LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

INSTITUT FÜR STATISTIK

Nicole Krämer, Juliane Schäfer, Anne-Laure Boulesteix

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

Technical Report Number 057, 2009
Department of Statistics
University of Munich

http://www.stat.uni-muenchen.de

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

Nicole Krämer* Juliane Schäfer†‡ Anne-Laure Boulesteix§¶

May 6, 2009

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.

Conclusions: 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. 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. 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. Furthermore, for data that violates 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.

*Machine Learning Group, Berlin Institute of Technology, nkraemer@cs.tu-berlin.de

†Basel Institute for Clinical Epidemiology & Biostatistics, University Hospital Basel, jschaefer@uhbs.ch

‡Seminar for Statistics, ETH Zurich

§Department of Medical Informatics, Biometry and Epidemiology, University of Munich, boulesteix@ibe.med.uni-muenchen.de

¶Department of Statistics, University of Munich

1 Introduction

Besides Bayesian networks [11], auto-regressive models [48], and state-space models [24], graphical Gaussian models (GGMs) are a popular method for modeling genetic networks based on microarray transcriptome data. In the GGM methodology [44], 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 [8, 10, 18, 23, 32, 33, 50]. These approaches are used in numerous applied studies, e.g., for estimating *Arabidopsis* gene networks [19] or for the study of genetically mediated cortical networks [35].

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. 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.

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) [44] 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 $\mathcal{Z} = \{\mathsf{Z}_1, \ldots, \mathsf{Z}_k\}$, the orthogonal complement of X with respect to $\mathcal{Z}$ is

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

where the projection $\mathcal{P}_{\mathcal{Z}}$ is defined with respect to the inner product $\langle \mathsf{X}_1, \mathsf{X}_2 \rangle = E[\mathsf{X}_1 \mathsf{X}_2]$ between two random variables X_1 and X_2. Here, we tacitly assume that all involved moments exist. The partial correlation $\rho_{\mathcal{Z}}(\mathsf{X}_1, \mathsf{X}_2)$ between X_1 and X_2 with respect to $\mathcal{Z}$ is the correlation of the orthogonal complements of X_1 and X_2 with respect to $\mathcal{Z}$:

$$\rho_{\mathcal{Z}}(\mathsf{X}_1, \mathsf{X}_2) \quad = \quad \mathrm{cor}\left(\mathsf{X}_{1\setminus\mathcal{Z}}, \mathsf{X}_{2\setminus\mathcal{Z}}\right). \tag{1}$$

In the context of gene regulatory networks, each of the p genes is represented by a random variable X_i $(i = 1, \ldots, p)$. 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 $\mathcal{Z}_{\setminus ij} = \{\mathsf{X}_1, \ldots, \mathsf{X}_p\} \setminus \{\mathsf{X}_i, \mathsf{X}_j\}$.

Given n observations (arrays) $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^p$ 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 [44], as outlined below.

Notations

In the rest of this article,

$$\boldsymbol{X} \;=\; (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\top \in \mathbb{R}^{n \times p} \tag{2}$$

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 $\boldsymbol{\Sigma}$ is then given as

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

Formulation 1: Inversion of the Covariance Matrix

If the estimate $\widehat{\boldsymbol{\Sigma}}$ is invertible, an unbiased estimate of the partial correlation between genes i and j is obtained as

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

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

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

Formulation 2: Least Squares Regression

Let us consider the p linear regression models

$$\mathsf{X}_i \;=\; \sum_{j \neq i} \beta_j^{(i)} \mathsf{X}_j + \varepsilon\,, \text{ for } i = 1, \ldots, p, \tag{4}$$

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 $i = 1, \ldots, p$, the least squares estimate $\widehat{\boldsymbol{\beta}}^{(i)} = (\beta_1^{(i)}, \ldots, \beta_{i-1}^{(i)}, \beta_{i+1}^{(i)}, \ldots, \beta_p^{(i)})^\top$ of the vector of regression coefficients is the solution of the optimization problem

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

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

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

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

In the $n > p$ setting, the two regression coefficients $\beta_j^{(i)}$ and $\beta_i^{(j)}$ always have the same sign. Hence, $\sqrt{\widehat{\beta}_j^{(i)} \widehat{\beta}_i^{(j)}}$ is well-defined. Moreover, it can be shown that both formulations 1 and 2 are equivalent [44] 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 $\widehat{\rho}_{ij}$.

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. First, 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 $\widehat{\boldsymbol{\Sigma}}$ is typically ill-conditioned or even singular, and its generalized (Moore-Penrose) inverse has large mean squared error [32]. 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.

Second, the testing approach described above breaks down in the $p \gg n$ setting, since the sampling distribution of estimates $\widehat{\rho}_{ij}$ under the null hypothesis of zero partial correlation will be 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 [9, 33, 37].

2 Methods

This section reviews the available strategies for estimating GGMs in the $p \gg n$ setting: biased large covariance estimation and 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 [32] adopt this approach and propose a ridge-type shrinkage estimator of the covariance matrix. This shrinkage estimator is constructed as a convex combination of the unrestricted sample covariance matrix $\widehat{\boldsymbol{\Sigma}}$ and an estimator $\widehat{\boldsymbol{T}}$ of a specified low-dimensional sub-model $\boldsymbol{T}$:

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

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 [9, 33, 37].

Estimates of the inverse covariance matrix can also be obtained using bootstrap aggregating (bagging) as a technique for variance reduction [5]. In some implicit way, the bootstrap procedure presumably helps to regularize the problem. However, bagging schemes are inferior to the shrinkage estimator [32], as computationally much more expensive. A recent extension using the augmented bootstrap [42] is in fact closely related to the shrinkage estimator [30,38] 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 [7, 10, 28, 45, 50]. 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 [49].

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* $\widehat{\boldsymbol{\beta}}^{(i)}_{\text{reg}}$ *of the linear regression model (4), we define the corresponding estimate of the partial correlations as*

$$\widehat{\rho}_{ij,\text{reg}} \;=\; \text{sign}\left(\widehat{\beta}^{(i)}_{j,\text{reg}}\right) \min\left\{1, \sqrt{\widehat{\beta}^{(i)}_{j,\text{reg}}\widehat{\beta}^{(j)}_{i,\text{reg}}}\right\} \tag{8}$$

if

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

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. [40] suggest Partial Least Squares (PLS) regression [46, 47] 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 [3, 29]. The main idea of PLS is to build a few orthogonal components from the original data $\boldsymbol{X}^{(\setminus i)}$ and to use them as predictors in a least squares fit. A PLS component $\boldsymbol{t} = \boldsymbol{X}^{(\setminus i)}\boldsymbol{w}$ is a linear combination of the original predictors that have maximal covariance with the response vector $\boldsymbol{X}^{(i)}$, under the additional assumption that the components are mutually orthogonal. Formally, the k-th PLS component is defined by

$$\begin{aligned} \boldsymbol{w}_k &= \arg\max_{\|\boldsymbol{w}\|=1} \operatorname{cov}\left(\boldsymbol{X}^{(\setminus i)}\boldsymbol{w}, \boldsymbol{X}^{(i)}\right)^2 \\ \text{s.t.} \qquad & \boldsymbol{w}^\top \boldsymbol{X}^{(\setminus i)\top}\boldsymbol{X}^{(\setminus i)}\boldsymbol{w}_l = 0 \quad \text{for } l < k\,. \end{aligned}$$

Hence, PLS regularizes the regression problem by compressing the p variables into a small number m of orthogonal components $\boldsymbol{T} = (\boldsymbol{t}_1, \ldots, \boldsymbol{t}_m)$ and regressing the response variable onto these components. After rescaling the weight vectors $\boldsymbol{w}_k$ $(k = 1, \ldots, m)$ such that $\boldsymbol{t}_k$ has length 1, this leads to the regression coefficients

$$\widehat{\boldsymbol{\beta}}^{(i)}_{\text{pls}} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m)\,\boldsymbol{T}^\top \boldsymbol{X}^{(i)}\,.$$

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 [26, 27]. 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 [16]. In the context of gene regulatory networks, Tenenhaus et.al. [40] 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 [6].) 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 exhaustiveness, let us mention in this section a variant of the PLS approach described above, which was recently suggested by Pihur et al. [23]. 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 $\widehat{\beta}^{(i)}_{j,pls}$ and $\widehat{\beta}^{(j)}_{i,pls}$ 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. [13]) 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

$$P(\boldsymbol{\beta}) = \lambda \|\boldsymbol{\beta}\|_2^2 = \lambda \sum_i \beta_i^2, \tag{9}$$

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 [37] 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 [20] propose to estimate the regression coefficients in Def. 1 with the Lasso [41] 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

$$P(\boldsymbol{\beta}) = \lambda \|\boldsymbol{\beta}\|_1 = \lambda \sum_i |\beta_i|, \tag{10}$$

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 $\widehat{\rho}_{ij,\text{lasso}} \neq 0$. 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 [20] 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 [32] 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. [51] show that the two-stage adaptive Lasso procedure [52] is consistent for high-dimensional model selection in graphical Gaussian models under rather general and less restrictive conditions. The adaptive Lasso [52] considers the Lasso with penalty weights

$$P(\boldsymbol{\beta}) = \lambda \sum_i \widehat{w}_i \, |\beta_i| \,, \tag{11}$$

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

$$\widehat{w}_i = 1/|\widehat{\beta}_{i,\text{lasso}}|. \tag{12}$$

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 $\widehat{\rho}_{ij,\text{adaptive lasso}} \neq 0$. 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.

3 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 ridge-type 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" [17], 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 [32]. The number of variables is fixed at $p = 100$. Partial correlation matrices $\boldsymbol{P}$ of size $p \times p$ with 248 non-zero entries are randomly generated. Various sample sizes ranging from 25 to 200 in steps of 25 are investigated successively. 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 $\boldsymbol{P}$.

For each generated data set, $\boldsymbol{P}$ 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 [12], 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 [15] for details). For ridge regression, we consider a logarithmically spaced sequence $l_1, \ldots, l_{1000}$ ranging from 10^{-10} to 10^{-1}. The candidate penalty parameters are then defined as $\lambda_s = l_s\, n\, p$ (with $s = 1, \ldots, 1000$). 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 Figure 1a, the MSE is displayed as a function of the sample size n.

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. However, 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 of Figure 1a, while the horizontal line is the number of true edges. 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.

The two lower panels in Figure 1a correspond to the power (left) and the true discovery rate (tdr, right) which are defined as

$$\begin{aligned} \text{power} &= \frac{\#\{\text{true edges that are selected}\}}{\#\{\text{true edges}\}} \quad \text{and} \\ \text{tdr} &= \frac{\#\{\text{true edges that are selected}\}}{\#\{\text{selected edges}\}}, \end{aligned}$$

respectively. The panels illustrate that the Lasso's comparatively high power comes at the prize of rather low true discovery rate. 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 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 in Figure 1b. It can be seen that 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.

Real Data Study

We compare the five different methods on diverse real world data sets: the `ecoli1` [14] and `ecoli2` [34], `Ara` [36], `t.cell10` and `t.cell34` [24], and `west` [43] 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 and the type of interaction between genes.

Figures 2a and 2b 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 be at least partly 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. 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 some implicit way by methods using an fdr-based assessment, where the distribution under the null hypothesis is estimated from the data.

Among the three methods with fdr-based assessment of the edges, i.e PLS, ridge regression and the approach based on 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. The estimated graphs show a moderate to strong overlap between 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 3a and 3b 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. Very interestingly, Lasso and adaptive Lasso yield less genes with a high number of connected genes, especially for the `ecoli2`, `Arabidopsis` and `west` data sets. This indicates that the Lasso might be less adequate for identifying so-called "hubs". This may also explain that the Lasso and adaptive Lasso find far less edges than the three other methods in the `Arabidopsis` and `west` data sets (in contrast to the simulation study).

Figure 4 displays the percentage of positive 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.

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 Lasso $\approx$ 17 hours, and the adaptive Lasso $\approx$ 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 observations. Furthermore, adaptive Lasso requires nested cross-validation. Partial relief may be found in a parallel implementation. Alternatively, for high-dimensional data, one might therefore consider to approximate the Lasso-based networks by first constructing a mildly sparse network without cross-validation (for example using the method described in [20]), and then to refine this network by running the (adaptive) Lasso with cross-validation.

4 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. 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 have significance testing using false discovery rate multiple testing to identify unusual outliers among estimated partial correlations 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. 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 ridge-type 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 [2], 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 approach using a ridge-type covariance estimator 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 [25] 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 [1, 22]. For instance, in the presence of longitudinal data, Arnold. et.al. [1] 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.

5 Available Software

The regularized estimation of partial correlations and the construction of gene association networks with (adaptive) Lasso, ridge regression and PLS have been implemented in the `R` package `parcor` [17] which is available from the CRAN repository http://cran.r-project.org/. The package relies heavily on the `lars` package [12]. For assigning statistical significance to the edges in the network, we used the `fdrtool` package [39]. An executable sheet for the simulations can be downloaded[1].

Authors' contribution

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

[1] http://ml.cs.tu-berlin.de/~nkraemer/software.html

concept and to the manuscript.

Acknowledgments

This work was supported in part by the BMBF grant FKZ 01-IS07007A (ReMind), and the FP7-ICT Programme of the European Community, under the PASCAL2 Network of Excellence, ICT-216886. Financial support from DSM Nutritional Products Ltd. (Basel, Switzerland) is gratefully acknowledged. We thank Lukas Meier and Mikio L. Braun for constructive comments on model selection, and Animesh Acharjee for helpful feedback on the R package "parcor".

References

[1] A. Arnold, Y. Liu, and N. Abe, *Temporal Causal Modeling with Graphical Granger Methods*, Proceedings of the Thirteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,, 2007.

[2] A. L. Boulesteix, A. Kondylis, and N. Krämer, *Comment on: Augmenting the bootstrap to analyze high dimensional genomic data*, TEST **17** (2008), 31–35.

[3] A.-L. Boulesteix and K. Strimmer, *Partial Least Squares: a Versatile Tool for the Analysis of High-Dimensional Genomic Data*, Briefings in Bioinformatics **8** (2007), no. 1, 32–44.

[4] Anne-Laure Boulesteix, Sophie Lambert-Lacroix, Julie Peyre, and Korbinian Strimmer, *plsgenomics: Pls analyses for genomics*, 2007, R package version 1.2-2.

[5] L. Breiman, *Bagging predictors*, Machine Learning **24** (1996), 123–140.

[6] H. Chun and S. Keles, *Sparse partial least squares for simultaneous dimension reduction and variable selection.*, Journal of the Royal Statistical Society (2009), to appear.

[7] A. d'Aspremont, O. Banerjee, and L. El Ghaoui, *First-Order Methods for Sparse Covariance Selection*, SIAM Journal on Matrix Analysis and its Applications **30** (2008), no. 1, 56–66.

[8] A. Dobra, C. Hans, B. Jones, J.R. Nevins, G. Yao, and M. West, *Sparse Graphical Models for Exploring Gene Expression Data*, Journal of Multivariate Analysis **90** (2004), no. 1, 196–212.

[9] B. Efron, *Large-Scale Simultaneous Hypothesis Testing: the Choice of a Null Hypothesis*, Journal of the American Statistical Association **99** (2004), 96–104.

[10] J. Friedman, T. Hastie, and R. Tibshirani, *Sparse Inverse Covariance Estimation with the Graphical Lasso*, Biostatistics **9** (2008), no. 3, 432–441.

[11] N. Friedman, *Inferring Cellular Networks using Probabilistic Graphical Models*, Science **303** (2004), no. 5659, 799–805.

[12] Trevor Hastie and Brad Efron, *lars: Least angle regression, lasso and forward stagewise*, 2007, R package version 0.9-7.

[13] A.E. Hoerl and R.W. Kennard, *Ridge Regression: Biased Estimation for Nonorthogonal Problems*, Technometrics **42** (2000), no. 1, 80–86.

[14] K.C. Kao, Y.L. Yang, R. Boscolo, C. Sabatti, V. Roychowdhury, and J.C. Liao, *Transcriptome-based Determination of Multiple Transcription Regulator Activities in Escherichia Coli by Using Network Component Analysis*, Proceedings of the National Academy of Sciences **101** (2004), no. 2, 641–646.

[15] N. Krämer, *On the Peaking Phenomenon in Model Selection for the Lasso*, preprint, available at `http://arxiv.org`, 2009.

[16] N. Krämer and M. L. Braun, *Kernelizing PLS, Degrees of Freedom, and Efficient Model Selection*, Proceedings of the 24th International Conference on Machine Learning (Z. Ghahramani, ed.), 2007, pp. 441–448.

[17] Nicole Krämer and Juliane Schäfer, *parcor: estimation of partial correlations based on regularized regression*, 2009, R package version 0.1.

[18] H. Li and J. Gui, *Gradient Directed Regularization for Sparse Gaussian Concentration Graphs, with Applications to Inference of Genetic Networks*, Biostatistics **7** (2008), no. 2, 302–317.

[19] S. Ma, Q. Gong, and H. J. Bohnert, *An Arabidopsis Gene Network Based on the Graphical Gaussian Model*, Genome Research **17** (2007), 1614–1625.

[20] N. Meinshausen and P. Bühlmann, *High Dimensional Graphs and Variable Selection with the Lasso*, Annals of Statistics **34** (2006), no. 3, 1436–1462.

[21] Rainer Opgen-Rhein and Korbinian Strimmer., *longitudinal: Analysis of multiple time course data*, 2008, R package version 1.1.4.

[22] Jean-Philippe Pellet and André Elisseeff, *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 (M. R. Berthold, J. Shawe-Taylor, and N. Lavrac, eds.), 2007, pp. 229–239.

[23] V. Pihur, S. Datta, and S. Datta, *Reconstruction of Genetic Association Networks from Microarray Data*, Bioinformatics **24** (2008), no. 4, 561–568.

[24] C. Rangel, J. Angus, Z. Ghahramani, M. Lioumi, E. Sotheran, A. Gaiba, D.L. Wild, and F. Falciani, *Modeling T-cell Activation using Gene Expression Profiling and State-Space Models*, Bioinformatics **20** (2004), 1361–1372.

[25] A. Reverter and E.K.F Chan, *Combining Partial Correlation and an Information Theory Approach to the Reversed-engineering of Gene Co-expression Networks*, Bioinformatics (2008), doi:10.1093/bioinformatics/btn482.

[26] R. Rosipal and N. Krämer, *Overview and Recent Advances in Partial Least Squares*, Subspace, Latent Structure and Feature Selection Techniques, Lecture Notes in Computer Science, Springer, 2006, pp. 34–51.

[27] R. Rosipal and L.J. Trejo, *Kernel Partial Least Squares Regression in Reproducing Kernel Hilbert Spaces*, Journal of Machine Learning Research **2** (2001), 97–123.

[28] A. Rothman, P. Bickel, E. Levina, and J. Zhu, *Sparse Permutation Invariant Covariance Estimation*, Electronic Journal of Statistics **2** (2008), 494–515.

[29] H. Saigo, N. Krämer, and K. Tsuda, *Partial Least Squares Regression for Graph Mining*, 14th International Conference on Knowledge Discovery and Data Mining (KDD2008), 2008, pp. 578–586.

[30] J. Schäfer, *Comments on: Augmenting the Bootstrap to Analyze High Dimensional Genomic Data*, TEST **17** (2008), no. 1, 28–30.

[31] J. Schäfer, R. Opgen-Rhein, and K. Strimmer, *Reverse Engineering Genetic Networks using the GeneNet Package*, R News **5/6** (2006), 50–53.

[32] J. Schäfer and K. Strimmer, *A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics*, Statistical Applications in Genetics and Molecular Biology **4** (2005), 32.

[33] J. Schäfer and K. Strimmer, *An Empirical Bayes Approach to Inferring Large-Scale Gene Association Networks*, Bioinformatics **21** (2005), 754–764.

[34] W. Schmidt-Heck, R. Guthke, S. Toepfer, H. Reischer, K. Duerrschmid, and K. Bayer, *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, pp. 407–441.

[35] J. E. Schmitt, R. K. Lenroot, G. L. Wallace, S. Ordaz, K. N. Taylor, N. Kabani, D. Greenstein, J. P. Lerch, K. S. Kendler, M. C. Neale, and J. N. Giedd, *Identification of Genetically Mediated Cortical Networks: A Multivariate Study of Pediatric Twins and Siblings*, Cerebral Cortex **18** (2008), no. 8, 1737–1747.

[36] S.M. Smith, D.C. Fulton, T. Chia, D. Thorneycroft, A. Chapple, H. Dunstan, C. Hylton, S.C. Zeeman, and A.M. Smith, *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 **136** (2004), no. 1, 2687–2699.

[37] K. Strimmer, *A Unified Approach to False Discovery Rate Estimation*, BMC Bioinformatics **9** (2008), 303.

[38] _____, *Comments on: Augmenting the Bootstrap to Analyze High Dimensional Genomic Data*, TEST **17** (2008), no. 1, 25–27.

[39] _____, *fdrtool: a versatile R package for estimating local and tail area-based false discovery rates*, Bioinformatics **24** (2008), 1461–1462.

[40] A. Tenenhaus, V. Guillemot, X. Gidrol, and V. Frouin, *Gene Association Networks from Microarray Data using a Regularized Estimation of Partial Correlation based on PLS Regression*, IEEE Transactions on Computational Biology and Bioinformatics **to appear** (2008).

[41] R. Tibshirani, *Regression Shrinkage and Selection via the Lasso*, Journal of the Royal Statistical Society, Series B **58** (1996), no. 1, 267–288.

[42] S. Tyekucheva and F. Chiaromonte, *Augmenting the Bootstrap to Analyze High Dimensional Genomic Data*, TEST **17** (2008), no. 1, 1–18.

[43] M. West, C. Blanchette, H. Dressman, E. Huang, S. Ishida, R. Spang, H. Zuzan, J.A. Olson Jr, J.R. Marks, and J.R. Nevins, *Predicting the Clinical Status of Human Breast Cancer by using Gene Expression Profiles*, Proceedings of the National Academy of Sciences **98** (2001), no. 2, 11462–11467.

[44] J. Whittaker, *Graphical Models in Applied Multivariate Statistics*, Wiley New York, 1990.

[45] D.M. Witten and R. Tibshirani, *Covariance-regularized regression and and classification for high-dimensional problems*, Journal of Royal Statistical Society, Series B, to appear.

[46] H. Wold, *Path Models with Latent Variables: The NIPALS Approach*, Quantitative Sociology: International Perspectives on Mathematical and Statistical Model Building (H. M. Blalock et al., ed.), Academic Press, 1975, pp. 307–357.

[47] S. Wold, H. Ruhe, H. Wold, and W. J. Dunn, III, *The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses*, SIAM Journal of Scientific and Statistical Computations **5** (1984), 735–743.

[48] M. K. S. Yeung, J. Tegnér, and J .J. Collins, *Reverse Engineering Gene Networks using Singular Value Decomposition and Robust Regression*, Proceedings of the National Academy of Sciences **99** (2002), no. 9, 6163–6168.

[49] M. Yuan, *Efficient Computation of ℓ_1 Regularized Estimates in Gaussian Graphical Models*, Journal of Computational and Graphical Statistics **17** (2008), no. 4, 809–826.

[50] M. Yuan and Y. Lin, *Model Selection and Estimation in the Gaussian Graphical Model*, Biometrika **94** (2007), no. 1, 19–35.

[51] S. Zhou, S. van de Geer, and P. Bühlmann, *Adaptive lasso for high dimensional regression and gaussian graphical modeling*, Preprint, arXiv:0903.2515v1 (2009).

[52] H. Zou, *The Adaptive Lasso and its Oracle Properties*, Journal of the American Statistical Association **101** (2006), no. 476, 1418–1429.

Figures

Figure 1a - MSE, number of edges, power and TDR

Mean squared error, number of selected edges, power and true discovery rate (TDR) for the various methods PLS, ridge regression, the approach based on shrinkage covariance estimation, Lasso, and adaptive Lasso.

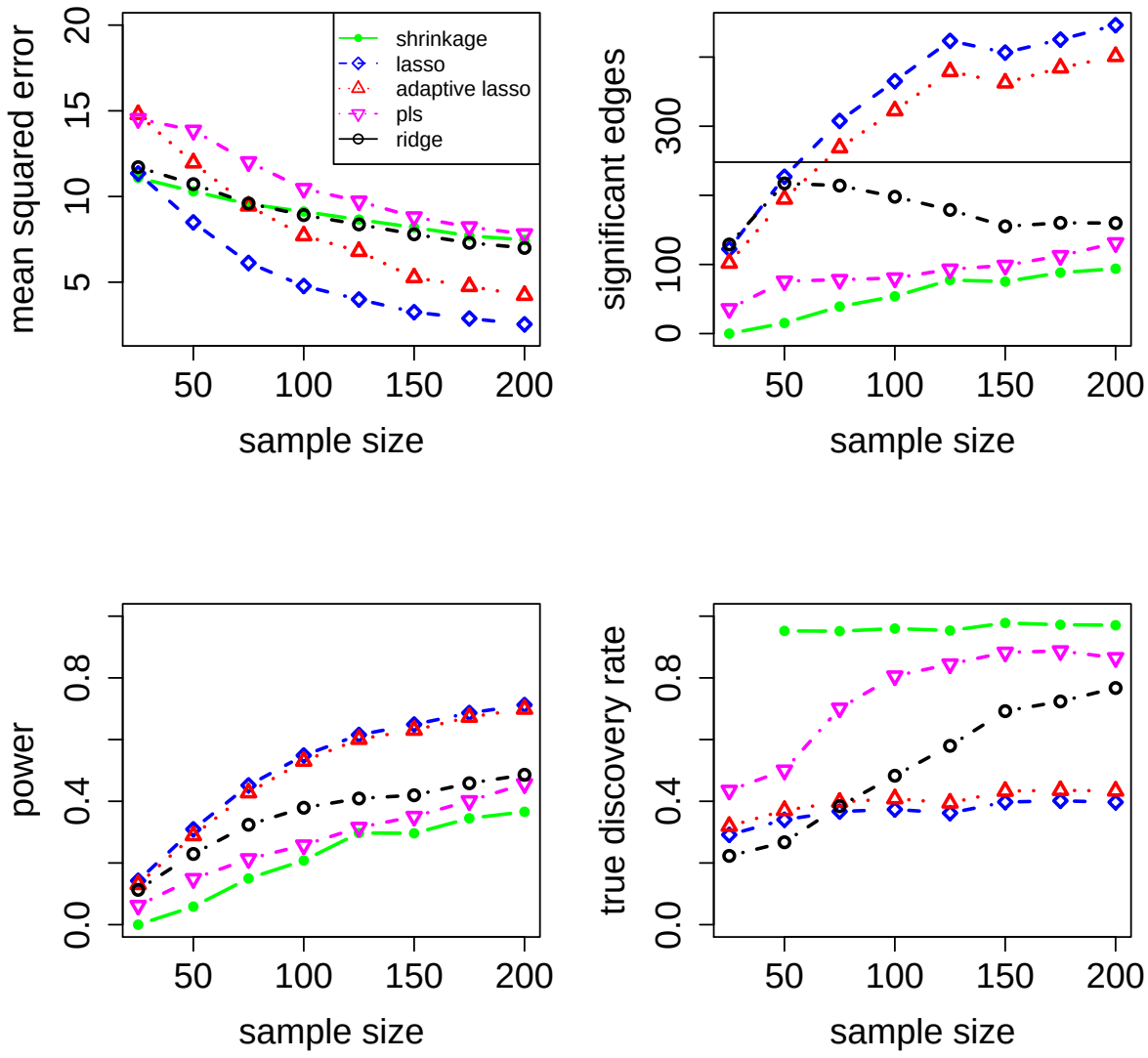


Figure 1b - ROC curves

ROC curves obtained by varying the fdr-threshold for PLS, Ridge Regression and Shrinkage. The sensitivity and specificity of Lasso and Adaptive Lasso are represented by a point.

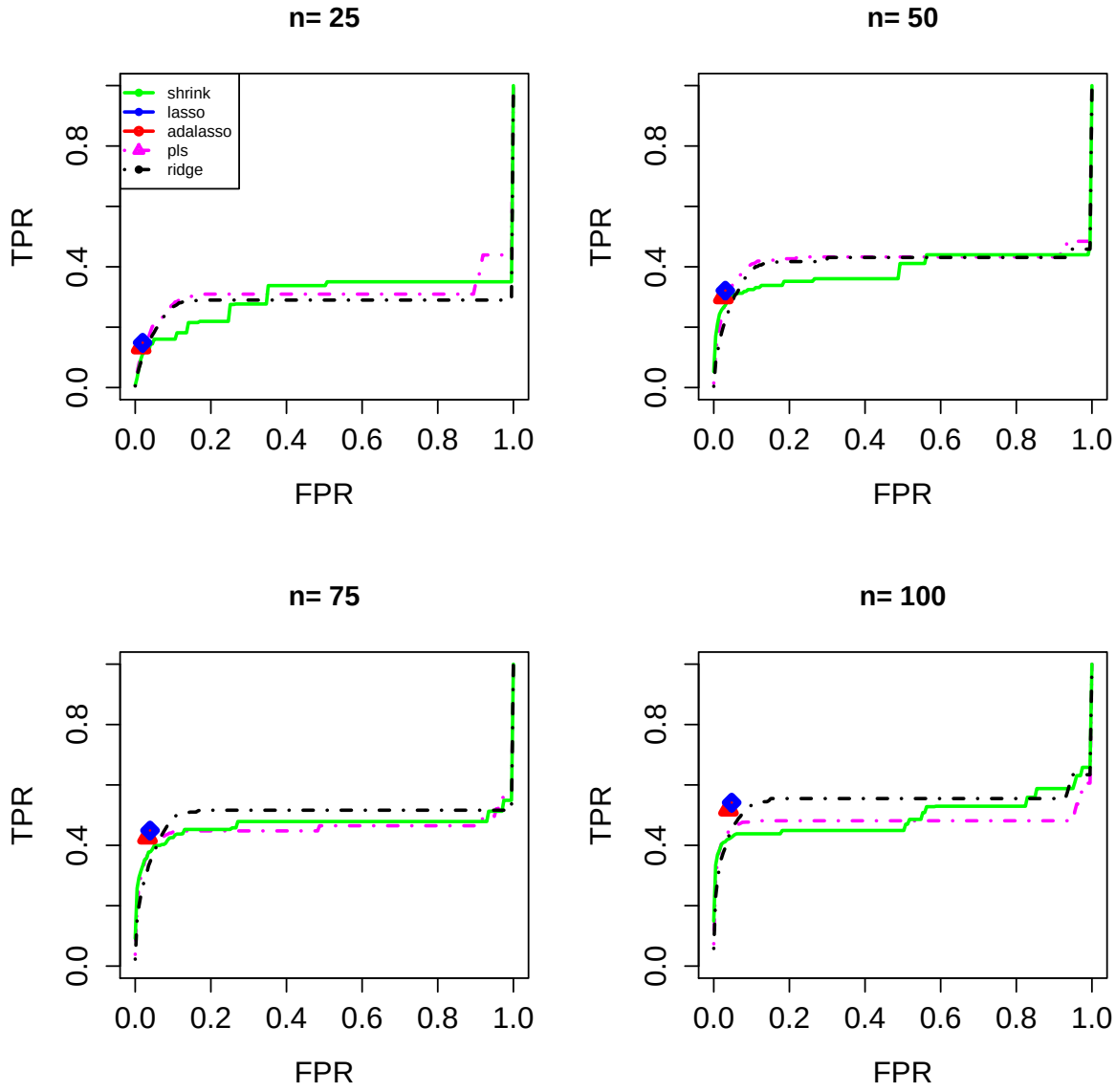


Figure 1b - ROC curves

ROC curves obtained by varying the fdr-threshold for PLS, Ridge Regression and Shrinkage. The sensitivity and specificity of Lasso and Adaptive Lasso are represented by a point.

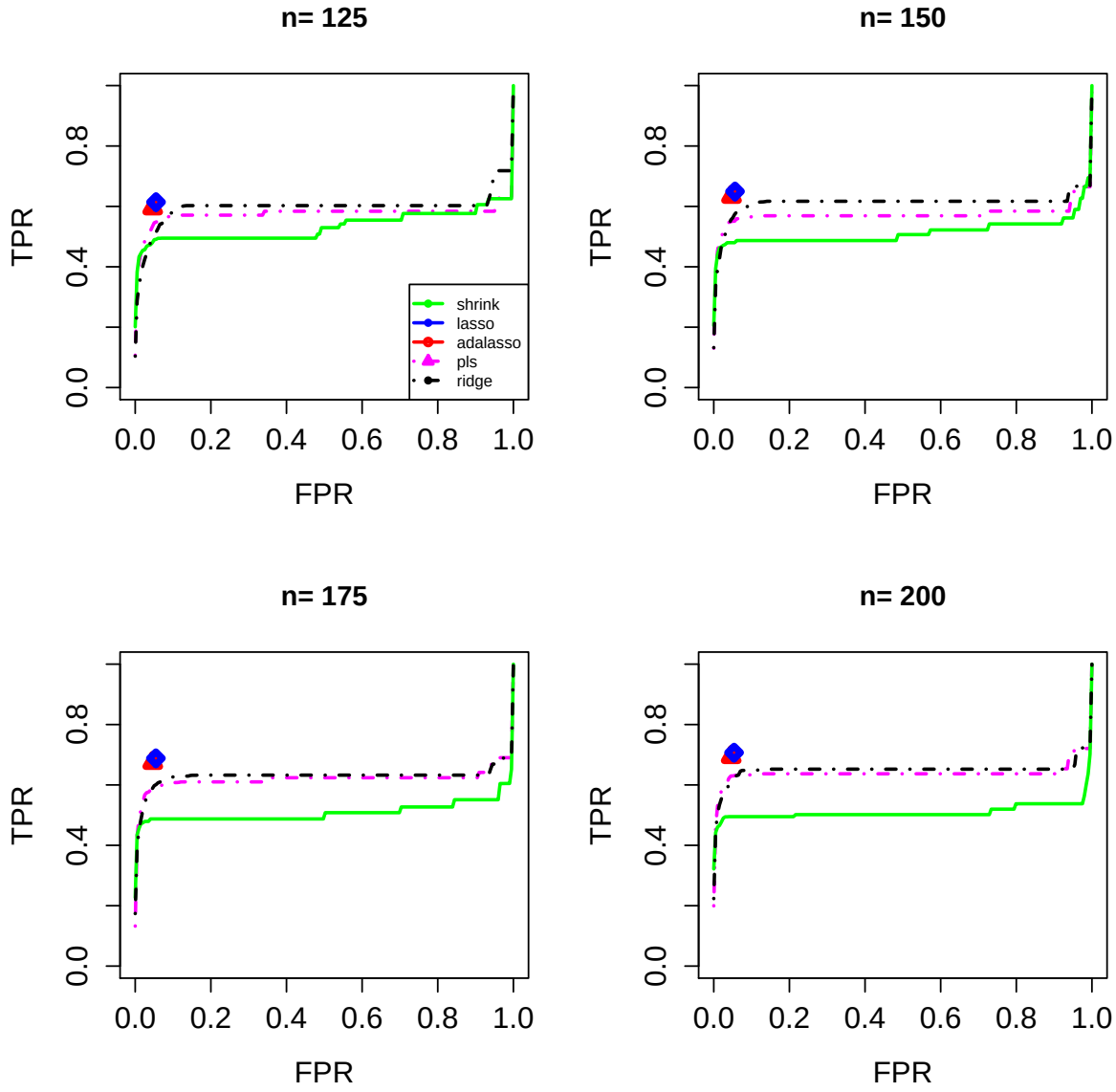


Figure 2a - Proportion of selected edges

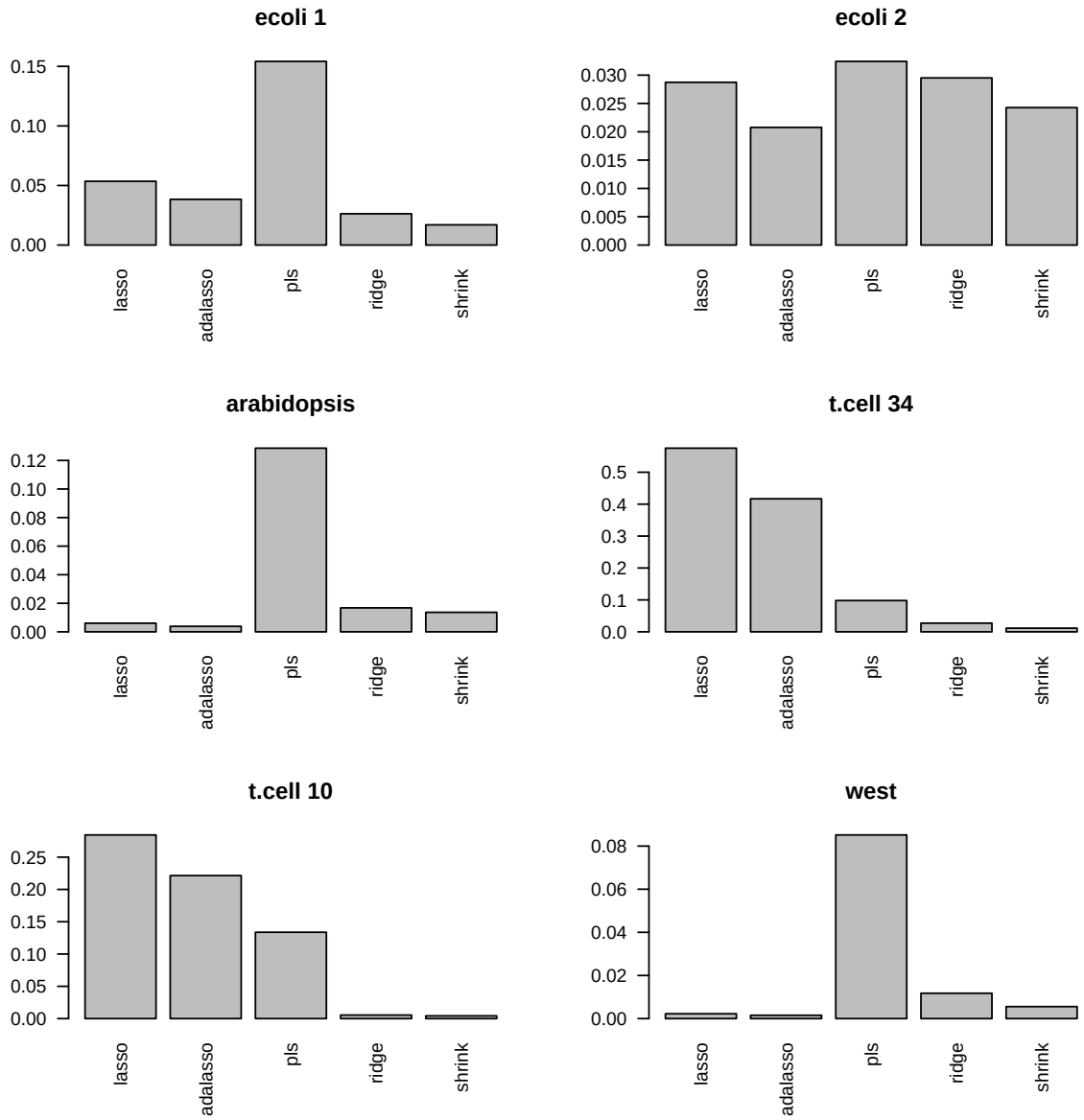


Figure 2b - Proportion of selected edges without PLS

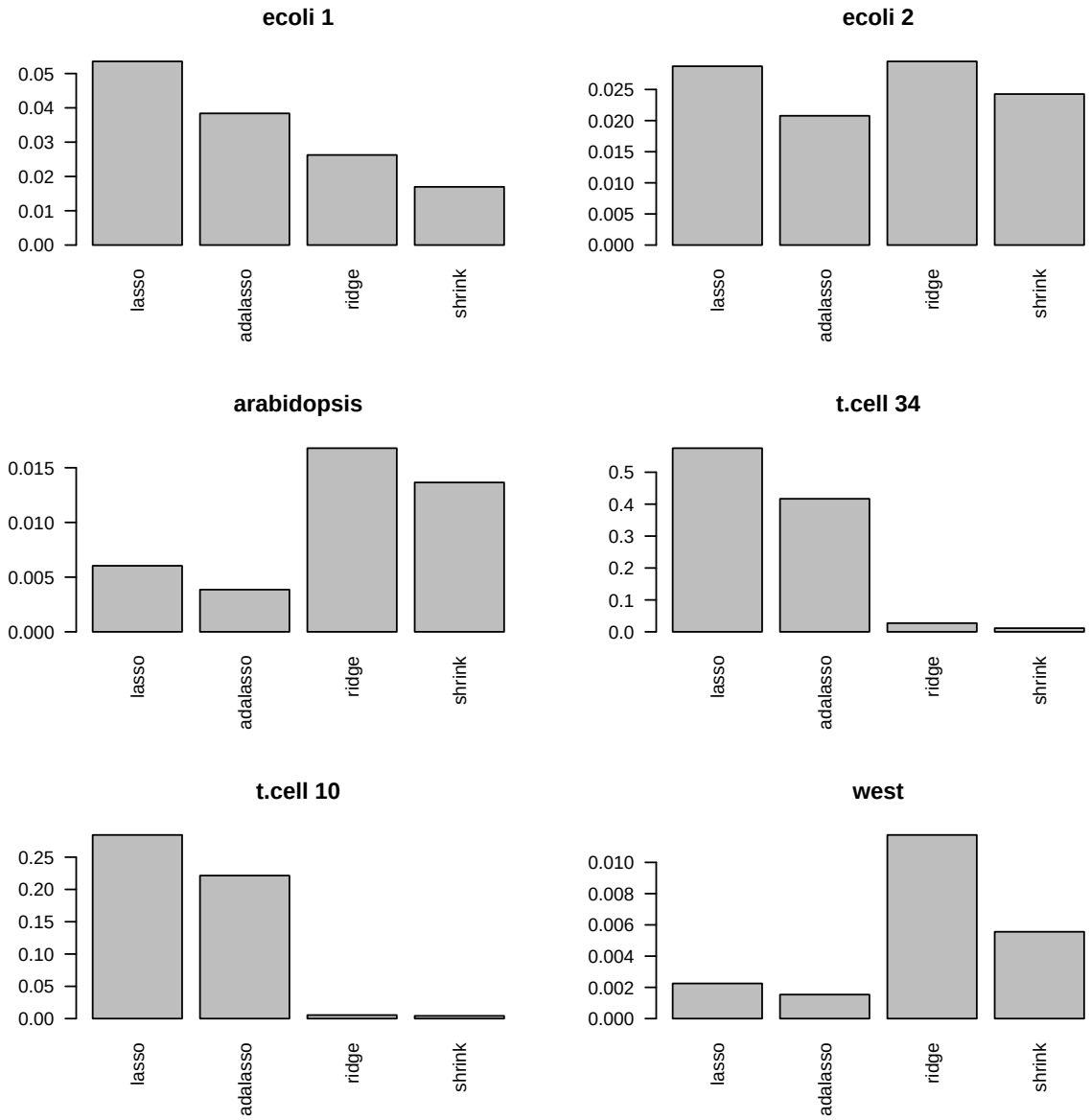


Figure 3a - Connectivity: Proportion of connected genes

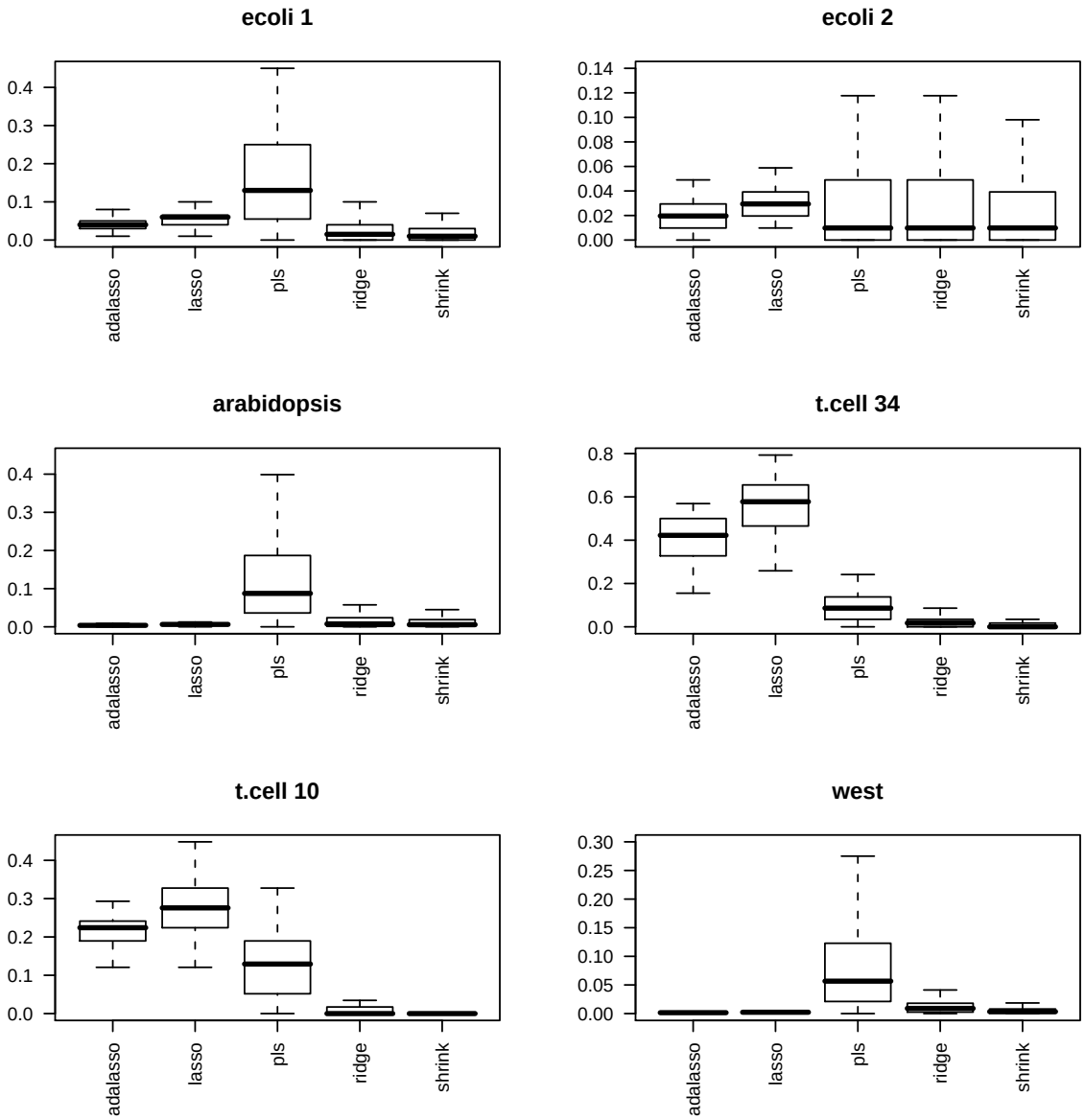


Figure 3b - Connectivity: Proportion of connected genes without PLS

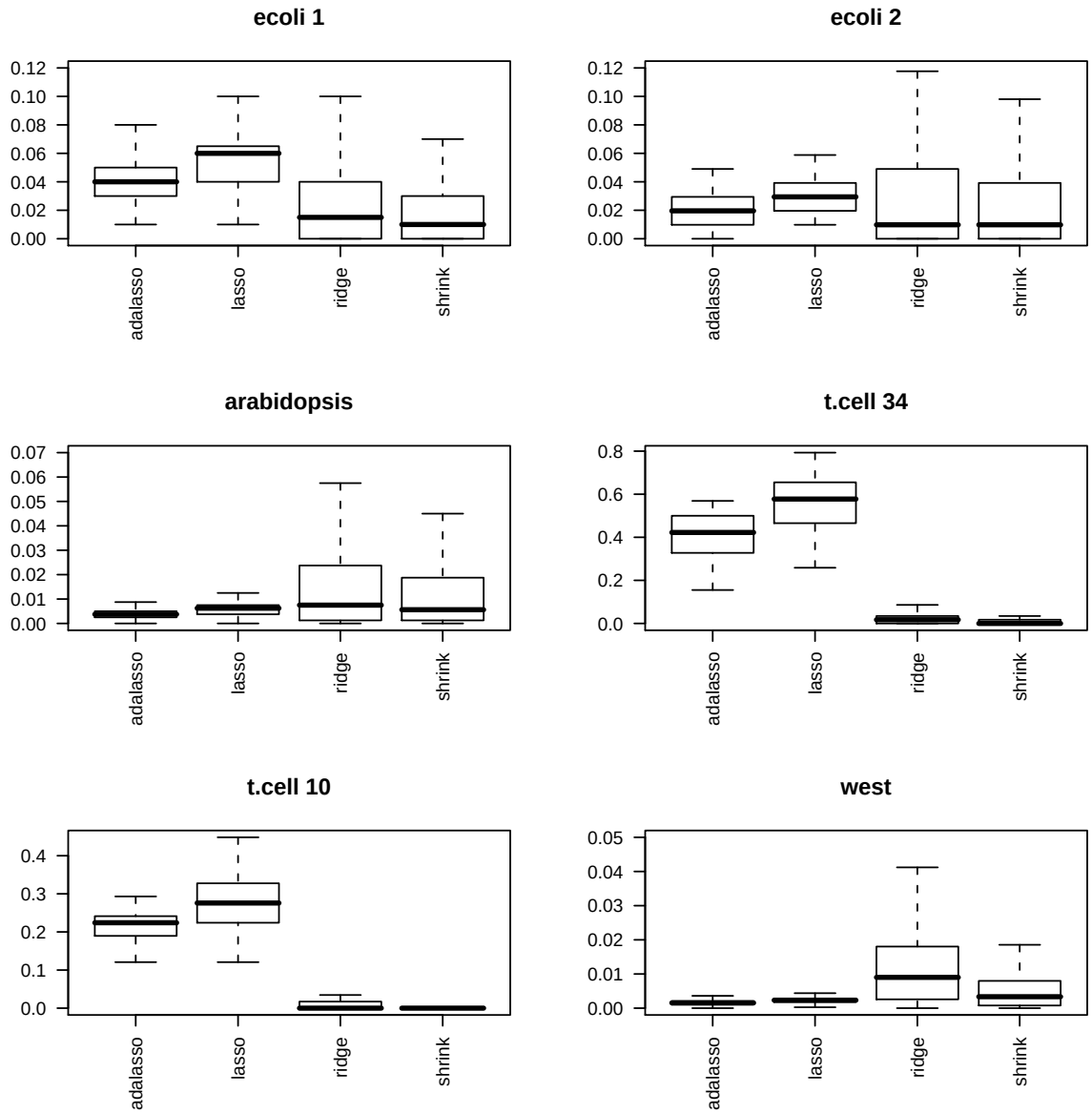
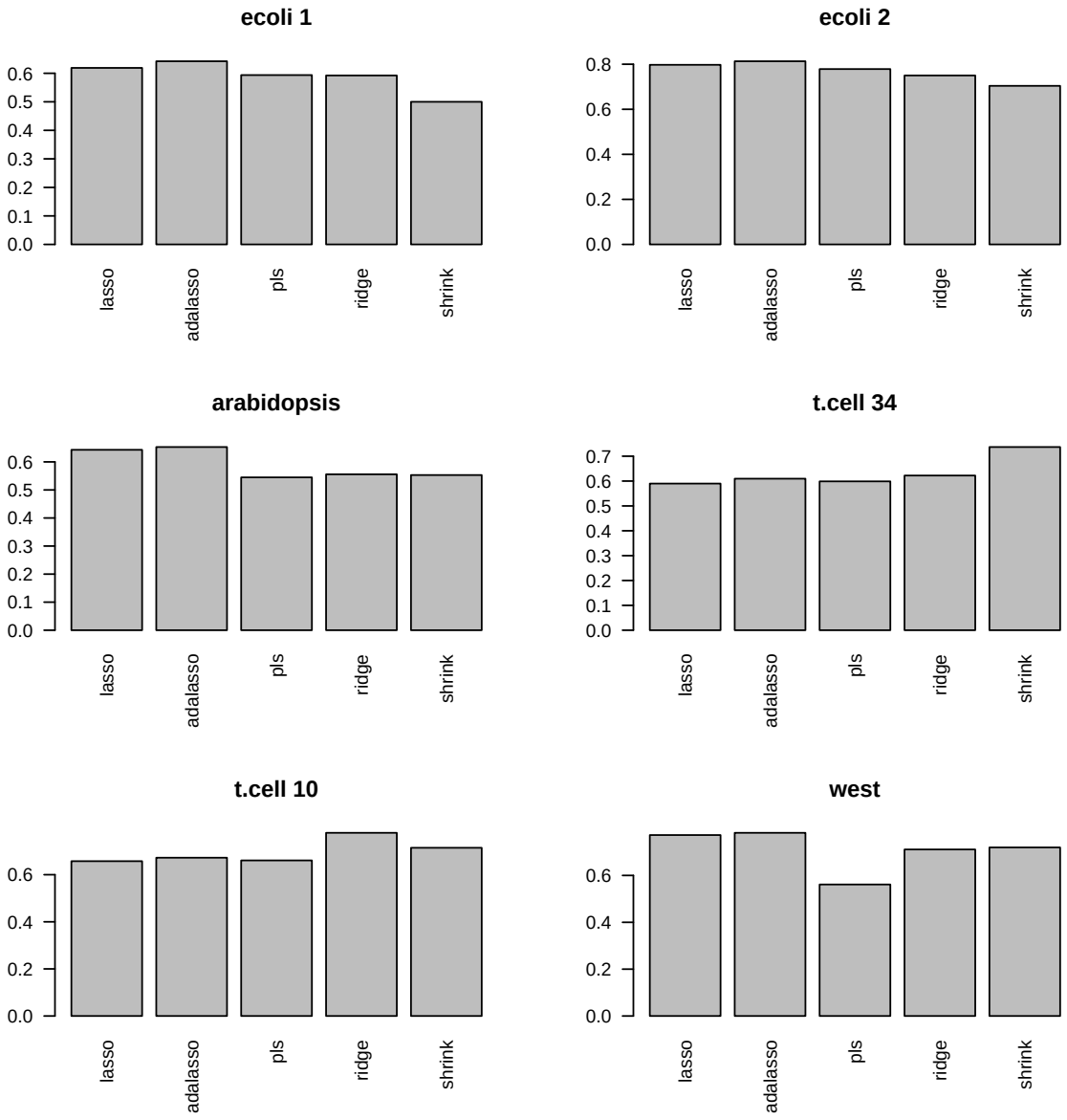


Figure 4 - Percentage of positive correlations



Tables

Table 1 - Overview of the methods

Characteristics of the five methods considered in this study: (1) shrinkage estimator of the covariance (shrink), (2) PLS regression (pls), (3) Ridge Regression (ridge), (4) Lasso regression (lasso), (5) Adaptive Lasso (adalasso). **2nd column:** Type of the method (inversion of a regularized estimate of the covariance matrix or regression-type method). **3rd column:** Parameter determining the amount of regularization. **4th column:** Method used to choose this(these) parameter(s). **5th column:** Method used to decide whether two genes should be edge-connected.

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	$\rho_{ij} \neq 0$
adalasso	regression	penalty λ ($\times 2$)	nested CV ($\times 2$)	$\rho_{ij} \neq 0$

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` [4]
`ecoli2`	9	102	yes	no	5 151	`R` package `GeneNet` [31]
`ara`	22	800	yes	no	319 600	`R` package `GeneNet`
`t.cell10`	100	58	yes	yes	1 653	`R` package `longitudinal` [21]
`t.cell34`	340	58	yes	yes	1 653	`R` package `longitudinal`
`west`	49	3883	no	no	7 536 903	http://strimmerlab.org/data.htm

Table 3 - 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.

data set		pls	ridge	lasso	adalasso	shrink
	pls	1.000	0.096	0.156	0.127	0.045
	ridge	0.686	1.000	0.600	0.457	0.390
`ecoli1`	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
	pls	1.000	0.593	0.263	0.156	0.305
	ridge	0.651	1.000	0.309	0.197	0.388
`ecoli2`	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
	pls	1.000	0.064	0.025	0.017	0.035
	ridge	0.590	1.000	0.151	0.108	0.377
`ara`	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
	pls	1.000	0.314	0.993	0.985	0.131
	ridge	0.956	1.000	1.000	1.000	0.422
`t.cell10`	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
	pls	1.000	0.053	0.762	0.670	0.031
	ridge	1.000	1.000	1.000	1.000	0.583
`t.cell34`	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
	pls	1.000	0.089	0.017	0.008	0.041
	ridge	0.667	1.000	0.118	0.062	0.236
`west`	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