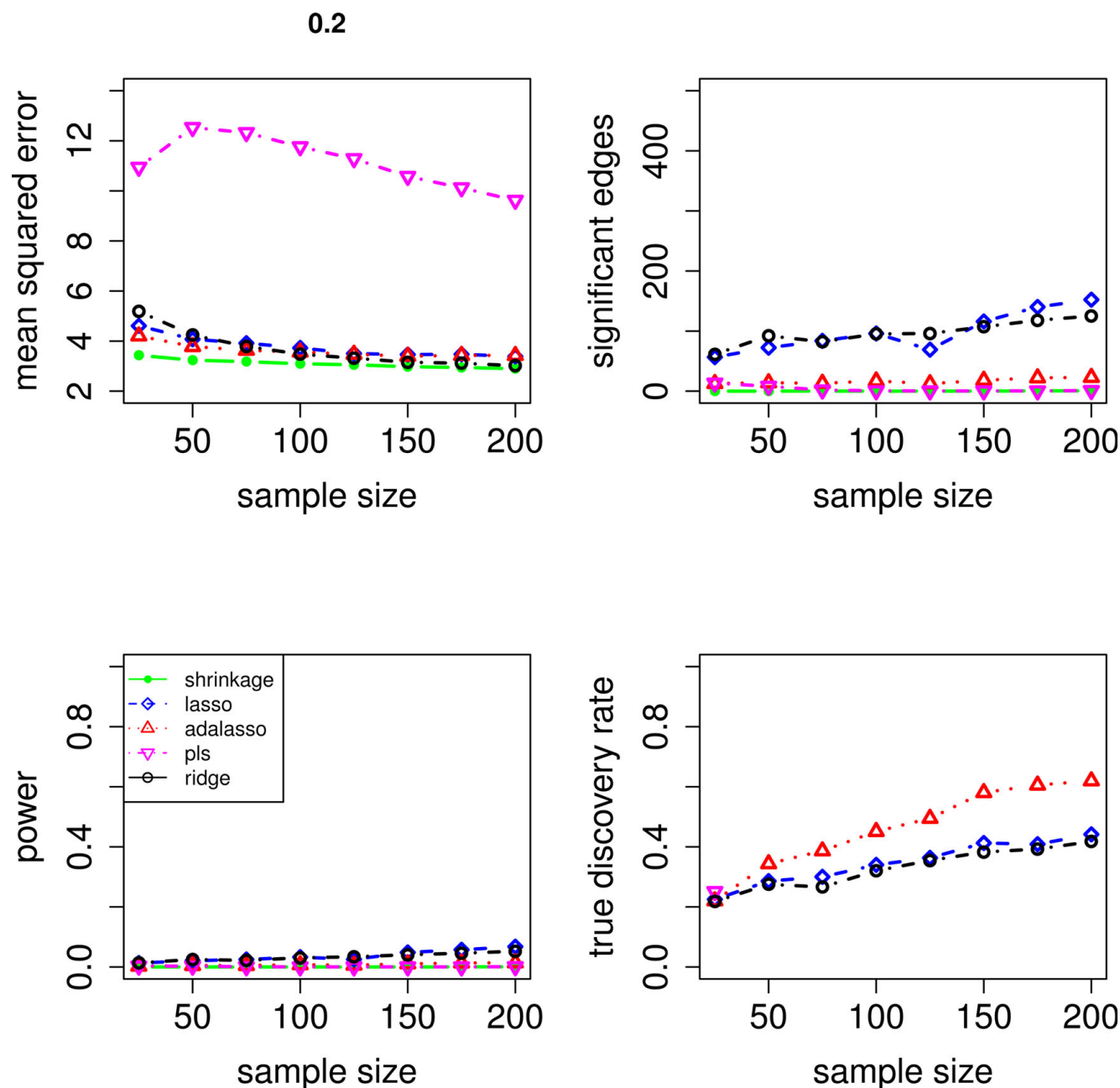
Figure 4
MSE, number of edges, power and true discovery rate for a density of 0.20.

the threshold value. To investigate the relative accuracy of the non-sparse methods independently of the particular fdr threshold, the same simulations are subsequently performed with other thresholds. In order to evaluate the ability of the three methods to detect non-zero partial correlations, their sensitivity and specificity are computed for these different fdr thresholds and displayed graphically in form of ROC curves (see additional files 3, 4, 5, 6, 7, 8,

9, 10, 11, 12). PLS and ridge regression yield very similar results. They slightly outperform the approach based on shrinkage covariance estimation. The sensitivity and specificity of the Lasso and the adaptive Lasso, which do not depend on a particular threshold, are depicted as single points. They are above the ROC curves of the three non-sparse methods, indicating good performance - especially for the adaptive Lasso.
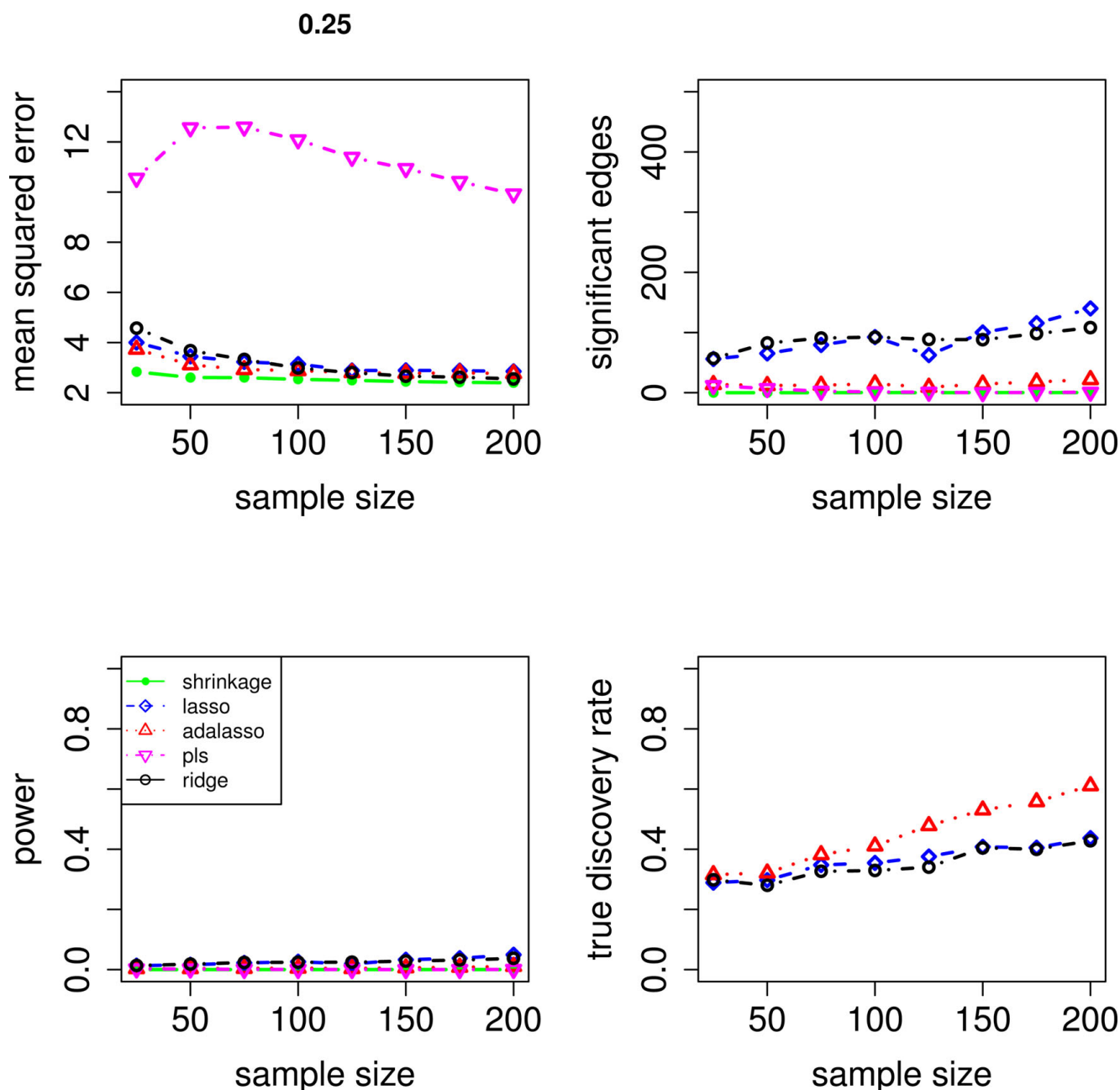
0.25

Figure 5
MSE, number of edges, power and true discovery rate for a density of 0.25.

Finally, we compare the runtime of the respective methods in Figure 6. Note that we do not display the runtime of the Lasso, as it was computed as an intermediate step in the R-function for the adaptive Lasso. The left part of Figure 6 clearly shows that the computational load for the adaptive Lasso is very high. This is due to the fact that we have to run the lasso algorithm k^2 times in k-fold cross-validation, and that

the (adaptive) lasso algorithm scales unfavorably in the number of variables - in contrast to PLS, Ridge Regression or shrinkage. The right part of Figure 6 only displays the runtime of the three latter methods. Shrinkage is faster than the regression based approaches as it circumvents both time-consuming cross-validation and the computation of p different regression models. The discrepancy with respect to the
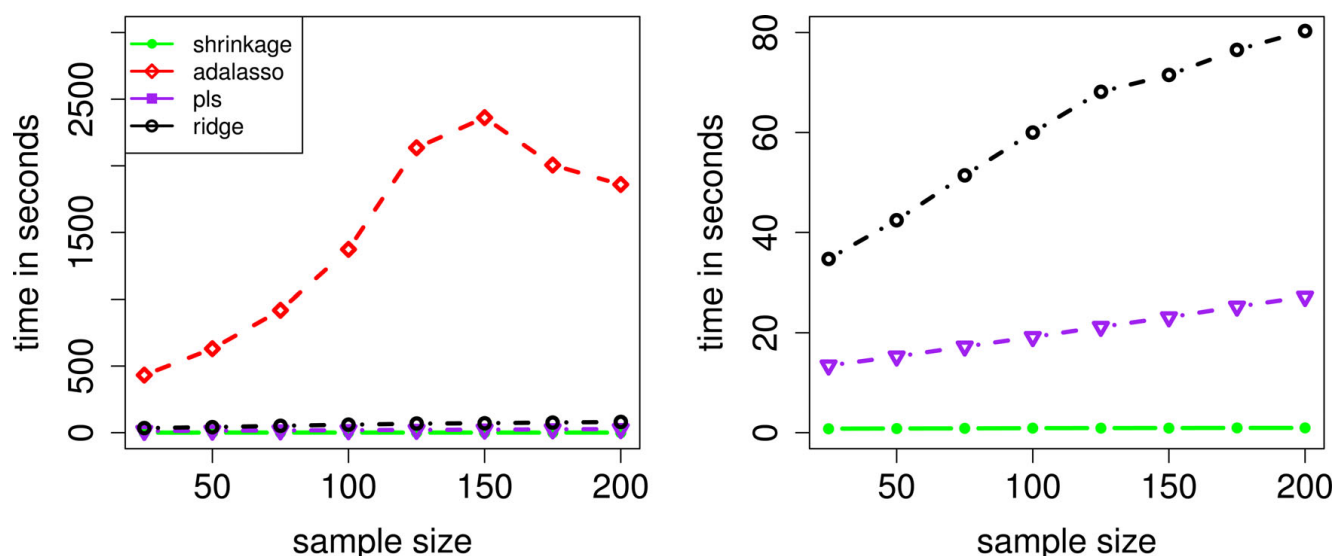
Figure 6
Runtime of the respective methods.

runtime becomes even more apparent in the real data study (see below).

Different network topologies
Next, we consider different network topologies. We simulate two different types of topologies (see additional file 1) The left part of the figure shows three clusters of genes. In each cluster, all genes are partially correlated, and genes from different clusters are not partially correlated. In the simulation, we consider networks with 1,2 and 3 clusters. The right part of the figure in clusters.pdf shows three star-shaped clusters. In each star, all genes are partially correlated to one gene, the center of the star. In the simulation, we consider a network with 3 stars. The MSE, the number of selected edges, the power and the true discovery rate are displayed in Figures 7, 8, 9 and 10. Again, we observe a high MSE for PLS in most scenarios. As explained above, this is probably due to the insufficient shrinkage of PLS towards 0. Overall, the Lasso and Ridge Regression perform best in these scenarios. So, in contrast to what is often conjectured/ reported in the literature, we do find in our simulations that sparse methods are able to reconstruct networks in the presence of cluster structures.

Real Data Study
We compare the five different methods on diverse real world data sets: the ecoli1[42] and ecoli2[43], Ara[44], t.cell10 and t.cell34[3], and west[45] data sets. All data sets are freely available. An overview of the size, characteristics and availability of the data sets is given in Table 2. The five considered methods (shrinkage covariance estimation, ridge

regression, PLS, Lasso, adaptive Lasso) including the model selection procedures for the regression-based approaches are exactly as in the simulation setting. For ecoli2, we use leave-one-out-cross-validation for model selection, and for west, we use $k = 5$-fold cross-validation. For the remaining 4 data sets, we use $k = 10$.

In real world scenarios, the ground truth, i.e. the true underlying network, is hardly ever known, and the performance of different methods cannot be determined in terms of MSE, power and tdr as in the simulation study. Nevertheless, it is possible to compare the performance of the different methods quantitatively. In particular, we investigate the size and the connectivity of the estimated graphs, their overlap the type of interaction between genes and their stability.

Figures 11 and 12 display the percentage of selected edges for each data set. As in the simulation study, the proportion of selected edges strongly depends on the chosen estimation method. More surprisingly, the relative levels of sparsity of the obtained graphs show very different patterns for the six investigated data sets. The Lasso and adaptive Lasso seem to behave very differently from the other methods. This can at least partly be explained by the fact that they rely on a completely different edge selection scheme which essentially depends on the sparsity of the regression method and not on the testing scheme.

In a nutshell, the Lasso and adaptive Lasso select less edges than the other methods for all data sets except for the two data sets t.cell10 and t.cell34 with repeated measurements.
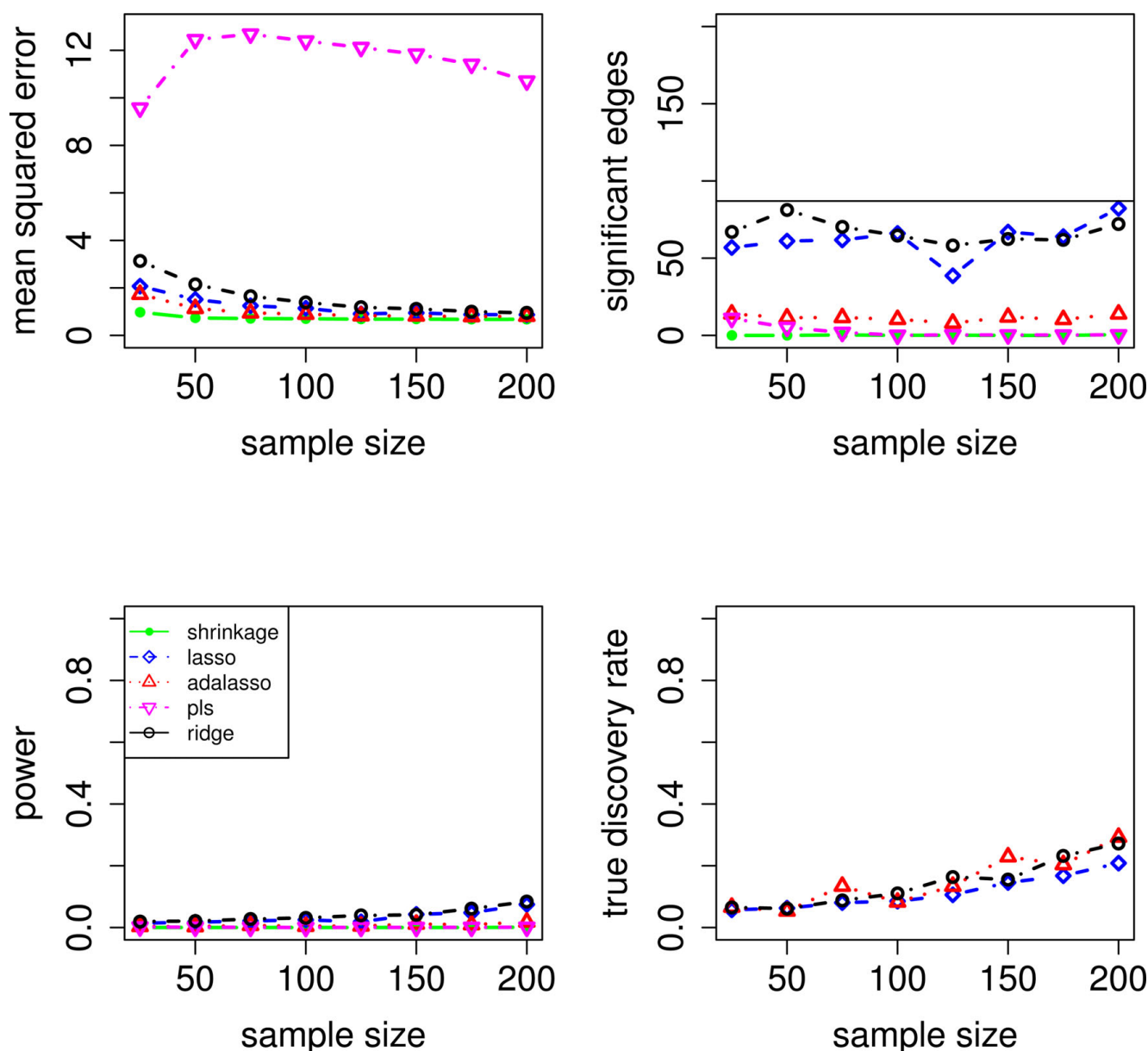
Figure 7
Network topology: 1 cluster.

With these two data sets, Lasso and adaptive Lasso yield complex graphs with as much as over 50% non-zero edges (t.cell34 data). This behavior is likely to be due to the longitudinal structure of the data that is not explicitly considered, since the standard Lasso regression method assumes independent observations. In contrast, longitudinal structures may be handled in an implicit way by methods using an fdr-based assessment, where the distribution under the null hypothesis is estimated from the data. To gather further evidence for our hypothesis, we average over the 10 replications in the two respective data

sets. This leads to 10 observations for t.cell10 and 34 observations for t.cell34. On the averaged data, both Lasso and adaptive Lasso indeed select far less edges: For the averaged t.cell10, we obtain: 4.2% (Lasso), 2.0% (adaptive Lasso), 12.2% (PLS), 0.2% (Ridge), 0.2% (shrinkage). For the averaged t.cell34, we obtain 12.3% (Lasso), 4.8% (adaptive Lasso), 11.9% (PLS), 2.7% (Ridge), 0.1% (shrinkage).

PLS reconstructs very dense networks for five out of the six data sets (ecoli1, ara, t.cell10, t.cell34 and west). In
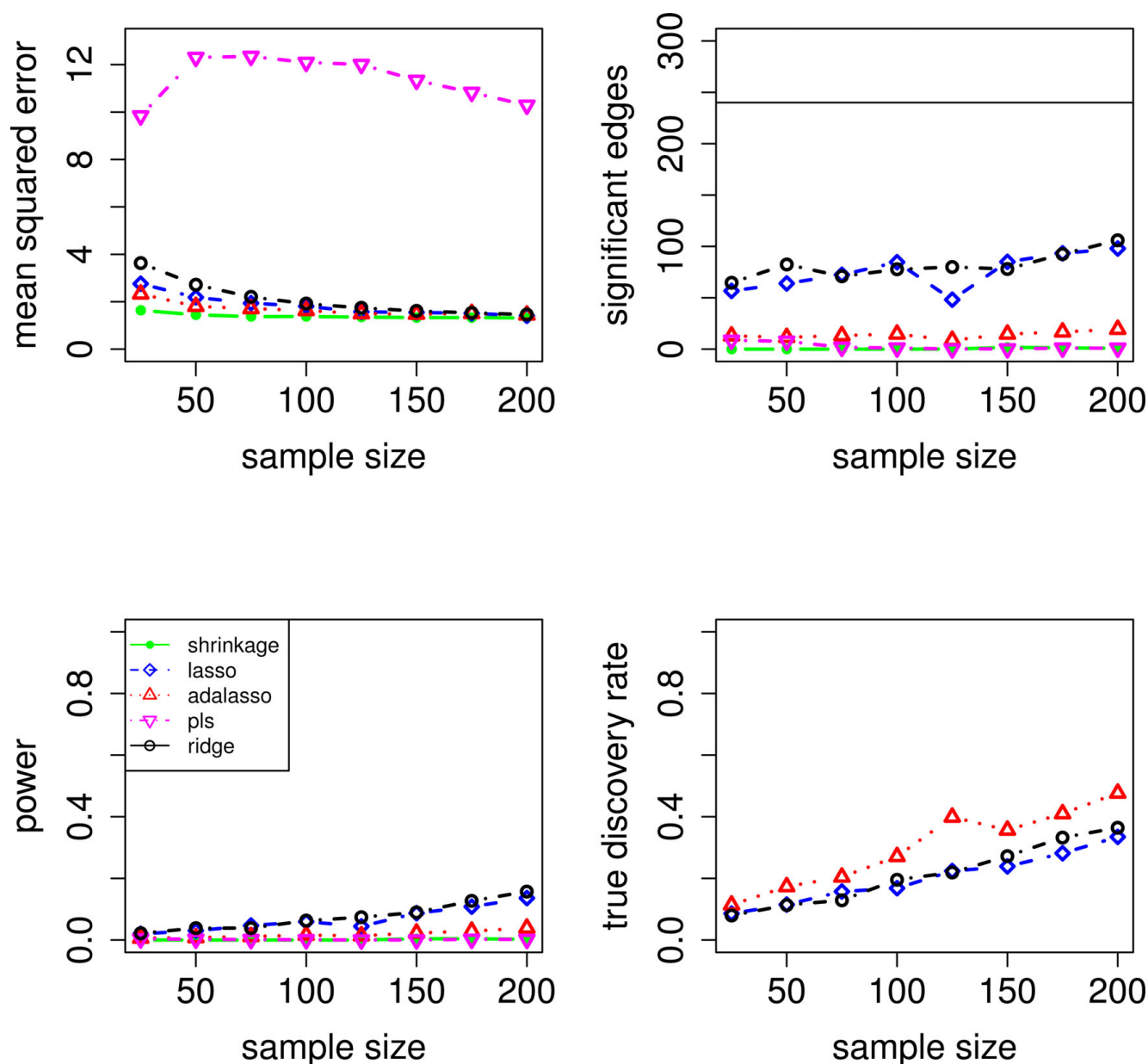
Figure 8
Network topology: 2 clusters.

combination with the high MSE that we observed in the simulations, we conjecture that PLS in combination with cross-validation is not the most reliable method for the reconstruction of networks. We believe that other model selection strategies or the incorporation of sparse PLS [31] are necessary in order to improve the performance of PLS.

Among the three methods with fdr-based assessment of the edges, i.e PLS, ridge regression and shrinkage covariance estimation, the latter procedure seems to be most conservative, whereas PLS identifies the highest number

of edges. This result is consistent for all six real data sets and yields a refinement of the results presented in the simulation study, where these three methods performed similarly.

Table 3 displays the overlap of the estimated graphs. (Example: On the ecoli1 data set, 68; 6% of the edges found by Ridge Regression are also found by PLS. For baseline comparison, the numbers in *italics* show the percentage of selected edges for the respective methods.) The estimated graphs show a moderate overlap between
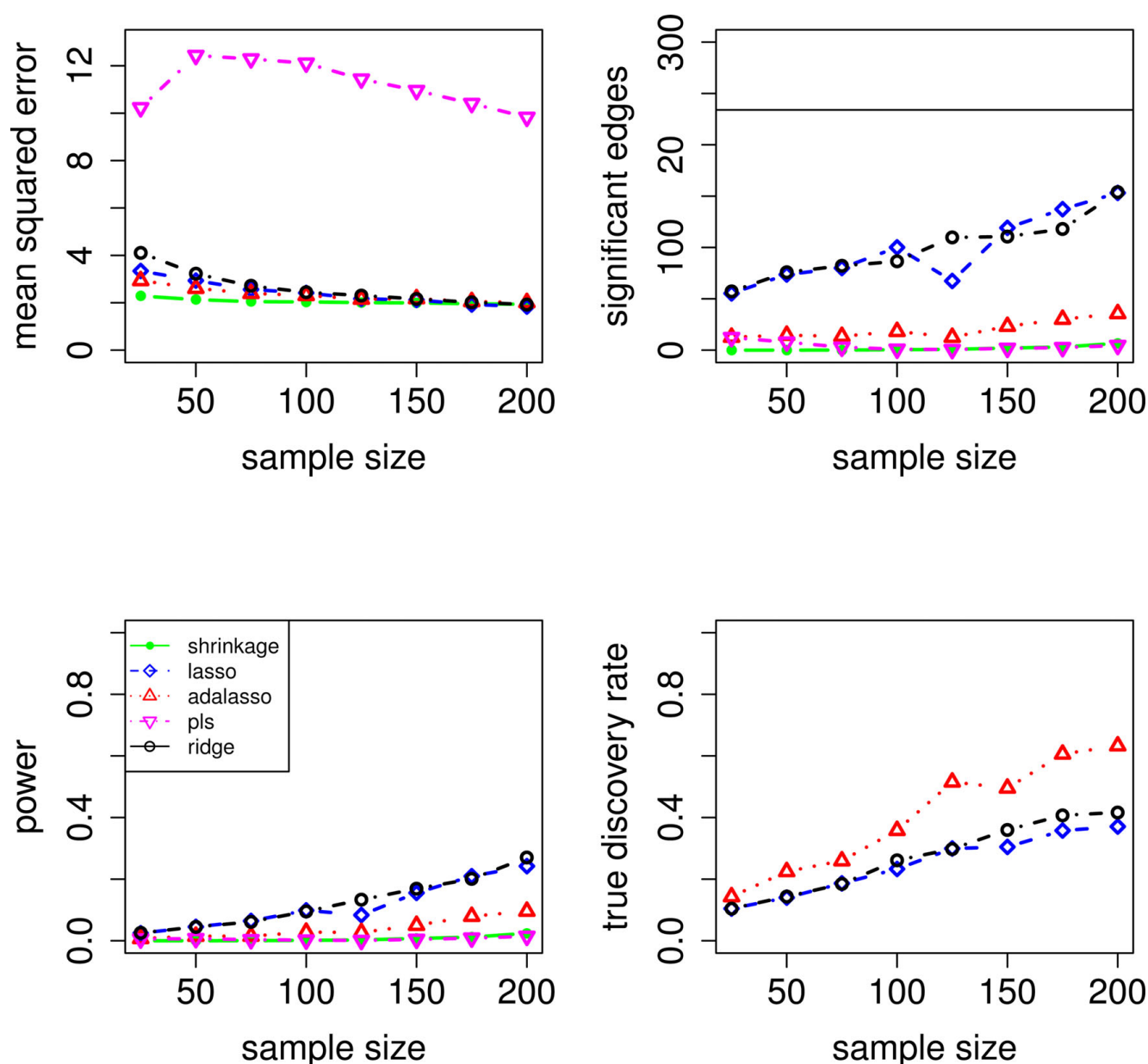
Figure 9
Network topology: 3 clusters.

the methods. While considering these results, one should keep in mind that the proportions of selected edges vary a lot across the five methods, which of course decreases the overlap considerably: a very sparse graph can obviously include only a very small proportion of the edges of a more complex graph. Interestingly, the overlap seems to be higher on average for the west data set including the highest number of genes than for the other five data sets. We remark that the Lasso and adaptive Lasso solutions are computed based on different, random cross-validation splits. This explains that, in

general, the graph found by adaptive Lasso is *not* exactly a subgraph of the solution found by Lasso.

Figures 13 and 14 display the connectivity of the estimated graphs for each of the six data sets. For each gene, we derive the proportion of genes that are connected to it through an edge, with each of the six data sets and each of the five methods. Each boxplot depicts the distribution of the proportion of connected genes for the considered method and the considered data set. As explained above, the assumption of i.i.d. observations is violated for the data
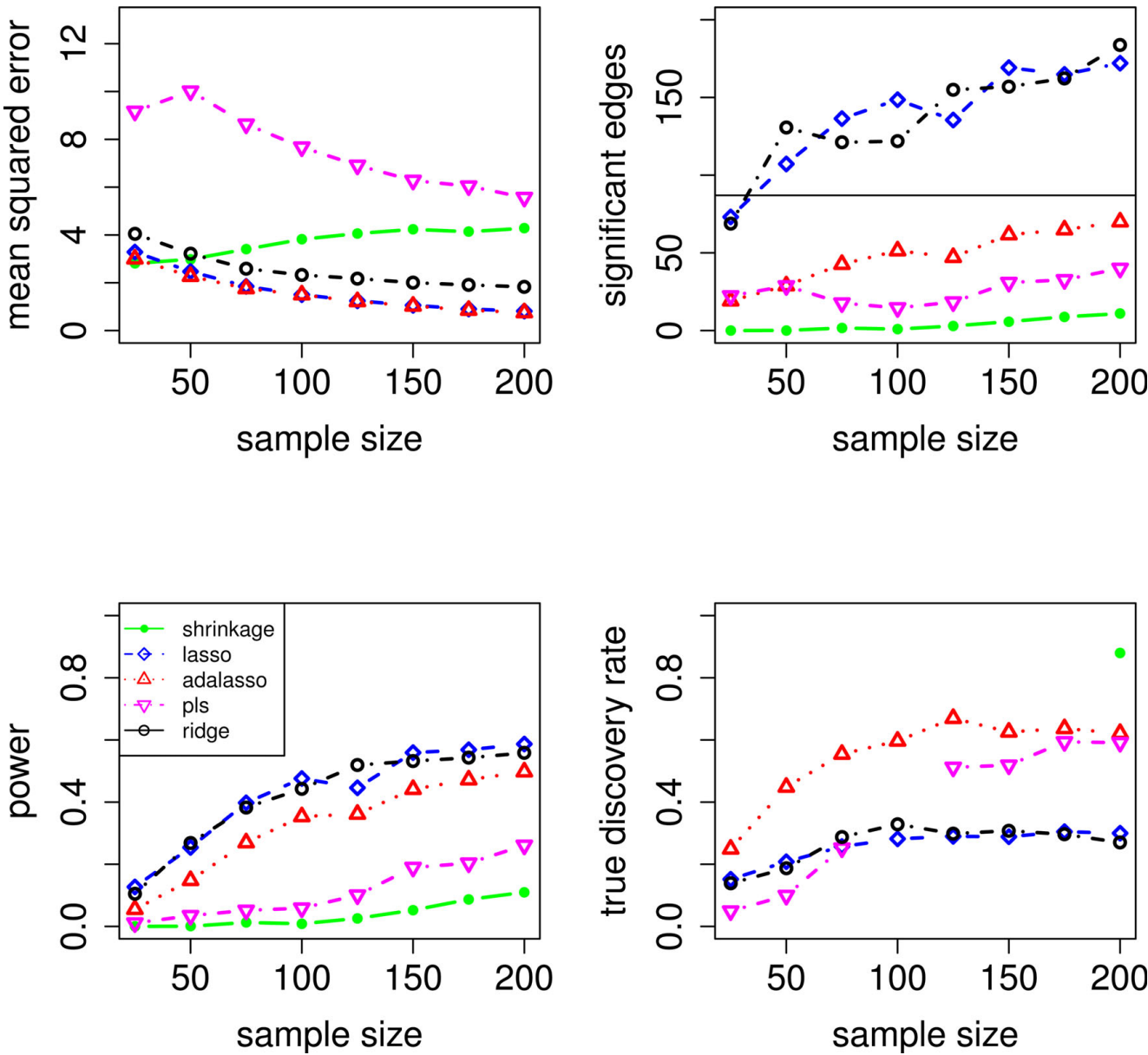
Figure 10
Network topology: 3 stars.

Table 2: Size of the data sets

| data set	arrays	genes	time series	repeated measurements	size of full graph	Availability
ecoli1	23	100	yes	no	4 950	R package plsgenomics[55]
ecoli2	9	102	yes	no	5 151	R package GeneNet[38]
ara	22	800	yes	no	319 600	R package GeneNet
t.cell10	100	58	yes	yes	1 653	R package longitudinal[56]
t.cell34	340	58	yes	yes	1 653	R package longitudinal
west	49	3883	no	no	7 536 903	http://strimmerlab.org/data.html
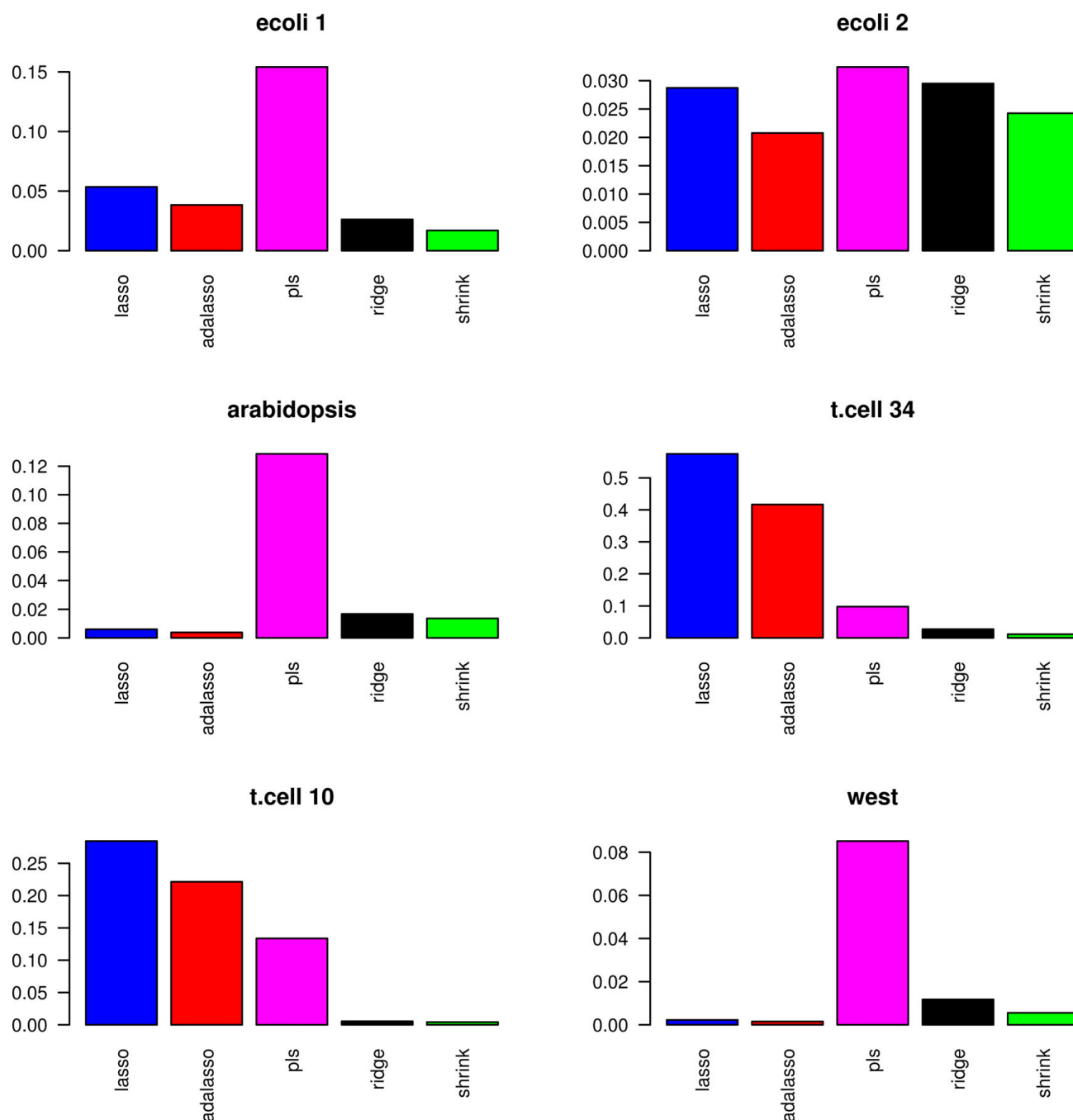
Figure 11
Proportion of selected edges.

sets t.cell10 and t.cell34. This leads to a high number of selected edges for the Lasso and adaptive Lasso (see figures 13 and 14), and consequently to a high number of connected genes for these methods.

Table 4 displays the percentage of positive (> 0) correlations among the edges identified by the five

methods for the six data sets. This proportion varies between 0.5 and 0.8. The results obtained using the five investigated methods seem much more consistent than the results on the number of identified edges. We also compare the methods with respect to their stability. This is an important issue in order to assess the reliability of competing methods. Recent research efforts have e.g.
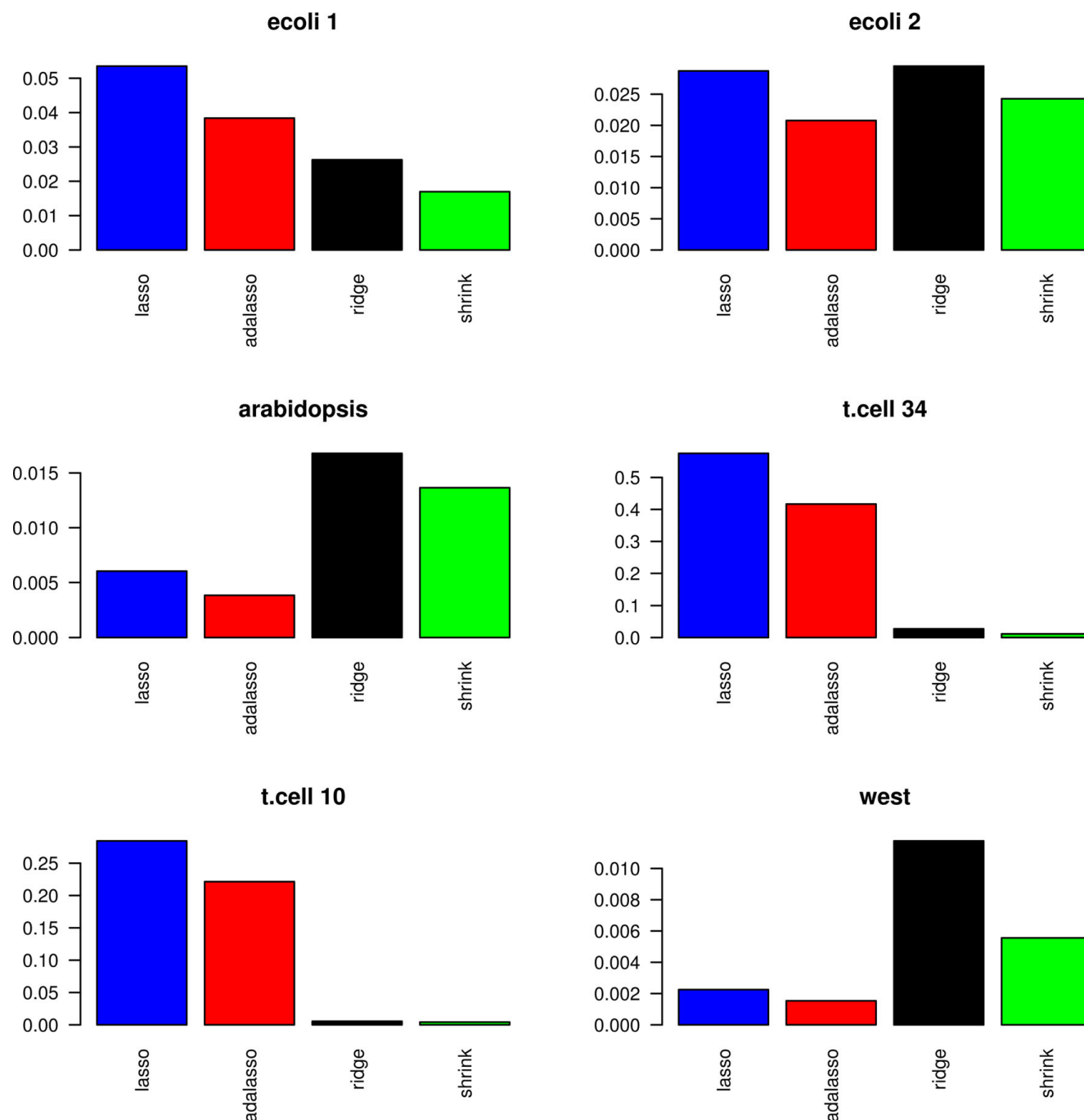
Figure 12
Proportion of selected edges without PLS.

concentrated on the stability of ranked gene lists, variable selection methods and Bayesian networks [46-48]. In our context, a good method is expected to yield a stable network in the sense that a slightly modified data set (for instance a subsample) does not lead to a completely different result. For data sets ecoli1,

ecoli2, t.cell10 and t.cell34, we draw subsamples by excluding ≈ 10% of the observations and compute the network based on each subsample using the five methods successively. The number of considered subsamples is fixed to $R = 10$ (only $R = 9$ for the data set ecoli2 that includes 9 observations). We do not analyze

Table 3: Overlap of the estimated graphs

data set		pls	ridge	lasso	adalasso	shrink
ecoli1	pls	1.000	0.096	0.156	0.127	0.045
	ridge	0.686	1.000	0.600	0.457	0.390
	lasso	0.496	0.267	1.000	0.581	0.165
	adalasso	0.597	0.302	0.862	1.000	0.189
	shrink	0.405	0.488	0.464	0.357	1.000
	% selected	*0.162*	*0.018*	*0.052*	*0.036*	*0.017*
ecoli2	pls	1.000	0.593	0.263	0.156	0.305
	ridge	0.651	1.000	0.309	0.197	0.388
	lasso	0.297	0.318	1.000	0.520	0.311
	adalasso	0.310	0.357	0.917	1.000	0.381
	shrink	0.408	0.472	0.368	0.256	1.000
	% selected	*0.032*	*0.030*	*0.029*	*0.020*	*0.024*
ara	pls	1.000	0.064	0.025	0.017	0.035
	ridge	0.590	1.000	0.151	0.108	0.377
	lasso	0.535	0.352	1.000	0.579	0.361
	adalasso	0.556	0.386	0.887	1.000	0.409
	shrink	0.335	0.392	0.161	0.119	1.000
	% selected	*0.126*	*0.018*	*0.006*	*0.004*	*0.014*
t.cell10	pls	1.000	0.314	0.993	0.985	0.131
	ridge	0.956	1.000	1.000	1.000	0.422
	lasso	0.141	0.047	1.000	0.795	0.020
	adalasso	0.170	0.057	0.965	1.000	0.024
	shrink	0.947	1.000	1.000	1.000	1.000
	% selected	*0.109*	*0.027*	*0.575*	*0.417*	*0.011*
t.cell34	pls	1.000	0.053	0.762	0.670	0.031
	ridge	1.000	1.000	1.000	1.000	0.583
	lasso	0.345	0.024	1.000	0.643	0.014
	adalasso	0.433	0.034	0.917	1.000	0.020
	shrink	1.000	1.000	1.000	1.000	1.000
	% selected	*0.134*	*0.005*	*0.284*	*0.221*	*0.004*
west	pls	1.000	0.089	0.017	0.008	0.041
	ridge	0.667	1.000	0.118	0.062	0.236
	lasso	0.643	0.611	1.000	0.407	0.404
	adalasso	0.673	0.694	0.884	1.000	0.458
	shrink	0.632	0.488	0.161	0.084	1.000
	% selected	*0.086*	*0.011*	*0.002*	*0.001*	*0.006*

the data sets ara and west, because repeated experiments would be computationally too expensive.

For each candidate edge i, $n_i^{(1)}$ counts how often this edge is selected across the R subsamples. Similarly, $n_i^{(0)} = R - n_i^{(1)}$ denotes the number of times the ith edge is not selected. These frequencies are summarized using Fleiss' κ-score [49] which measures the degree of agreement among the R subsamples of the data. The measure is defined as follows. We first compute the average proportion of assignments

$$p^{(l)} = \frac{1}{R \times \# \text{ edges}} \sum_{i=1}^{\# \text{ edges}} n_i^{(l)}, l = 0, 1.$$

Further, the degree of agreement of the R subsamples for the ith edge is measured as

$$P_i = \frac{1}{R(R+1)} \left[\sum_{l=0}^{1} \left(n_i^{(l)} \right)^2 - R \right]$$

Finally, with $\bar{P}$ denoting the average of the P_i's and with $P_c = \left(p^{(0)} \right)^2 + \left(p^{(1)} \right)^2$ denoting the agreement expected by chance, Fleiss κ is defined as

$$\kappa = \frac{\bar{P} - P_c}{1 - P_c}.$$

The score is always ≤ 1, and the higher the value of κ, the more stable the methods are with respect to subsampling.

The κ-score of the methods is given in Table 5. As the absolute values are hard to compare between data sets, we also display the ranking on each data set. The
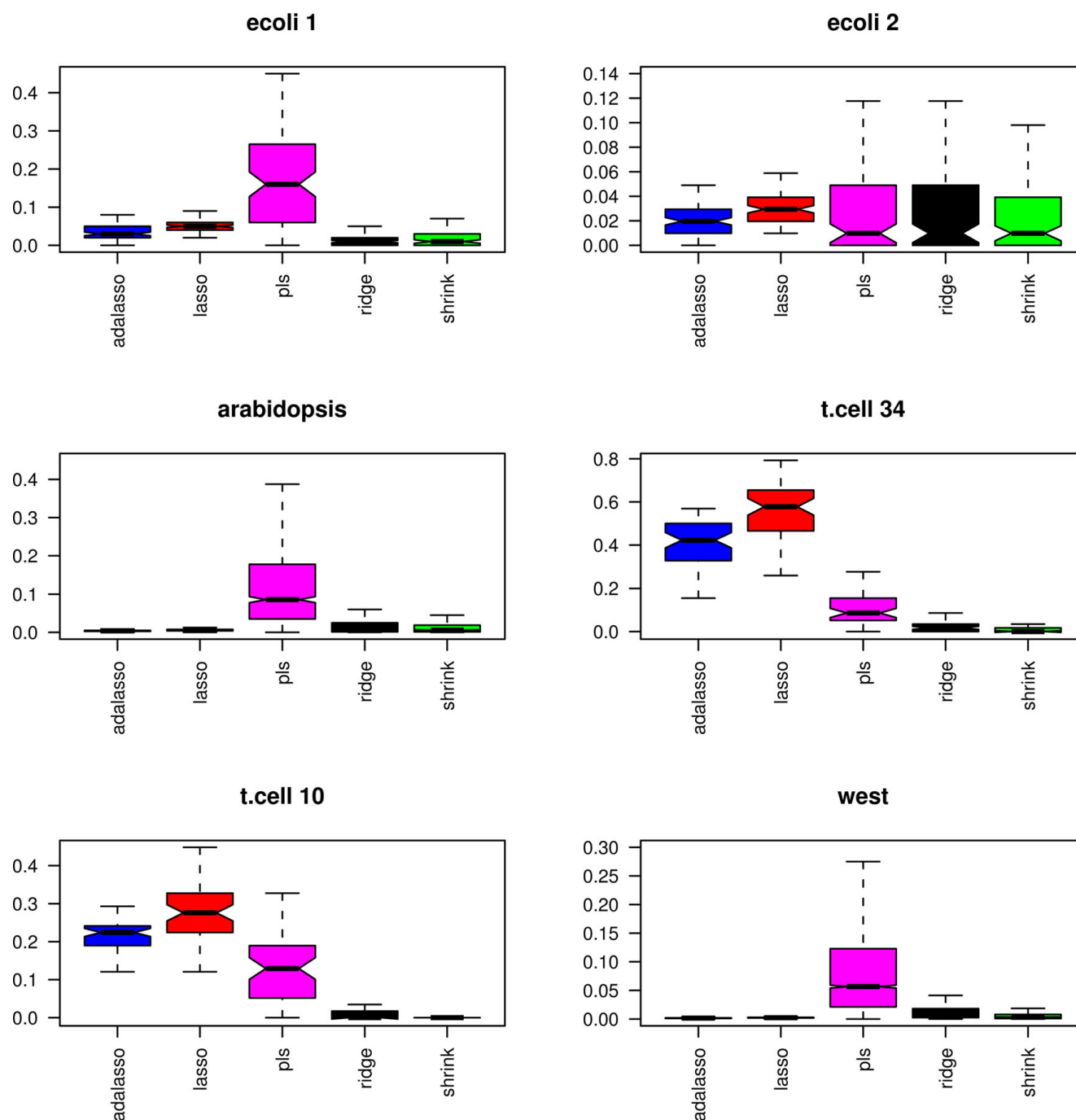
Figure 13
Connectivity: Proportion of connected genes.

shrinkage approach is the most stable, probably because it does not rely on additional subsampling in form of cross-validation splits. The regression based approaches are less stable, but among them, the degree of stability is comparable. In particular, in this experiment, we cannot see any difference between sparse and non-sparse approaches.

Finally, the considered methods differ quite dramatically with respect to their run-time. As an illustration, we compared the run-time on the west data set, which contains 3883 genes. The approach based on shrinkage covariance estimation is by far the most efficient one ($\approx$ 2 min), and all other methods scale within several *hours*: PLS $\approx$ 7.5 hours, ridge regression $\approx$ 10 hours, the
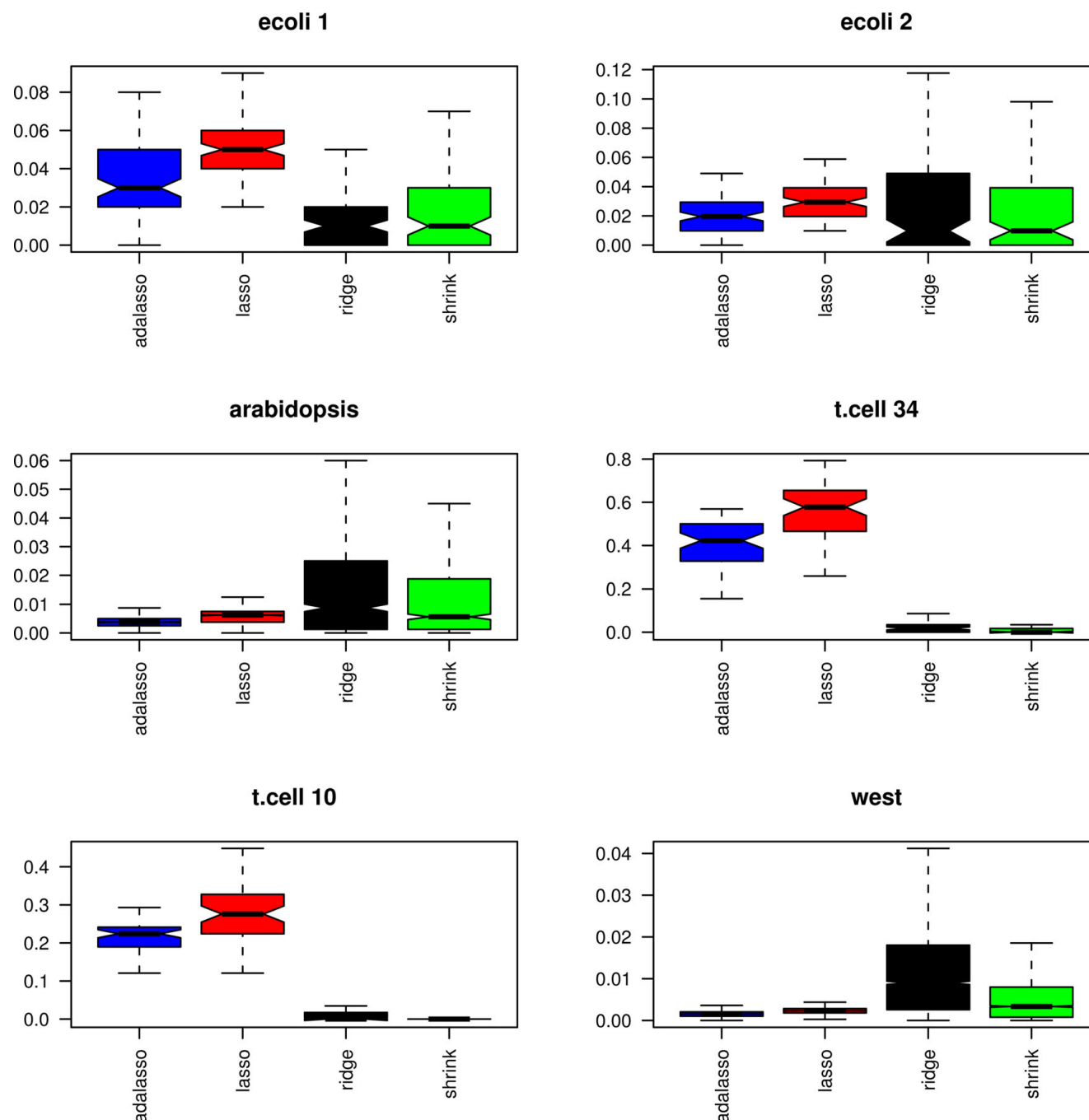
Figure 14
Connectivity: Proportion of connected genes without PLS.

Lasso ≈ 17 hours, and the adaptive Lasso ≈ 3.5 *days*. This can be seen as a major drawback of the methods relying on cross-validation schemes, especially the Lasso-based methods. While Ridge Regression and PLS allow a representation that only scales in the number of observations, Lasso and adaptive Lasso scale in the number of variables. Furthermore, adaptive Lasso requires nested cross-validation. Partial relief can be found in a parallel implementation. Alternatively, for high-dimensional data, one might consider to approximate the Lasso-based networks by first constructing a mildly sparse network without cross-validation (for

Table 4: Percentage of positive correlations

	ecoli1	ecoli2	ara	t.cell10	t.cell34	west
lasso	61.2%	79.6%	63.6%	59.0%	65.7%	77.2%
adalasso	61.9%	81.9%	65.3%	61.0%	67.2%	78.2%
pls	57.5%	77.8%	54.7%	60.0%	66.1%	56.1%
ridge	58.9%	75.0%	55.6%	62.2%	77.8%	71.1%
shrinkage	50.0%	70.4%	55.3%	73.7%	71.4%	72.0%

Table 5: Stability of the Methods

data set	measure	pls	ridge	lasso	adalasso	shrink
ecoli1	κ	0.630	0.510	0.550	0.550	0.593
	ranking of κ	1	5	3.5	3.5	2
ecoli2	κ	0.242	0.280	0.469	0.450	0.486
	ranking of κ	5	4	2	3	1
t.cell10	κ	0.656	0.797	0.670	0.674	0.742
	ranking of κ	5	1	4	3	2
t.cell34	κ	0.655	0.555	0.625	0.619	0.702
	ranking of κ	2	5	3	4	1
	mean ranking of κ	3.25	3.75	3.125	3.375	1.5

example using the method described in [33]), and then to refine this network by running the (adaptive) Lasso with cross-validation.

Discussion

In this paper, we proposed and compared different methods to estimate partial correlation coefficients based on regularized regression techniques with applications to genetic networks. It is remarkable that while we focus on the framework of graphical Gaussian models (and do not consider alternative frameworks as e.g. Bayesian networks), the investigated methods nevertheless show clear differences. Hence, the employed regularization technique for graphical Gaussian models has a considerable effect.

In a simulation study, we assessed the performance of the considered methods in terms of estimation accuracy (MSE) and in terms of reverse engineering of the true underlying networked topology. As a result, the investigated non-sparse methods (PLS, ridge regression, and the approach based on shrinkage covariance estimation that served as a reference method) were found to perform similarly. It is to be noted that these methods have fdr-based significance testing in common. They are rather conservative with respect to the inclusion of edges when used with the standard fdr threshold 0.2. The Lasso tends to produce too "dense" structures, while the adaptive Lasso compensates for that by selecting edges in a two-step approach, therefore leading to sparser graphs. The latter two-stage approach is able to select relevant edges, even for small samples, while

at the same time preventing to be too dense. For denser networks, the performances of the five methods are very similar. On real world data, the behavior of the non-sparse methods is again similar, except that PLS is less conservative than ridge regression and the approach using a shrinkage covariance estimator. A remarkable difference with respect to the different data sets is the behavior of the Lasso and the adaptive Lasso on the t.cell data sets. In contrast to the four other data sets, the t.cell data include replications, thus violating the assumption of independent samples. Consequently, the (adaptive) Lasso does not handle the underlying data structure correctly, while empirical null modeling seems to account for the decreased "effective" sample size in an implicit way.

Note that all investigated methods require the specification of tuning parameters that need to be optimized based on the available data. The choice of the model selection criterion itself strongly influences the results of the methods [50], especially for small n. As an example, the model selection procedure introduces a substantial amount of variation for the Lasso and the adaptive Lasso. In the real world study, we estimate the two graphs on two different random cross-validation splits, which leads to an overlap of only 88.4% on the west data, although the adaptive Lasso graph is defined as a subgraph of the Lasso graph. Hence, tuning parameters should be given much attention in future research when new methods are developed. Moreover, setting the parameters to fixed values without proper selection

procedure (such as cross-validation) and just because they "yield nice results" is an incorrect and biased strategy which may favor the proposed novel method. Furthermore, from a computational point of view, a major strength of the shrinkage approach is that the optimal amount of regularization can be estimated from the data using an analytic formula, thus making time-consuming cross-validation procedures unnecessary.

We want to emphasize that there are interesting alternatives for the detection of significant edges that do not depend on sparsity penalties or testing based on local false discovery rates. For instance, Reverter & Chan [51] propose information theoretic measures for the reconstruction of gene co-expression networks. The comparative performance of these methods and their connections to the approaches investigated above may be explored in future research.

Finally, the methods discussed in this paper can potentially be used for detecting causal interactions [52,53]. For instance, in the presence of longitudinal data, Arnold et al. [53] propose to identify the direction of interactions between variables by investigating partial correlations between time-shifted copies of the variables. Amongst others, they propose to estimate these partial correlations using Lasso regression, but other regression methods might be promising alternatives.

Conclusion
We briefly summarize our findings. A summary of our findings can be found in Table 6.

Performance
In the simulation, the investigated non-sparse regression methods, i.e. Ridge Regression and Partial Least Squares, exhibit rather conservative behavior when combined with (local) false discovery rate multiple testing in order to decide whether or not an edge is present in the network. For networks with higher densities, the

difference in performance of the methods decreases. Both sparse and non-sparse methods can deal with cluster topologies in the network.

For PLS, we observe both a high MSE in the simulations and a high percentage of selected edges in some of the real data. In our opinion, this is an indication that PLS itself might not be too well-suited for the reconstruction of networks. The reasons are that PLS is not sparse by design, and that it does not shrink arbitrarily close to zero. Therefore, we suggest to incorporate sparse versions of PLS instead in future research.

On six real data sets, we also clearly distinguish the results obtained using the non-sparse methods and those obtained using the sparse methods where specification of the regularization parameter automatically means model selection. For data that violate the assumption of uncorrelated observations (due to replications), the Lasso and the adaptive Lasso yield very complex structures, indicating that they might not be suited under these conditions.

Stability
We compared the stability of the methods under two aspects. All regression-based methods are less stable than the shrinkage approach over different subsamples of the data, and within the regression-based approaches, there is no clear difference between sparse and non-sparse methods. However, the two sparse regression methods seem to be unstable with respect to violations of the i.i.d assumption of the samples.

Runtime
The computational load for the Lasso and in particular for the adaptive Lasso is considerable. For very high-dimensional data, this can constitute a severe limitation. The runtime might be decreased by applying parallel computation techniques or by preselecting a coarse network topology that does not rely on cross-validation. While PLS and Ridge Regression are slower than

Table 6: Comparison of the investigated methods

	lasso	adaptive lasso	PLS	Ridge Regression	shrinkage
properties					
testing needed	no	no	yes	yes	yes
run-time	high	very high	medium	medium	low
simulation results					
density	very dense	between lasso & non-sparse methods	very sparse	very sparse	very sparse
mean-squared-error	low	low	high	medium	medium
real-world-study					
density	too dense on t.cell	too dense on t.cell	very dense	-	-
repeated measurements	too dense	too dense	dense	-	-
stability under resampling	medium	medium	medium	medium	good

shrinkage, both of them are fairly fast to compute, as they allow a kernel representation, i.e. most of the computation scales in the number of samples and not in the number of variables.

Available Software

The regularized estimation of partial correlations and the construction of gene association networks with (adaptive) Lasso, ridge regression and PLS are implemented in the R package parcor[37] which is available from the CRAN repository http://cran.r-project.org/. The package relies heavily on the lars package [40]. For assigning statistical significane to the edges in the network, we use the fdrtool package [54]. We also provide an executable sheet for the simulations (additional file 13) and the real-world data (additional file 14).

Authors' contributions

NK and ALB initiated the study. NK wrote the initial version of the manuscript. JS and NK implemented the R package, NK and ALB performed the analyses. All authors contributed to the concept and to the manuscript.

Additional material

Additional file 1
Cluster structure. The figure illustrates the two different cluster structure (denoted by "clusters" and "stars") that is used in the simulation study.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-10-384-S1.PDF]

Additional file 2
Histogram of partial correlations. The figure displays the histogram of the non-zero partial correlations in the simulation study for different density levels.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-10-384-S2.PDF]

Additional file 3
ROC-curves for a density of 5%, part I. The figures display the ROC-curves for a density of 5%, and for n = 25; 50; 75; 100
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-10-384-S3.PDF]

Additional file 4
ROC-curves for a density of 10%, part I. The figures display the ROC-curves for a density of 10%, and for n = 25; 50; 75; 100
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-10-384-S4.PDF]

Additional file 5
ROC-curves for a density of 15%, part I. The figures display the ROC-curves for a density of 15%, and for n = 25; 50; 75; 100
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-10-384-S5.PDF]

Additional file 6
ROC-curves for a density of 20%, part I. The figures display the ROC-curves for a density of 20%, and for n = 25; 50; 75; 100
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-10-384-S6.PDF]

Additional file 7
ROC-curves for a density of 25%, part I. The figures display the ROC-curves for a density of 25%, and for n = 25; 50; 75; 100
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-10-384-S7.PDF]

Additional file 8
ROC-curves for a density of 5%, part II. The figures display the ROC-curves for a density of 5%, and for n = 125; 150; 175; 200
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-10-384-S8.PDF]

Additional file 9
ROC-curves for a density of 10%, part II. The figures display the ROC-curves for a density of 10%, and for n = 125; 150; 175; 200
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-10-384-S9.PDF]

Additional file 10
ROC-curves for a density of 15%, part II. The figures display the ROC-curves for a density of 15%, and for n = 125; 150; 175; 200
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-10-384-S10.PDF]

Additional file 11
ROC-curves for a density of 20%, part II. The figures display the ROC-curves for a density of 20%, and for n = 125; 150; 175; 200
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-10-384-S11.PDF]

Additional file 12
ROC-curves for a density of 25%, part II. The figures display the ROC-curves for a density of 25%, and for n = 125; 150; 175; 200
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-10-384-S12.PDF]

Additional file 13
Simulation. This file contains the R-script to run the simulations that are described in the paper.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-10-384-S13.R]

Additional file 14

Real-world. This file contains the R-script to run the real-world studies that are described in the paper.

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2105-10-384-S14.R]

References

1. Friedman N: **Inferring Cellular Networks using Probabilistic Graphical Models.** *Science* 2004, **303(5659):**799–805.
2. Yeung MKS, Tegnér J and Collins JJ: **Reverse Engineering Gene Networks using Singular Value Decomposition and Robust Regression.** *Proceedings of the National Academy of Sciences* 2002, **99 (9):**6163–6168.
3. Rangel C, Angus J, Ghahramani Z, Lioumi M, Sotheran E, Gaiba A, Wild D and Falciani F: **Modeling T-cell Activation using Gene Expression Profiling and State-Space Models.** *Bioinformatics* 2004, **20:**1361–1372.
4. Whittaker J: *Graphical Models in Applied Multivariate Statistics.* Wiley New York; 1990.
5. Dobra A, Hans C, Jones B, Nevins J, Yao G and West M: **Sparse Graphical Models for Exploring Gene Expression Data.** *Journal of Multivariate Analysis* 2004, **90:**196–212.
6. Schäfer J and Strimmer K: **A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics.** *Statistical Applications in Genetics and Molecular Biology* 2005, **4:**32.
7. Schäfer J and Strimmer K: **An Empirical Bayes Approach to Inferring Large-Scale Gene Association Networks.** *Bioinformatics* 2005, **21:**754–764.
8. Li H and Gui J: **Gradient Directed Regularization for Sparse Gaussian Concentration Graphs, with Applications to Inference of Genetic Networks.** *Biostatistics* 2008, **7(2):**302–317.
9. Yuan M and Lin Y: **Model Selection and Estimation in the Gaussian Graphical Model.** *Biometrika* 2007, **94:**19–35.
10. Pihur V, Datta S and Datta S: **Reconstruction of Genetic Association Networks from Microarray Data.** *Bioinformatics* 2008, **24(4):**561–568.
11. Ma S, Gong Q and Bohnert HJ: **An Arabidopsis Gene Network Based on the Graphical Gaussian Model.** *Genome Research* 2007, **17:**1614–1625.
12. Schmitt JE, Lenroot RK, Wallace GL, Ordaz S, Taylor KN, Kabani N, Greenstein D, Lerch JP, Kendler KS, Neale MC and Giedd JN: **Identification of Genetically Mediated Cortical Networks: A Multivariate Study of Pediatric Twins and Siblings.** *Cerebral Cortex* 2008, **18(8):**1737–1747.
13. Efron B: **Large-Scale Simultaneous Hypothesis Testing: the Choice of a Null Hypothesis.** *Journal of the American Statistical Association* 2004, **99:**96–104.
14. Strimmer K: **A Unified Approach to False Discovery Rate Estimation.** *BMC Bioinformatics* 2008, **9:**303.
15. Breiman L: **Bagging predictors.** *Machine Learning* 1996, **24:**123–140.
16. Tyekucheva S and Chiaromonte F: **Augmenting the Bootstrap to Analyze High Dimensional Genomic Data.** *TEST* 2008, **17:**1–18.
17. Strimmer K: **Comments on: Augmenting the Bootstrap to Analyze High Dimensional Genomic Data.** *TEST* 2008, **17:**25–27.
18. Schäfer J: **Comments on: Augmenting the Bootstrap to Analyze High Dimensional Genomic Data.** *TEST* 2008, **17:**28–30.
19. d'Aspremont A, Banerjee O and Ghaoui LE: **First-Order Methods for Sparse Covariance Selection.** *SIAM Journal on Matrix Analysis and its Applications* 2008, **30:**56–66.
20. Rothman A, Bickel P, Levina E and Zhu J: **Sparse Permutation Invariant Covariance Estimation.** *Electronic Journal of Statistics* 2008, **2:**494–515.
21. Witten D and Tibshirani R: **Covariance-regularized regression and classification for high-dimensional problems.** *Journal of Royal Statistical Society, Series B* 2009, **71(3):**615–636.
22. Yuan M: **Efficient Computation of ℓ_1 Regularized Estimates in Gaussian Graphical Models.** *Journal of Computational and Graphical Statistics* 2008, **17(4):**809–826.
23. Tenenhaus A, Guillemot V, Gidrol X and Frouin V: **Gene Association Networks from Microarray Data using a Regularized Estimation of Partial Correlation based on PLS Regression.** *IEEE Transactions on Computational Biology and Bioinformatics* 2008 http://doi.ieeecomputersociety.org/10.1109/TCBB.2008.87.
24. Wold H: **Path Models with Latent Variables: The NIPALS Approach.** *Quantitative Sociology: International Perspectives on Mathematical and Statistical Model Building.* Academic Press: HMB et al 1975, 307–357.
25. Wold S, Ruhe H, Wold H and Dunn WJ III: **The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses.** *SIAM Journal of Scientific and Statistical Computations* 1984, **5:**735–743.
26. Saigo H, Krämer N and Tsuda K: **Partial Least Squares Regression for Graph Mining.** *14th International Conference on Knowledge Discovery and Data Mining (KDD2008)* 2008, 578–586.
27. Boulesteix AL and Strimmer K: **Partial Least Squares: a Versatile Tool for the Analysis of High-Dimensional Genomic Data.** *Briefings in Bioinformatics* 2007, **8:**32–44.
28. Rosipal R and Trejo L: **Kernel Partial Least Squares Regression in Reproducing Kernel Hilbert Spaces.** *Journal of Machine Learning Research* 2001, **2:**97–123.
29. Rosipal R and Krämer N: **Overview and Recent Advances in Partial Least Squares.** *Subspace, Latent Structure and Feature Selection Techniques, Lecture Notes in Computer Science.* Springer; 2006, 34–51.
30. Krämer N and Braun ML: **Kernelizing PLS, Degrees of Freedom, and Efficient Model Selection.** *Proceedings of the 24th International Conference on Machine Learning.* Ghahramani Z 2007, 441–448.
31. Chun H and Keles S: **Sparse partial least squares for simultaneous dimension reduction and variable selection.** *Journal of the Royal Statistical Society* 2009, **182(1):**79–90.
32. Hoerl A and Kennard R: **Ridge Regression: Biased Estimation for Nonorthogonal Problems.** *Technometrics* 2000, **42:**80–86.
33. Meinshausen N and Bühlmann P: **High Dimensional Graphs and Variable Selection with the Lasso.** *Annals of Statistics* 2006, **34 (3):**1436–1462.
34. Tibshirani R: **Regression Shrinkage and Selection via the Lasso.** *Journal of the Royal Statistical Society, Series B* 1996, **58:**267–288.
35. Zhou S, Geer van de S and Bühlmann P: **Adaptive Lasso for High Dimensional Regression and Gaussian Graphical Modeling.** 2009 in press, arXiv:0903.2515v1.
36. Zou H: **The Adaptive Lasso and its Oracle Properties.** *Journal of the American Statistical Association* 2006, **101(476):**1418–1429.
37. Krämer N and Schäfer J: *parcor: estimation of partial correlations based on regularized regression.* 2009, [R package version 0.1].
38. Schäfer J, Opgen-Rhein R and Strimmer K: **Reverse Engineering Genetic Networks using the GeneNet Package.** *R News* 2006, **5/6:**50–53.
39. Ruschhaupt M: **Erzeugung von positiv definiten Matrizen mit Nebenbedingungen zur Validierung von Netzwerkalgorithmen für Microarray-Daten.** *PhD thesis* University of Munich; 2008.
40. Hastie T and Efron B: *lars: Least Angle Regression, Lasso and Forward Stagewise.* 2007, [R package version 0.9-7].
41. Krämer N: **On the Peaking Phenomenon in Model Selection for the Lasso.** 2009 in press, http://arxiv.org/abs/0904.4416.
42. Kao K, Yang Y, Boscolo R, Sabatti C, Roychowdhury V and Liao J: **Transcriptome-based Determination of Multiple Transcription Regulator Activities in Escherichia Coli by Using Network Component Analysis.** *Proceedings of the National Academy of Sciences* 2004, **101(2):**641–646.

43. Schmidt-Heck W, Guthke R, Toepfer S, Reischer H, Duerrschmid K and Bayer K: **Reverse engineering of the stress response during expression of a recombinant protein.** *EUNITE 2004 European Symposium on Intelligent Technologies, Hybrid Systems and their Implementation on Smart Adaptive Systems.* 2004, 407–441.

44. Smith S, Fulton D, Chia T, Thorneycroft D, Chapple A, Dunstan H, Hylton C, Zeeman S and Smith A: **Diurnal Changes in the Transcriptome Encoding Enzymes of Starch Metabolism Provide Evidence for Both Transcriptional and Posttranscriptional Regulation of Starch Metabolism in Arabidopsis Leaves 1.** *Plant Physiology* 2004, **136:**2687–2699.

45. West M, Blanchette C, Dressman H, Huang E, Ishida S, Spang R, Zuzan H, Olson J Jr, Marks J and Nevins J: **Predicting the Clinical Status of Human Breast Cancer by using Gene Expression Profiles.** *Proceedings of the National Academy of Sciences* 2001, **98 (2):**11462–11467.

46. Boulesteix AL and Slawski M: **Stability and aggregation of ranked gene lists.** *Briefings in Bioinformatics* 2009, **10(5):**556–68.

47. Saeys Y, Inza I and Larranaga P: **A review of feature selection techniques in bioinformatics.** *Bioinformatics* 2007, **23(19):**2507.

48. Scutari M: **Structure variability in Bayesian networks.** 2009 in press, http://arxiv.org/abs/0909.1685.

49. Fleiss J: **Measuring nominal scale agreement among many raters.** *Psychological Bulletin* 1971, **76(5):**378–382.

50. Boulesteix AL, Kondylis A and Krämer N: **Comment on: Augmenting the bootstrap to analyze high dimensional genomic data.** *TEST* 2008, **17:**31–35.

51. Reverter A and Chan E: **Combining Partial Correlation and an Information Theory Approach to the Reversed-engineering of Gene Co-expression Networks.** *Bioinformatics* 2008, **24 (21):**2491–2497.

52. Pellet JP and Elisseeff A: **A Partial Correlation-Based Algorithm for Causal Structure Discovery with Continuous Variables.** *Advances in Intelligent Data Analysis VII, 7th International Symposium on Intelligent Data Analysis.* Berthold MR, Shawe-Taylor J, Lavrac N 2007, 229–239.

53. Arnold A, Liu Y and Abe N: **Temporal Causal Modeling with Graphical Granger Methods.** *Proceedings of the Thirteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM; 2007, 66–75.

54. Strimmer K: **fdrtool: a versatile R package for estimating local and tail area-based false discovery rates.** *Bioinformatics* 2008, **24:**1461–1462.

55. Boulesteix AL, Lambert-Lacroix S, Peyre J and Strimmer K: *plsgenomics: PLS analyses for genomics.* 2007, [R package version 1.2-2].

56. Opgen-Rhein R and Strimmer K: *longitudinal: Analysis of Multiple Time Course Data.* 2008, [R package version 1.1.4].