The 'Un-Shrunk' Partial Correlation in Gaussian Graphical Models

Victor Bernal
University of Groningen: Rijksuniversiteit Groningen
Rainer Bischoff
University of Groningen: Rijksuniversiteit Groningen
Peter Horvatovich (p.l.horvatovich@rug.nl)
Department of Analytical Biochemistry, Groningen Research Institute of Pharmacy, University of Groningen, Groningen, 9713 AV, The Netherlands. https://orcid.org/0000-0003-2218-1140
Victor Guryev
University of Groningen: Rijksuniversiteit Groningen
Marco Grzegorczyk
Bernoulli Institute, University of Groningen, Groningen, 9747 AG, The Netherlands

Research article

Keywords: Gaussian graphical models, Partial correlations, Shrinkage, Gene regulatory networks

Posted Date: September 23rd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-76682/v1

Title: The 'un-shrunk' partial correlation in Gaussian graphical models

Authors: Victor Bernal[1,2], Rainer Bischoff[2], Peter Horvatovich[2]*, Victor Guryev[3]*, Marco Grzegorczyk[1]*

Address: [1]Bernoulli Institute, University of Groningen, Groningen, 9747 AG, The Netherlands. [2]Department of Analytical Biochemistry, Groningen Research Institute of Pharmacy, University of Groningen, Groningen, 9713 AV, The Netherlands. [3]European Research Institute for the Biology of Ageing, University Medical Center Groningen, University of Groningen, Groningen, 9713 AV, The Netherlands.

Email: p.l.horvatovich@rug.nl, v.guryev@umcg.nl, m.a.grzegorczyk@rug.nl

*Corresponding author. These authors contributed equally to this work.

Abstract

Background: In systems biology, it is important to reconstruct regulatory networks from quantitative molecular profiles. Gaussian graphical models (GGMs) are one of the most popular methods to this end. A GGM consists of nodes (representing the transcripts, metabolites or proteins) inter-connected by edges (reflecting their partial correlations). Learning the edges from quantitative molecular profiles is statistically challenging, as there are usually fewer samples than nodes ('high dimensional problem'). Shrinkage methods address this issue by learning a regularized GGM. However, it is an open question how the shrinkage affects the final result and its interpretation.

Results: We show that the shrinkage biases the partial correlation in a non-linear way. This bias does not only change the magnitudes of the partial correlations but also affects their order. Furthermore, it makes networks obtained from different experiments incomparable and hinders their biological interpretation. We propose a method, referred to as the 'un-shrunk' partial correlation, which corrects for this non-linear bias. Unlike traditional methods, which use a fixed shrinkage value, the new approach provides partial correlations that are closer to the actual (population) values and that are easier to interpret. We apply the 'un-shrunk' method to two gene expression datasets from *Escherichia coli* and *Mus musculus.*

Conclusions: GGMs are popular undirected graphical models based on partial correlations. The application of GGMs to reconstruct regulatory networks is commonly performed using shrinkage to overcome the "high-dimensional" problem. Besides it advantages, we have identified that the shrinkage introduces a non-linear bias in the partial correlations. Ignoring this type of effects caused by the shrinkage can obscure the interpretation of the network, and impede the validation of earlier reported results.

Keywords: Gaussian graphical models, Partial correlations, Shrinkage, Gene regulatory networks.

Supplementary information: Supplementary data are available at *BMC Bioinformatics* online.

1 Background

An important goal in systems biology is to elucidate gene regulatory and protein interaction patterns. To this end, in the literature many network models have been proposed, such as Relevance networks (RNs) [1], Bayesian networks (BNs) [2], and Gaussian graphical models (GGMs) [3]. For more details, see [4].

A GGM consists of a network structure of nodes representing transcripts, metabolites, or proteins, which are inter-connected by edges reflecting significant partial correlations. Partial correlations measure linear associations between pairs of random variables, where the contribution from the remaining variables is adjusted for. Unlike RNs, which are based on Pearson's correlation, GGMs remove the spurious patterns caused by confounded variables (e.g. when genes share a common regulator). Compared to BNs, GGMs scale up more efficiently to large network analyses (i.e. to omics' data) and often yield comparable network reconstruction accuracies [5]. Although the edges in a GGM are undirected, there are various methods to learn their directions [6, 7]. These aforementioned features made GGMs a popular tool in bioinformatics and biomedical studies of colon cancer [8], immunological diseases [9], diabetes [10], respiratory diseases [11], functional connectivity between brain regions [12], and chronic mental disorders [13].

Partial correlations can be computed from the (standardized) inverse of the covariance matrix (i.e. the precision matrix). In principle, the covariance matrix is unknown and has to be estimated from data. The estimated covariance matrix must be well-conditioned to ensure that its inverse exists and that numerical (or estimation) errors are not magnified during its computation. The sample covariance, as obtained from a dataset of n samples and p variables, is (i) invertible and well-conditioned when n is greater than p, (ii) invertible but ill-conditioned when n is comparable to p, and (iii) not invertible when n is smaller than p [14]. The last case is known as a 'high-dimensional problem', 'small n, large p', or '$n \ll p$'. This scenario is common in omics' studies, where often a large set of genes, proteins or metabolites is quantified in relatively few samples.

Shrinkage approaches are widely applied to deal with the 'high-dimensional problem'. They produce a more stable estimator at the cost of some bias. The most popular shrinkage approaches for estimating GGMs are Glasso [15] and the Ledoit-Wolf (LW) shrinkage [14, 16]. Glasso estimates a 'shrunk' precision matrix using L1 regularization, which forces some entries to be equal to zero. The LW-shrinkage estimates a 'shrunk' covariance (or 'shrunk' correlation) matrix, which is invertible and henceforth allows the indirect

computation of the partial correlations. Although both approaches have been successfully applied in bioinformatics, to the best of our knowledge, it has not been studied in the literature yet how the shrinkage affects/biases the estimated partial correlations.

In this paper, we present an improvement for the 'shrunk' partial correlations inferred with the LW-shrinkage [17]. In our previous work we focused on the estimation of shrinkage-based p-values [18], and here we study a fundamental source of bias, namely the effect of the shrinkage value on the 'shrunk' partial correlation. Most importantly, we show that this effect is non-linear, so that not only the magnitudes but also the order of the estimated partial correlations changes with the shrinkage value. This has unexpected consequences on the results, as GGMs learnt from different experiments (e.g. differing in their sample sizes or number of genes) have unequal shrinkage values, and thus are affected differently. Therefore, to compare studies, or to decide whether partial correlations are relevant associations, the users require (i) a 'shrunk' test of significance, and (ii) an accurate assessment of the 'shrunk' partial correlation coefficients. The first was addressed in our recent study [18], and the second is the focus of this article.

2 Methods

In this section, we review Gaussian graphical models (GGMs) with the Ledoit-Wolf (LW) shrinkage [14, 16]. We illustrate how the shrinkage modifies the network structure (i.e. the magnitudes and orders of the partial correlations) in a non-linear way. To overcome this pitfall, we propose the 'un-shrunk' partial correlation and discuss its properties. Throughout the text, uppercase bold letters are used for matrices (e.g. $\mathbf{P}$ is the matrix of partial correlations) and the *hat* symbol denotes statistical estimators (e.g. $\widehat{\mathbf{P}}$ is an estimator of $\mathbf{P}$).

2.1 The 'shrunk' partial correlation

The partial correlation is a measure of (linear) association between Gaussian variables, where confounding effects coming from the other variables are removed (i.e. a full-conditional correlation). A GGM is represented by a matrix $\mathbf{P}$ of partial correlations (Whittaker, 1990). The matrix element $\mathbf{P}_{ij}$ is the partial correlation between the variables i and j and can be computed via the relationship

$$\mathbf{P}_{ij} = \frac{-\mathbf{\Omega}_{ij}}{\sqrt{\mathbf{\Omega}_{ii}\mathbf{\Omega}_{jj}}} \qquad (1)$$

where $\mathbf{\Omega}$ is the inverse of the covariance matrix $\mathbf{C}$ (or equivalently, the inverse of the correlation matrix $\mathbf{R}$, see **Supplementary material S1**). For a dataset $\mathbf{D}$ that consists of p variables and n samples, $\mathbf{C}$ is a p × p matrix that can be estimated from data, e.g. by the sample covariance matrix $\hat{\mathbf{C}}^{\mathrm{SM}}$. However, estimating $\mathbf{C}$ is challenging when n is comparable to or smaller than p, as the estimator then becomes ill-conditioned (numerically unstable) or, even worse, non-invertible.

The LW-shrinkage estimator $\hat{\mathbf{C}}^{[\lambda]}$ consists of a convex linear combination of $\hat{\mathbf{C}}^{\mathrm{SM}}$ and a target matrix $\mathbf{T}$ (e.g. a diagonal matrix of variances), and it is defined as

$$\hat{\mathbf{C}}^{[\lambda]} := (1-\lambda)\hat{\mathbf{C}}^{\mathrm{SM}} + \lambda\mathbf{T} \qquad (2)$$

where $\lambda \in (0,1)$ represents the weight allocated to $\mathbf{T}$. The inverse of $\hat{\mathbf{C}}^{[\lambda]}$, denoted by $\hat{\mathbf{\Omega}}^{[\lambda]}$, can then be plugged into **Equation (1)**, yielding

$$\hat{\mathbf{P}}^{[\lambda]}{}_{\mathrm{ij}} = \frac{-\hat{\mathbf{\Omega}}^{[\lambda]}{}_{\mathrm{ij}}}{\sqrt{\hat{\mathbf{\Omega}}^{[\lambda]}{}_{\mathrm{ii}}\hat{\mathbf{\Omega}}^{[\lambda]}{}_{\mathrm{jj}}}} \qquad (3)$$

This is the 'shrunk' partial correlation [17], and an edge in the GGM structure is selected according to its magnitude and/or statistical significance. The operations involved in **Equation (3)** (i.e. matrix inversion, square roots, and standardization) suggest that $\hat{\mathbf{P}}_{\mathrm{ij}}{}^{[\lambda]}$ is a non-linear function of λ. In addition, the value of λ is usually optimized by minimizing the mean square error between $\hat{\mathbf{C}}^{[\lambda]}$ and $\mathbf{C}$, but this λ does not necessarily minimize the MSE between $\hat{\mathbf{\Omega}}^{[\lambda]}$ and $\mathbf{\Omega}$.

Pitfalls of the 'shrunk' partial correlation

Let us consider the following covariance matrix $\mathbf{C}$ and its inverse $\mathbf{\Omega}$,

$$\mathbf{C} = \begin{pmatrix} 1 & \frac{1}{2} & \frac{-1}{4} & \frac{-1}{8} \\ \frac{1}{2} & 1 & \frac{-3}{4} & \frac{-3}{4} \\ \frac{-1}{4} & \frac{-3}{4} & 1 & \frac{3}{4} \\ \frac{-1}{8} & \frac{-3}{4} & \frac{3}{4} & 1 \end{pmatrix}, \quad \mathbf{\Omega} = \begin{pmatrix} \frac{160}{97} & \frac{-152}{97} & \frac{-8}{97} & \frac{-88}{97} \\ \frac{-152}{97} & \frac{416}{97} & \frac{124}{97} & \frac{200}{97} \\ \frac{-8}{97} & \frac{124}{97} & \frac{272}{97} & \frac{-112}{97} \\ \frac{-88}{97} & \frac{200}{97} & \frac{-112}{97} & \frac{320}{97} \end{pmatrix} \tag{4}$$

The matrix $\mathbf{C}$ is invertible, its eigenvalues are 2.66, 0.94, 0.25, and 0.15, its determinant is 0.09, and its condition number 16.89. From **Equation (1)** we find the partial correlations $\mathbf{P}_{12} = 152/\sqrt{160}\sqrt{416} = 0.63$ and $\mathbf{P}_{34} = 112/\sqrt{272}\sqrt{320} = 0.37$, concluding that $\mathbf{P}_{12}$ is *stronger* than $\mathbf{P}_{34}$. On the other hand, computing $\hat{\mathbf{C}}^{[\lambda]}$ with **Equation (2)** and substituting it into **Equation (3)** gives the 'shrunk' partial correlations. As λ increases, the value of the two 'shrunk' partial correlations change. **Figure 1a** shows that for λ greater than 0.3, $\hat{\mathbf{P}}^{[\lambda]}{}_{12}$ gets *weaker* than $\hat{\mathbf{P}}^{[\lambda]}{}_{34}$ and their relative order reverses. Although **Equation (2)** is a linear shrinkage of $\hat{\mathbf{C}}^{[\lambda]}$, the matrix inversion and the standardization in **Equation (3)** propagate the effect from λ to $\mathbf{P}^{[\lambda]}$ in a non-linear way.

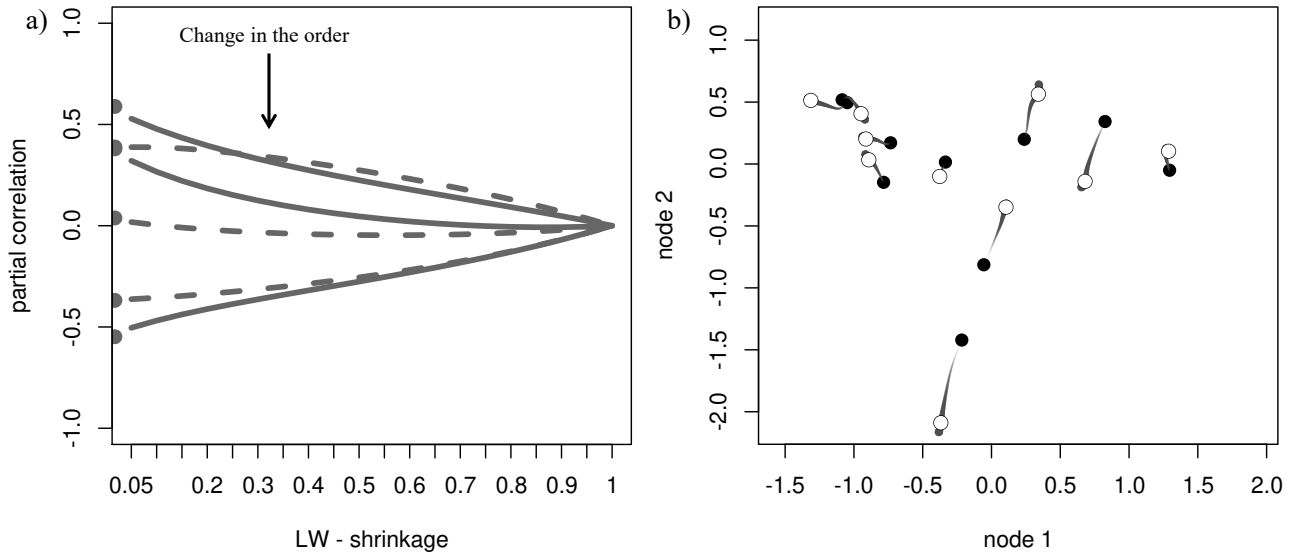

Fig. 1. Non-linear effects of the shrinkage. Panel a) partial correlations obtained from **Equations (3-4)** while varying the shrinkage value. Lines represent 'shrunk' partial correlations, and their intersection reflects the changes in their order. Panel b) scatter plot of 'shrunk' data points (for two variables) that change as the shrinkage increases. We simulated data (n = 10) from a random network (p = 8), and the effect of the shrinkage at the data level is obtained via the Singular Value Decomposition of the data matrix, see **Supplementary material S3.** In black: the original data points. In grey: the data points changing their positions for $\lambda \in (0, 1)$. The dots' sizes are proportional to λ. In white: the 'shrunk' data points for the optimal shrinkage of 0.65.

2.2 **Additional properties**

Without loss of generality, we now switch to the correlation matrix $\mathbf{R}$ (the standardized $\mathbf{C}$). Using **Equation (2), R** (instead of **C**) can be 'shrunk' towards its diagonal ($\mathbf{T}$ is the identity matrix). For small samples sizes, the sample correlation $\widehat{\mathbf{R}}^{\mathrm{SM}}$ (the standardized $\widehat{\mathbf{C}}^{\mathrm{SM}}$) is not positive definite. Some eigenvalues of $\widehat{\mathbf{R}}^{\mathrm{SM}}$ can be (i) near to zero, (ii) equal to zero, or (iii) slightly negative. This translates into $\widehat{\mathbf{R}}^{\mathrm{SM}}$ being (i) ill-conditioned, (ii) singular, or (iii) indefinite, respectively. The case of an indefinite matrix arises from a numerical inaccuracy that can make zero eigenvalues slightly negative.

Let α_k $(k = 1, 2, \ldots, p)$ denote the eigenvalues of $\widehat{\mathbf{R}}^{\mathrm{SM}}$. **Equation (2)** yields $\widehat{\mathbf{R}}^{[\lambda]}$, which has eigenvalues:

$$\alpha^{[\lambda]}{}_k = (1-\lambda)\alpha_k + \lambda \quad (k = 1, 2, \ldots, p) \qquad (5)$$

and the shrinkage $\lambda \in (0, 1)$ transforms each α_k into a positive $\alpha^{[\lambda]}{}_k$, so that $\widehat{\mathbf{R}}^{[\lambda]}$ becomes positive definite (see **Supplementary material S2**). Accordingly, the eigenvalues of $\widehat{\mathbf{R}}^{[\lambda]^{-1}}$ that are obtained by inversion:

$$\frac{1}{\alpha^{[\lambda]}{}_k} = \frac{1}{(1-\lambda)\alpha_k + \lambda} \qquad (k = 1, 2, \ldots, p) \qquad (6)$$

are positive as well. This ensures that $\widehat{\mathbf{R}}^{[\boldsymbol{\lambda}]^{-1}}$ is a symmetric positive definite precision matrix (i.e. the covariance matrix of the full-conditioned random variables).

Shrinking the data

Traditionally, the shrinkage is interpreted as a modification of the statistical model (a 'shrunk' covariance/correlation/partial correlation), where the data remains unchanged. However, most research questions need to be interpreted in terms of the dataset. We therefore propose to discuss the shrinkage from a different perspective, namely from the data level. To this end, we use the Singular Value Decomposition (SVD) of the data matrix **D** and the fact that the shrinkage only modifies the eigenvalues, while the eigenvectors stay identical, see **Supplementary material S2**. As singular values are the positive square roots of the eigenvalues $\alpha^{[\lambda]}$ given in **Equation (5)**, we can derive the SVD of the 'shrunk' data matrix $\mathbf{D}^{[\lambda]}$ as

$$\mathbf{D}^{[\lambda]} = \mathbf{U}\mathrm{diag}\left(\sqrt{(n-1)\alpha^{[\lambda]}}\right)\mathbf{V}^{t} \qquad (7)$$

Here $\mathbf{U}$ and $\mathbf{V}$ are the matrices of (left and right) singular vectors of $\mathbf{D}$, and the singular values are replaced by their 'shrunk' counterparts. This relationship allows us to study the shrinkage effect at the data level. That is, analyzing the original dataset $\mathbf{D}$ with a 'shrunk' model $\hat{\mathbf{C}}^{[\lambda]}$ is equivalent to analyzing $\mathbf{D}^{[\lambda]}$ with the classical model $\hat{\mathbf{C}}^{SM}$. To illustrate this, we generate data from a network (p = 8, n =10) and investigate what happens to the original data points as the shrinkage increases; see **Figure 1b** and **Supplementary material S3** for more details.

2.3 **The 'un-shrunk' partial correlations**

Here we propose the concept of 'un-shrunk' partial correlation, which we define as the limit of $\hat{\mathbf{P}}^{[\lambda]}{}_{ij}$ as λ approaches zero,

$$\mathbf{P}^{[0]}{}_{ij} := \lim_{\lambda \to 0} \mathbf{P}^{[\lambda]}{}_{ij} \qquad (8)$$

where $\mathbf{P}^{[\lambda]}{}_{ij}$ is a continuous and bounded function of λ, see **Supplementary material S4**. For a general proof of the existence of this limit, see **Supplementary material S5**. The key idea is that there is no divergence in **Equation (8)**, and to illustrate this we consider the eigen-decomposition,

$$\hat{\mathbf{R}}^{[\lambda]^{-1}} = \mathbf{V}\,\mathrm{diag}\left(\frac{1}{\alpha^{[\lambda]}}\right)\mathbf{V}^{t} \qquad (9)$$

where $\mathbf{V}$ is a matrix, whose columns are the eigenvectors of $\hat{\mathbf{R}}^{[\lambda]}$, and $\mathrm{diag}\left(1/\alpha^{[\lambda]}\right)$ is the diagonal matrix of eigenvalues $1/\alpha^{[\lambda]}{}_{k}$ $(k = 1, 2, \ldots, p)$.

Let us assume that $\hat{\mathbf{R}}^{SM}$ is singular, e.g. $\alpha_k = 0$, and recall two facts from the previous subsection. First, any $\alpha_k = 0$ is transformed into $\alpha^{[\lambda]}{}_{k} = \lambda$ in **Equation (5)**, and the corresponding eigenvalue of $\hat{\mathbf{R}}^{[\lambda]^{-1}}$ is

$1/\alpha^{[\lambda]}{}_k = 1/\lambda$ by **Equation (6)**. Second, $\widehat{\mathbf{R}}^{SM}$ and $\widehat{\mathbf{R}}^{[\lambda]}$ have the same eigenvectors, because the shrinkage only changes the eigenvalues (see **Supplementary material S2**). Substituting **Equation (9)** in **Equation (3)**, and factorizing out the term $1/\lambda$ gives

$$\mathbf{P}^{[\lambda]}{}_{ij} = \frac{-\frac{1}{\lambda}\left[\mathbf{V}\,\mathrm{diag}\left(\frac{\lambda}{\alpha^{[\lambda]}}\right)\mathbf{V}^t\right]_{ij}}{\frac{1}{\lambda}\sqrt{\left[\mathbf{V}\,\mathrm{diag}\left(\frac{\lambda}{\alpha^{[\lambda]}}\right)\mathbf{V}^t\right]_{ii}\left[\mathbf{V}\,\mathrm{diag}\left(\frac{\lambda}{\alpha^{[\lambda]}}\right)\mathbf{V}^t\right]_{jj}}} \qquad (10)$$

Any singularity disappears by cancelling the term $1/\lambda$. As $\alpha^{[\lambda]}{}_k > \lambda$, the diagonal elements $\lambda/\alpha^{[\lambda]}$ in **Equation (10)** have limits equal to (i) zero for $\alpha_k \neq 0$, or (ii) one for $\alpha_k = 0$, see **Equations (S43, S48, S52, S55, S58)**. In this sense, the ratio that defines the 'shrunk' partial correlation in **Equation (10)** does not diverge when removing the shrinkage. For a general proof of the existence of this limit, see **Supplementary material S5**. We propose this limit as a generalization of the classical partial correlation. The idea resembles the classical example from Calculus, where the limits to zero of x^{-1} and x^{-2} are both infinite; while the limit of their ratio x^{-1}/x^{-2} is finite (i.e. zero). For illustration, let us consider a 3×3 correlation matrix of ones (all variables are maximally correlated). The matrix is singular as all $\alpha_k = 0$ and **Equation (2)** gives

$$\mathbf{R}^{[\lambda]} = \begin{pmatrix} 1 & (1-\lambda) & (1-\lambda) \\ (1-\lambda) & 1 & (1-\lambda) \\ (1-\lambda) & (1-\lambda) & 1 \end{pmatrix} \qquad (11)$$

using **Equation (3)** we obtain

$$\mathbf{P}^{[\lambda]} = \frac{\frac{(1-\lambda)}{\lambda}}{\frac{1}{\lambda}} \begin{pmatrix} 1 & (2-\lambda)^{-1} & (2-\lambda)^{-1} \\ (2-\lambda)^{-1} & 1 & (2-\lambda)^{-1} \\ (2-\lambda)^{-1} & (2-\lambda)^{-1} & 1 \end{pmatrix} \qquad (12)$$

and we see that $\mathbf{P}^{[0]}{}_{ij} = \lim_{\lambda \to 0} \mathbf{P}^{[\lambda]}{}_{ij} = 1/2$. Here it is worth noting that a simple linear re-scaling by $(1-\lambda)$ is *not sufficient* to remove the shrinkage effect (see **Figure S1**). In this example, $\mathbf{P}^{[\lambda]}{}_{ij}$ would become $1/(2-$

λ) which for $\lambda = 1/3$ gives $\mathbf{P}^{[1/3]}{}_{ij} = 3/5 \approx 0.6$, and for $\lambda = 2/3$ is $\mathbf{P}^{[2/3]}{}_{ij} = 3/4 \approx 0.75$. More toy examples can be found in **Supplementary material S6**.

2.4 **Practical implementation**

From a mathematical perspective, **Equation (8)** can be computed by means of the analytical results in **Equations (S43, S48, S52, S55, S58)**. However, as numerical inaccuracies make the elements of **V** unreliable, and often render zero eigenvalues (slightly) positive/negative, these results are un-practical. To circumvent numerical issues, we apply a simple approximation. We compute the 'shrunk' partial correlation $\mathbf{P}^{[\lambda]}{}_{ij}$ for different $\lambda \in (0, 1)$ values, and then we fit a polynomial function to the points. By extrapolating the polynomial to $\lambda = 0$ the limit in **Equation (8)** is approximated. For more details we refer to **Supplementary material S7**. All computations are performed with the R package *GeneNet* version 1.2.13.

3 Data

3.1 **Escherichia coli microarray data**

This data set consists of *Escherichia coli* microarray gene-expression measurements [19] from a study of the temporal stress response upon expression of recombinant human superoxide dismutase (SOD). SOD expression was induced by isopropyl β-D-1-thiogalactopyranoside (IPTG), which is a lactose analogue inducer of the lac operon, and measured at 8, 15, 22, 45, 68, 90, 150, and 180 minutes. The authors identified 102 out of 4289 protein coding genes as differentially expressed in one or more samples after induction. Data preprocessing included $\log_2$-ratio transformation with respect to the first time point. The final data set consists of expression values of 102 genes with 9 time points and was obtained from the R package *GeneNet* version 1.2.13. Accessed May 15, 2020.

3.2 **Mus musculus RNA sequencing data**

This dataset corresponds to single end RNA-Seq reads from 21 male mice from two strains (B6, n = 10 and D2, n = 11), and is available at ReCount: http://bowtie-bio.sourceforge.net/recount/ExpressionSets/bottomly_eset.RData [20]. Accessed May 15, 2020. Genes with low count's averages across samples (less than 5) were excluded. After correcting by strain type, 223 genes

out of 9431 were identified as differentially expressed using the R package *limma* and Benjamini-Hochberg (BH) adjusted *p-values* < 0.05 [21, 22]. We applied upper quartile normalization, $\log_2$-transformation and a correction by strain type using linear models. The final data set consists of 223 genes with 21 samples.

4 Results

4.1 Analysis of simulated data

Here we demonstrate the advantages of the new 'un-shrunk' estimator over the 'shrunk' estimator that uses a fixed shrinkage value λ. The evaluation consists of comparing the reconstructed partial correlations to their actual (population) values. In total, 1251 data sets were simulated (see **Table S2**).

First, we study the shrinkage distortion for partial correlations coefficient with magnitudes between -1 and 1. To this end, we simulate two networks and 25 datasets in the high dimensional (or $n << p$) scenario (p = 50, 100, and n = 40). **Figure 2** presents the partial correlations estimated with (i) the optimal shrinkage, (ii) the new 'un-shrunk' approach, and (iii) their actual values. It can be seen that the new 'un-shrunk' coefficients are consistently closer to the actual values than the 'shrunk' coefficients. In **Figure S2** we see how a GGM structure (i.e. its edges) changes with varying λ. Here we used p = 10 and n = 1000 (a large sample size) to reduce the sampling variability, such that the observed effect on the edges order can be attributed to λ. **Table S1** lists the edges sorted by their magnitudes.

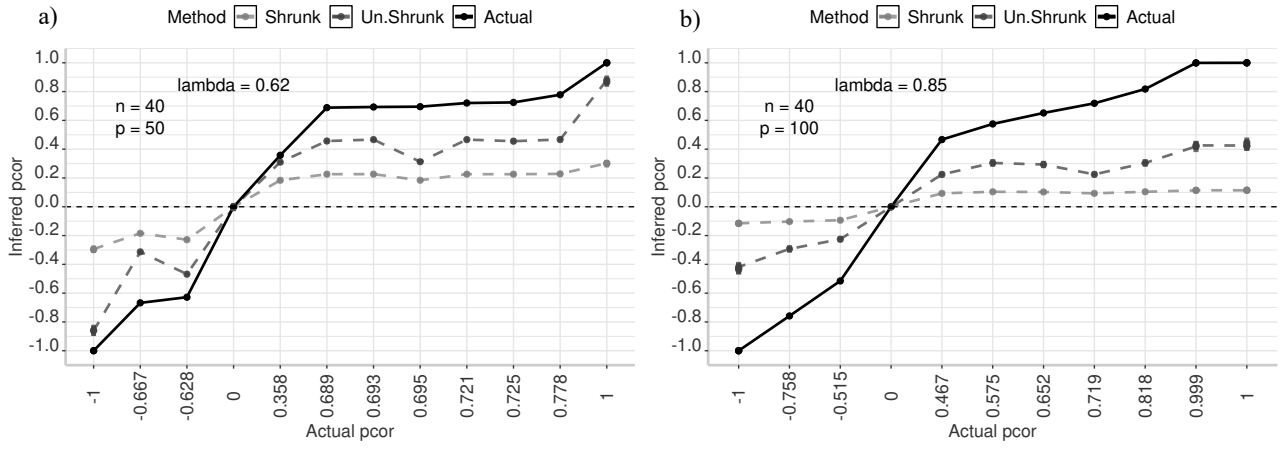

Fig. 2. Partial correlations versus their actual (population) values. We simulated 25 datasets from two network structures (n= 40) and reconstructed the 'shrunk' and 'un-shrunk' partial correlations. Panel a) corresponds to p = 50, and panel b) to p = 100. In black: the actual partial correlations. In light grey (dashed): the average of the 'shrunk' method. In dark grey (dashed): the average of the new 'un-shrunk' method. Error bars represent 2 standard errors.

Second, we study whether GGMs from different experimental conditions are comparable. We create a (random) network structure and simulate 10 datasets for each n = 10, 20, …, 90. **Figure 3** shows the 'shrunk' and 'un-shrunk' partial correlations as a function of n.

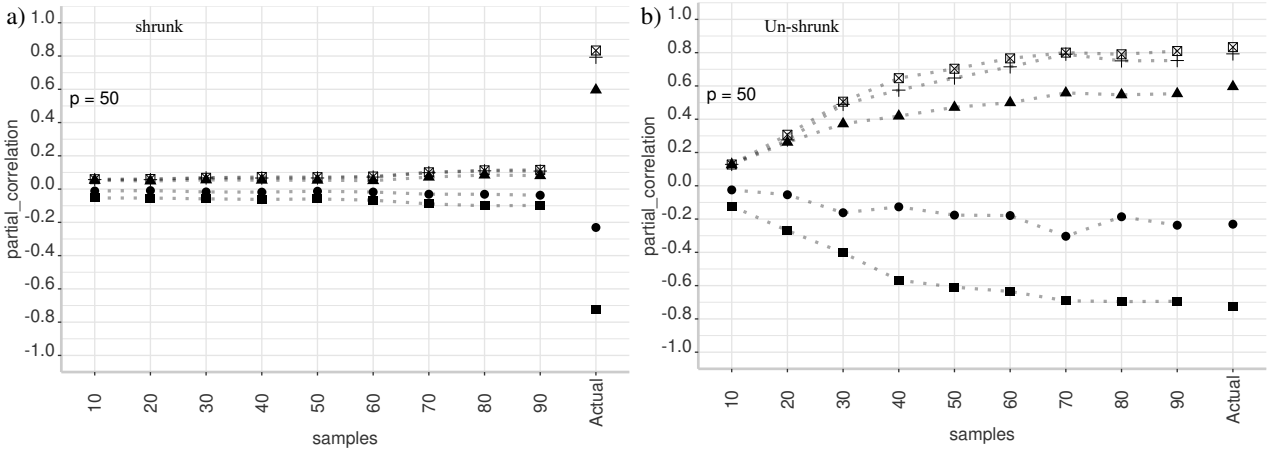

Fig. 3. Partial correlation versus sample size. We create a fixed network structure for p = 50, and simulated data for sample size n ranging from 10 to 90. Panels a, b) show the average 'shrunk' and 'un-shrunk' partial correlations (over 10 simulations). The proposed 'un-shrunk' partial correlations are closer to their actual values.

The new 'un-shrunk' partial correlation approaches the actual (population) values as the sample size increases, while the shrunk counterpart stays far away from the actual values. **Figure S3** compares the performance of both methods for different combinations of p and n for an actual partial correlation of 0.5. The results are presented in the form of a heatmap over a p-n grid. The color scale shows the L1 distances to the population value of 0.5. We see that the proposed 'un-shrunk' estimation is consistently closer to the population values. In **Figure S4** we illustrate how, in general, partial correlations are deflated when the samples size is very small, e.g. n = *10*.

4.2 **Analysis of experimental data**

Effects of human superoxide dismutase (SOD) protein expression on transcript expression in *E. coli*

Following previous works, the dataset is treated as static and nominal p-values are considered significant at the 0.05 level [17, 18]. The 'shrunk' approach infers an optimal shrinkage of 0.18. A scatter plot of the partial correlations in absolute value is presented in **Figure 4a, b.** The panels are segmented into four regions using a threshold of |pcorr| = 0.1 or p-values = 0.05. Clustering of points around the diagonal indicates that the 'shrunk' and 'un-shrunk' coefficients (as well as the p-values) are similar.

The 'shrunk' network has 238 edges (involving 74 genes), and the 'un-shrunk' network has 34 additional edges, for a total of 272 edges (involving 78 genes). We evaluate the enrichment of the connected genes using the protein-protein interaction STRING database version 10 [23, 24]. The enriched Gene Ontologies (GOs), and Kyoto Encyclopedia of Genes and Genomes (KEGGs) pathways are reported in **Supplementary Tables S4 a-d**. The complete set of 102 genes was set as background, which mapped 96 proteins.

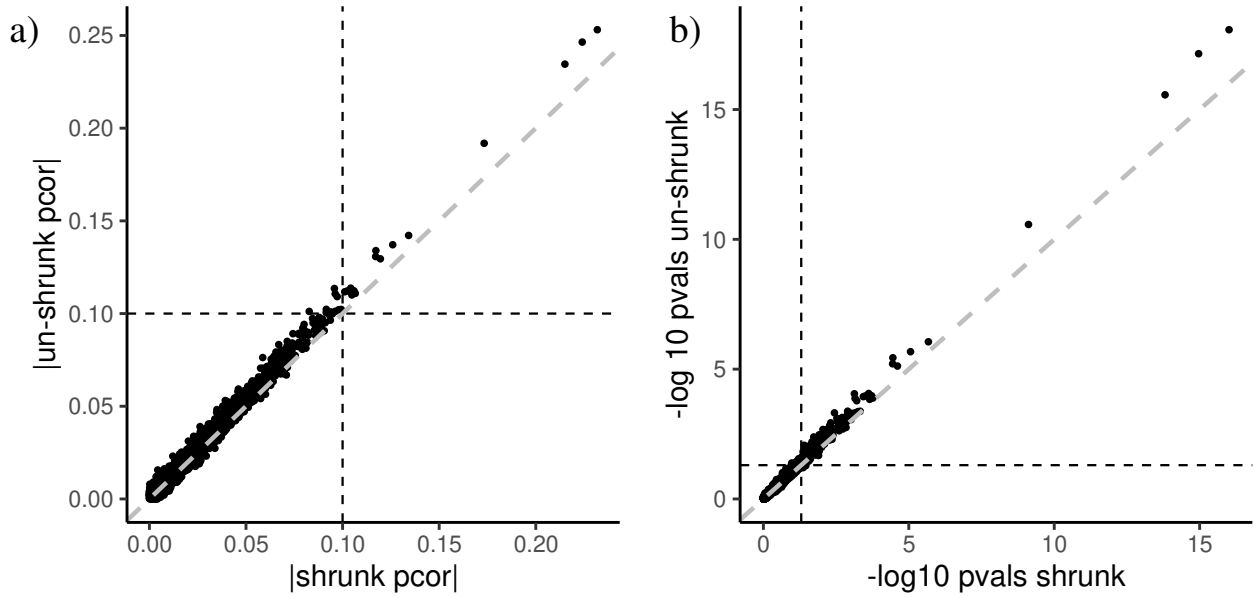

Fig. 4. Comparison of partial correlations from *Eschericha coli* microarray data. Panel a) Scatter plot of the partial correlation in absolute value |pcorr|. Panel b) Scatter plot of –log10(p-values). P-values were computed with the shrunk approach (Bernal et al. 2019). The panels are segmented into four regions using a threshold of |pcorr| = 0.1 or p-values = 0.05 (in dashed lines). The 'un-shrunk' model has 34 additional significant edges at p-values < 0.05.

The expected number of interactions in STRINGdb for a random set of proteins of similar size was 581. The 102 genes retrieved 668 interactions (*p-value* = 0.00025). The connected genes for the 'shrunk' method mapped onto 69 proteins with 332 interactions (expected = 286, *p-value* = 0.00438). For the 'un-shrunk' it mapped onto 73 proteins with 372 interactions (expected = 320, *p-value* = 0.00272). **Tables S4 a-b** show the GO enrichment, where the new the 'un-shrunk' method has 9 additional GOs and one KEGG pathway (all related to pyrimidine's metabolism). **Tables S4 c-d** show that the connected genes from both methods are enriched for (i) galactitol catabolic process (GO:0019404, FDR = 0.048) and (ii) galactitol metabolic process (GO:0019402, FDR = 0.048) and (iii) the galactose metabolism KEGG pathway (FDR = $1.96\ 10^{-7}$, $1.16\ 10^{-7}$). The network structures can be found in **Figure S6.**

Analysis of M. musculus RNA-seq dataset

Figure 5 presents scatter plots of the partial correlation (in absolute values) and their p-values. The 'shrunk' approach infers an optimal shrinkage of 0.69. Points clustering near the diagonal indicate that the partial correlations are similar in magnitude, while scattering away from it reflects changes in their order.

The 'shrunk' network has no connections, and the 'un-shrunk' network has one connection using BH adjusted *p-values* < 0.05. As the networks are empty, we proceed with nominal *p-values* at 0.001, where the 'shrunk' and 'un-shrunk' networks have 19 and 33 edges, involving 29 and 40 genes, respectively. The 223 genes mapped onto 208 proteins (STRINGdb) with 478 interactions (*p-value* = 0.09), while the expected number of interactions (for a random set of proteins of similar size) was 449. The connected genes in the 'shrunk' network mapped onto 29 proteins with 3 interactions (expected = 4, *p-value* = 0.84). The new 'un-shrunk' method mapped 36 proteins with 5 interactions (expected = 6, *p-value* = 0.78). The Gene Ontologies (GOs), and Kyoto Encyclopedia of Genes and Genomes (KEGGs) pathways are reported in Supplementary **Tables S5 a-d**. The 'shrunk' and 'un-shrunk' connected genes are enriched for 562 and 647 GOs, and for 7 and 13 KEGGs pathways at FDR 0.05, respectively. The GOs belong mainly to regular cellular processes. The 'un-shrunk' method retrieves (additional) KEGGs related to degenerative diseases (5010: Alzheimer and 5012: Parkinson, both with FDR = $7.17\ 10^{-4}$). The network structure can be found in **Figure S7**.

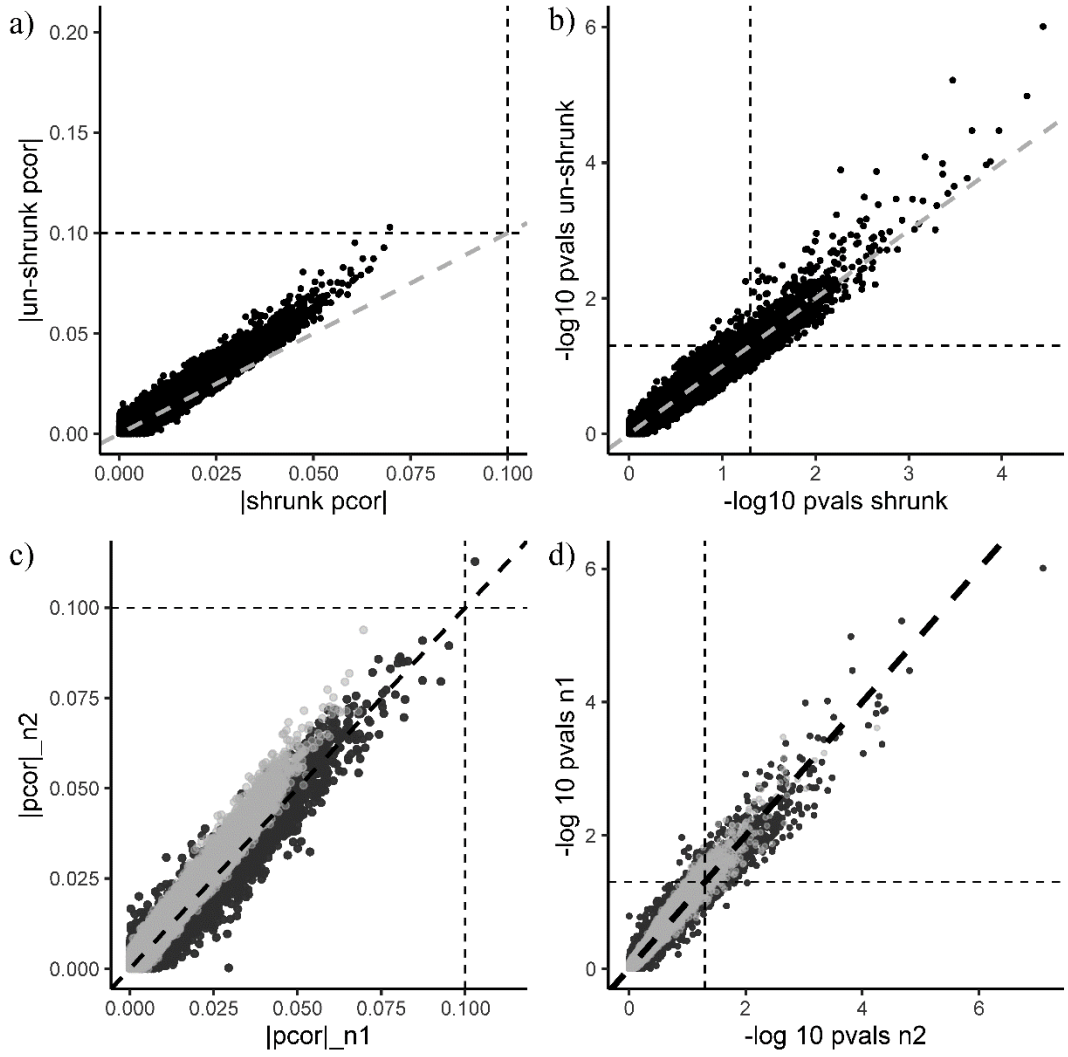

Fig. 5. Comparison of partial correlations from Mus musculus RNA-seq data. Panel a) Scatter plot of the partial correlation in absolute value |pcorr|. Panel b) scatter plot of –log10 (p-values). Panel c, d) scatter plots of |pcorr|, and of –log10 (p-values) for different sample sizes ($n_1 = 21$, $n_2 = 31$). To generate an additional dataset ($n_2 = 31$) from the original one ($n_1 = 21$), we use a sampling approach. The dashed lines show the threshold of |pcorr| = 0.1 and p-values = 0.05, respectively. P-values were computed with the shrunk approach (Bernal et al. 2019). In light grey: the 'shrunk' method with $\lambda = 0.69$ and 0.42. In dark grey: the 'un- shrunk' method. The 'shrunk' partial correlations suffer a slope shift for different sample sizes and are not comparable. The 'un-shrunk' partial correlations are concentrated around the diagonal, what reflects that they are comparable.

To compare the methods for different shrinkage values, we generate a second dataset as follows: (i) we randomly choose 10 samples out of the original 21 samples, (ii) we add Gaussian noise (mean = 0, sd = 0.05) to all the expression levels, and (iii) we merged the ten newly generated samples to the original dataset, which results in 31 samples. The 'shrunk' approach retrieves a new optimal shrinkage of 0.42. Both methods were applied to the new data, and plotted against the previous results as shown in **Figure 5c-d**. While the shrinkage change (from 0.69 to 0.42) made the 'shrunk' method deviate from the diagonal line (and hard to compare directly), the new method provides coefficients that are more similar.

5 Discussion

GGMs are undirected graphical models that represent pairwise partial correlations in the form of a network. They are widely used in many fields, because they are computationally fast and simple to interpret. Despite of that, the estimation of partial correlations from gene-expression data is challenging whenever there are fewer samples than genes. This motivated the development of estimators based on shrinkage; however, we have observed some unexpected effect of the shrinkage and we have investigated it.

In particular, we have identified a bias in the 'shrunk' partial correlations. The bias is a non-linear effect caused by the shrinkage value, which modifies the magnitudes and order of the partial correlations, see **Figure 1, S1, S2**. As the order of the edges is affected non-linearly, re-scaling the 'shrunk' partial correlations (e.g. dividing it by $1 - \lambda$) is not sufficient (see **Figure S1** and **Section 2.3**). Consequently, as GGMs learnt from different experimental conditions use different shrinkage values, they infer networks that are not comparable.

To correct for this bias, we have introduced the concept of 'un-shrinking' the partial correlation. On the theoretical side, the 'un-shrunk' partial correlation is a generalization of the classical partial correlation, and it is defined even for singular matrices. On the applied side, the 'un-shrunk' partial correlation is easy to interpret, because its magnitude is between -1 and 1, and it significantly closer to its actual (population) value (**Figures 2, 3, and S3**). For the well-conditioned case, we have also shown that the shrinkage can be considered as a transformation of the original dataset into a new 'shrunk' dataset, see **Figure 1b**. Hence, 'un-shrinking' could be interpreted as the limit of the partial correlations when the 'shrunk' data points approach their original values.

Our empirical results show that they are local (edge-wise) shrinkage distortions. These can be identified by comparing the scatter plots of the partial correlation coefficients differing in their shrinkage value. For instance, (i) the slope reflects whether the scales/magnitudes of the coefficients are different, and (ii) points scattering away from the diagonal line reflect that their order has changed. For the *E. coli* dataset, the strongest connections (in both models) were *lacA-lacZ, lacY-lacZ*, and *lacA-lacY*, all related to the lac operon (that was induced by IPTG in the experiment). The result is in agreement with previous analyses [17, 18], however, it includes 34 additional edges. For the *E. coli's* and *M. musculus's* networks in **Figure 4-5** the edges are highly correlated, but their relative order is different. **Figure 5c** shows that the slope of the 'shrunk' coeffi-

cients (in light grey) is shifted away from the diagonal, and thus the two networks are not comparable. In contrast, the new 'un-shrunk' coefficients (in dark grey) are closer to the diagonal line.

A non-linear effect on the partial correlations translates into a non-linear effect on the *p-values*, and consequently affects the inferred network structure. The 'shrunk' *p-values* [18] account for the change in the magnitudes, but do not include a correction for the non-linear effects that alter the edge order. This is apparent from **Figure 5d**, where the *p-values* are around the diagonal line in both cases, but their deviations are different. One may be tempted to say that the 'shrunk' points (in light grey) are less dispersed (or more similar) in **Figure 5d,** however, this is an effect of the variance-bias trade-off. Due to the shrinkage, partial correlations are less variable, but their order is biased.

In **Figure S3** we see that for every p and $n \geq 30$, the 'un-shrunk' coefficients are closer to their actual values. From this figure the new method seems particularly superior when $p / n > 2$. For very small samples, e.g. $n = 10$ or 20, we see that both methods are suboptimal because the model's assumptions are not necessarily fulfilled, see **Figure S4a**. Very often in bioinformatics the data is transformed e.g. log transformed, scaled. The transformed data is *approximately* a sample from a Gaussian distribution, where the approximation improves for larger samples. This mismatch is not necessarily negligible for very small samples. For instance, the *Law of large numbers* ensures that the sample mean converges to the population mean when n approaches infinity. The *Central limit theorem* states that the scaled sample mean is *asymptotically* normal, with an error that depends on n, see *Berry–Esseen theorem* [25]. The same applies to the sample correlation [26], and consequently for very small samples the distributional assumptions of GGMs are not always met.

To the best of our knowledge, this is the first study aiming to *de-regularize* the estimates. While the 'un-shrunk' partial correlation can be found algebraically (see **Section 2.1**), we employed a polynomial approximation in order to achieve feasible computational costs for large scale applications. In principle, large samples can cause weak associations to be statistically significant, and small samples can cause strong associations to be non-significant. A proper discussion should therefore report the magnitude and significance of the estimates, as they provide different pieces of information. In this case, it must be accounted for the fact that the partial correlations are 'shrunk' (biased), before concluding about the result. Ignoring the shrinkage value would divorce the analysis from the original data (and its biological meaning) what could obscure the interpretation, and impede the validation of earlier reported results (e.g. in biomarker's discovery).

Declarations

Ethics approval and consent to participate: Not applicable

Consent for publication: Not applicable

Availability of data and materials: The R code that generates the results is available in https://github.com/V-Bernal/UnShrunk. The datasets supporting the conclusions of this article are available in the R package *GeneNet* version 1.2.13 (*E. coli* dataset) and at http://bowtie-bio.sourceforge.net/recount/ExpressionSets/bottomly_eset.RData. (*M. musculus* dataset).

Competing interests: None

Funding: This work was supported by the Data Science and System Complexity Centre (DSSC) of the University of Groningen. This research was part of the Netherlands X-omics Initiative and partially funded by NWO, project 184.034.019.

Authors' contributions: VB. PH, VG, MG designed the study. VB developed the concept and implemented the code. VB and MG wrote the mathematical proofs and other theoretical results. VG, RB, and PH contributed to the analysis and interpretation of the results. All authors participated in the writing of the manuscript.

Acknowledgements: Not applicable

References

1. Butte AJ, Kohane IS. Relevance Networks: A First Step Toward Finding Genetic Regulatory Networks Within Microarray Data. In: Parmigiani G., Garrett E.S., Irizarry R.A. ZSL, editor. The Analysis of Gene Expression Data. NY: Springer, New York, NY; 2003. p. 428–46.

2. Friedman N, Linial M, Nachman I, Pe'er D. Using Bayesian Networks to Analyze Expression Data. J Comput Biol. 2000;7:601–20.

3. Edwards D. Introduction to Graphical Modelling 2nd edition. 2nd edition. Springer-Verlag New York; 2000.

4. Huynh-Thu VA, Sanguinetti G. Gene Regulatory Network Inference: An Introductory Survey. In: Gene Regulatory Networks. 2019. p. 1–23.

5. Werhli A V., Grzegorczyk M, Husmeier D. Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical gaussian models and bayesian networks.

Bioinformatics. 2006;22:2523–31.
6. Opgen-Rhein R, Strimmer K. Learning causal networks from systems biology time course data: An effective model selection procedure for the vector autoregressive process. BMC Bioinformatics. 2007.
7. Opgen-Rhein R, Strimmer K. From correlation to causation networks: A simple approximate learning algorithm and its application to high-dimensional plant gene expression data. BMC Syst Biol. 2007;1:37.
8. Beerenwinkel N, Antal T, Dingli D, Traulsen A, Kinzler KW, Velculescu VE, et al. Genetic progression and the waiting time to cancer. PLoS Comput Biol. 2007;3:e225.
9. Benedetti E, Pučić-Baković M, Keser T, Wahl A, Hassinen A, Yang JY, et al. Network inference from glycoproteomics data reveals new reactions in the IgG glycosylation pathway. Nat Commun. 2017;8:1483.
10. Keller MP, Choi Y, Wang P, Davis DB, Rabaglia ME, Oler AT, et al. A gene expression network model of type 2 diabetes links cell cycle regulation in islets with diabetes susceptibility. Genome Res. 2008;:gr. 074914.107.
11. Imkamp K, Bernal V, Grzegorzcyk M, Horvatovich P, Vermeulen CJ, Heijink IH, et al. Gene network approach reveals co-expression patterns in nasal and bronchial epithelium. Sci Rep. 2019;:1–13. doi:10.1038/s41598-019-50963-x.
12. Das A, Sampson AL, Lainscsek C, Muller L, Lin W, Doyle JC, et al. Interpretation of the precision matrix and its application in estimating sparse brain connectivity during sleep spindles from human electrocorticography recordings. Neural Computation. 2017.
13. McNally RJ, Robinaugh DJ, Wu GWY, Wang L, Deserno MK, Borsboom D. Mental disorders as causal systems: A network approach to posttraumatic stress disorder. Clinical Psychological Science. 2015.
14. Ledoit O, Wolf M. A well-conditioned estimator for large-dimensional covariance matrices. J Multivar Anal. 2004;88:365–411.
15. Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. Biostatistics. 2008.
16. Ledoit O, Wolf M. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. J Empir Financ. 2003;10:603–21.
17. Schäfer J, Strimmer K. A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and

Implications for Functional Genomics. Stat Appl Genet Mol Biol. 2005;4:1175–89. doi:10.2202/1544-6115.1175.

18. Bernal V, Bischoff R, Guryev V, Grzegorczyk M, Horvatovich P. Exact hypothesis testing for shrinkage-based Gaussian graphical models. Bioinformatics. 2019; May:1–7.

19. Schmidt-Heck W, Guthke R, Toepfer S, Reischer H, Duerrschmid K, Bayer K, et al. Reverse engineering of the stress response during expression of a recombinant protein. Proc EUNITE Symp. 2004;:10–12. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.129.539&rep=rep1&type=pdf.

20. Bottomly D, Walter NAR, Hunter JE, Darakjian P, Kawane S, Buck KJ, et al. Evaluating gene expression in C57BL/6J and DBA/2J mouse striatum using RNA-Seq and microarrays. PLoS One. 2011;6.

21. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015.

22. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. 1995;57:289–300. doi:10.2307/2346101.

23. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. 2017;:gkw937.

24. Franceschini A, Simonovic M, Roth A, Von Mering C, Szklarczyk D, Frankild S, et al. STRING v9.1: Protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 2013.

25. Berry AC. The Accuracy of the Gaussian Approximation to the Sum of Independent Variates. Trans Am Math Soc. 1941.

26. Zimmerman DW, Zumbo BD, Williams RH. Bias in Estimation and Hypothesis Testing of Correlation. Psicológica. 2003;24.

Figures

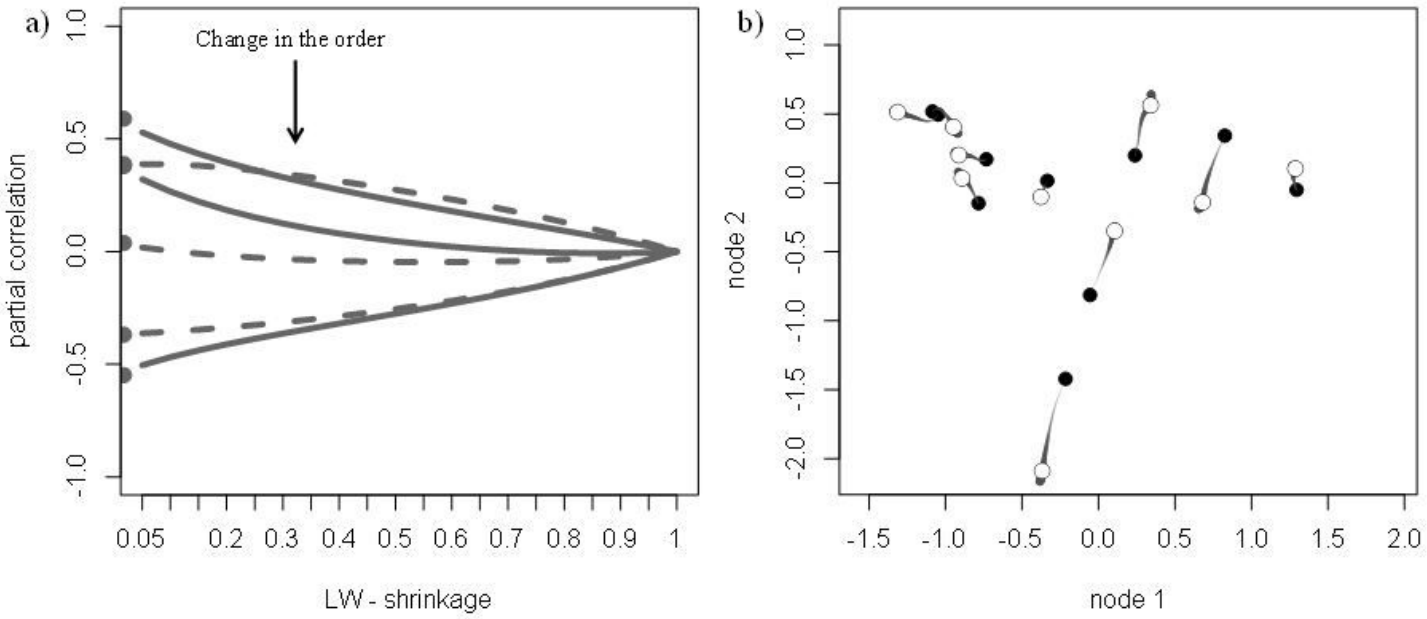

Figure 1

Non-linear effects of the shrinkage. Panel a) partial correlations obtained from Equations (3-4) while varying the shrinkage value. Lines represent 'shrunk' partial correlations, and their intersection reflects the changes in their order. Panel b) scatter plot of 'shrunk' data points (for two variables) that change as the shrinkage increases. We simulated data (n = 10) from a random network (p = 8), and the effect of the shrinkage at the data level is obtained via the Singular Value Decomposition of the data matrix, see Supplementary material S3. In black: the original data points. In grey: the data points changing their positions for λ ∈ (0,1). The dots' sizes are proportional to λ. In white: the 'shrunk' data points for the optimal shrinkage of 0.65.

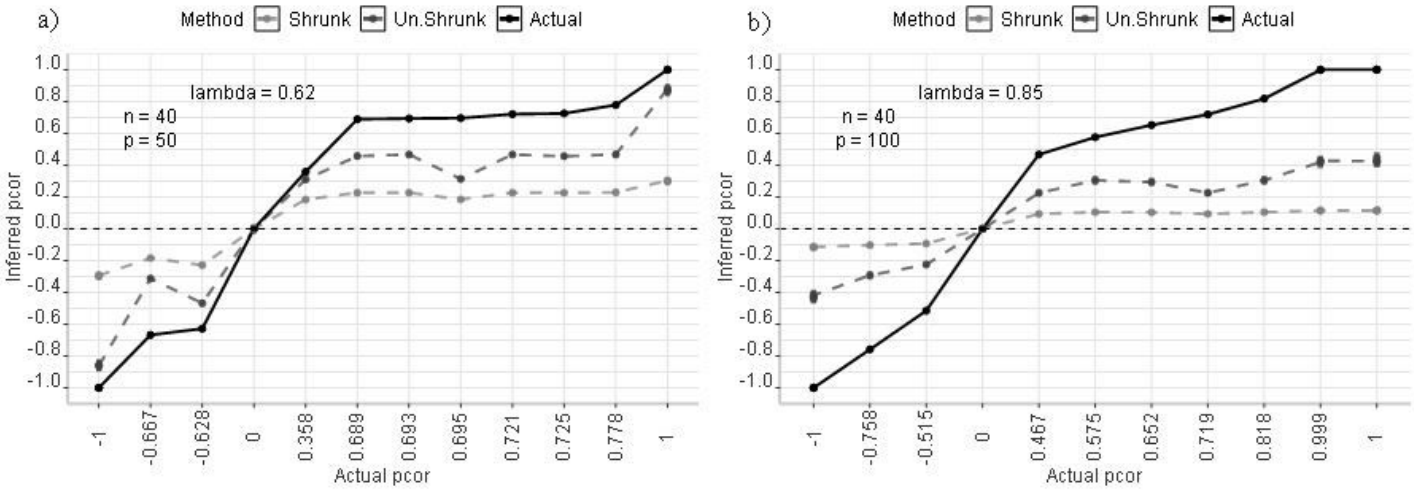

Figure 2

Partial correlations versus their actual (population) values. We simulated 25 datasets from two network structures (n= 40) and recon-structed the ‘shrunk’ and ‘un-shrunk’ partial correlations. Panel a) corresponds to p = 50, and panel b) to p = 100. In black: the actual partial correlations. In light grey (dashed): the average of the ‘shrunk’ method. In dark grey (dashed): the average of the new ‘un-shrunk’ method. Error bars represent 2 standard errors.

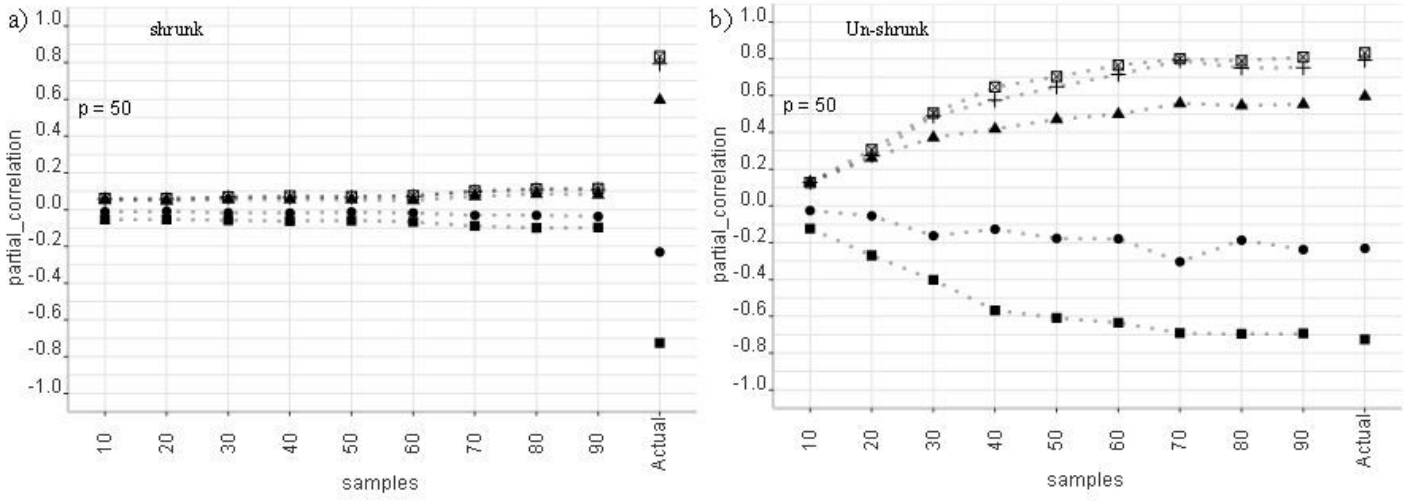

Figure 3

Partial correlation versus sample size. We create a fixed network structure for p = 50, and simulated data for sample size n ranging from 10 to 90. Panels a, b) show the average ‘shrunk’ and ‘un-shrunk’ partial correlations (over 10 simulations). The proposed ‘un-shrunk’ partial correlations are closer to their actual values.

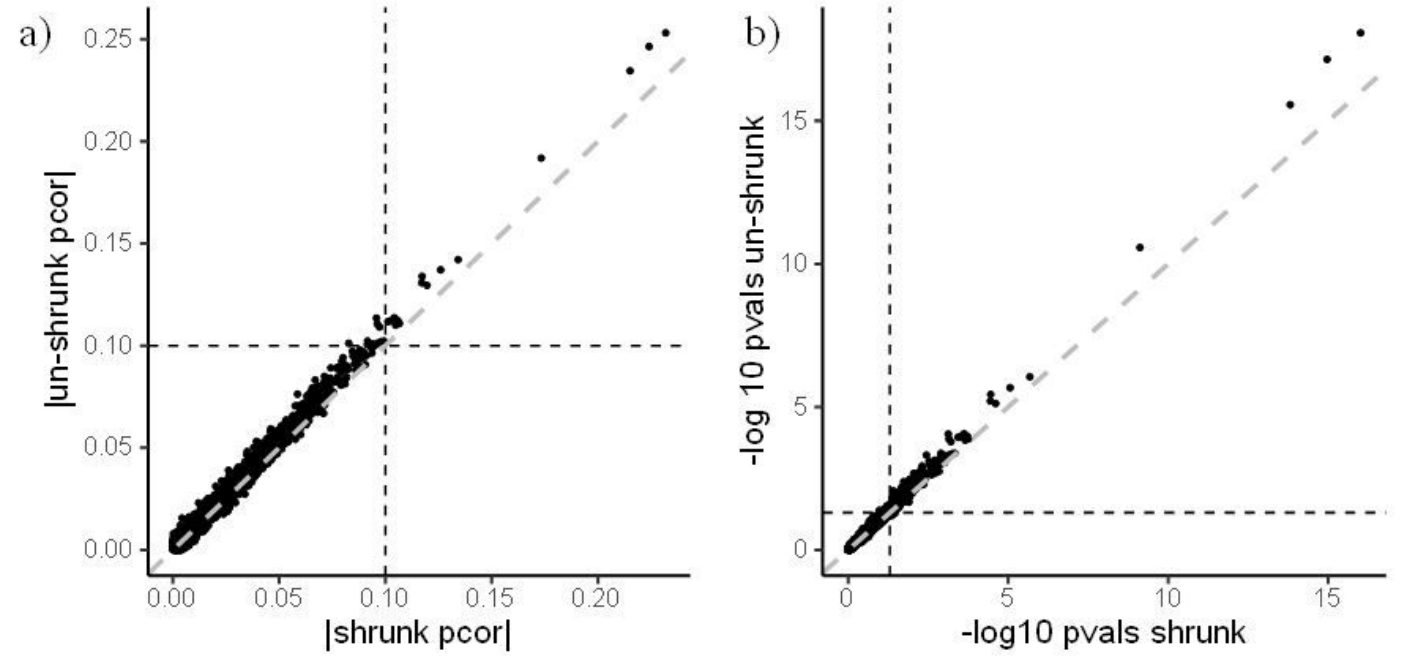

Figure 4

Comparison of partial correlations from Eschericha coli microarray data. Panel a) Scatter plot of the partial correlation in absolute value |pcorr|. Panel b) Scatter plot of −log10(p-values). P-values were computed with the shrunk approach (Bernal et al. 2019). The panels are segmented into four regions using a threshold of |pcorr| = 0.1 or p-values = 0.05 (in dashed lines). The 'un-shrunk' model has 34 additional significant edges at p-values < 0.05.

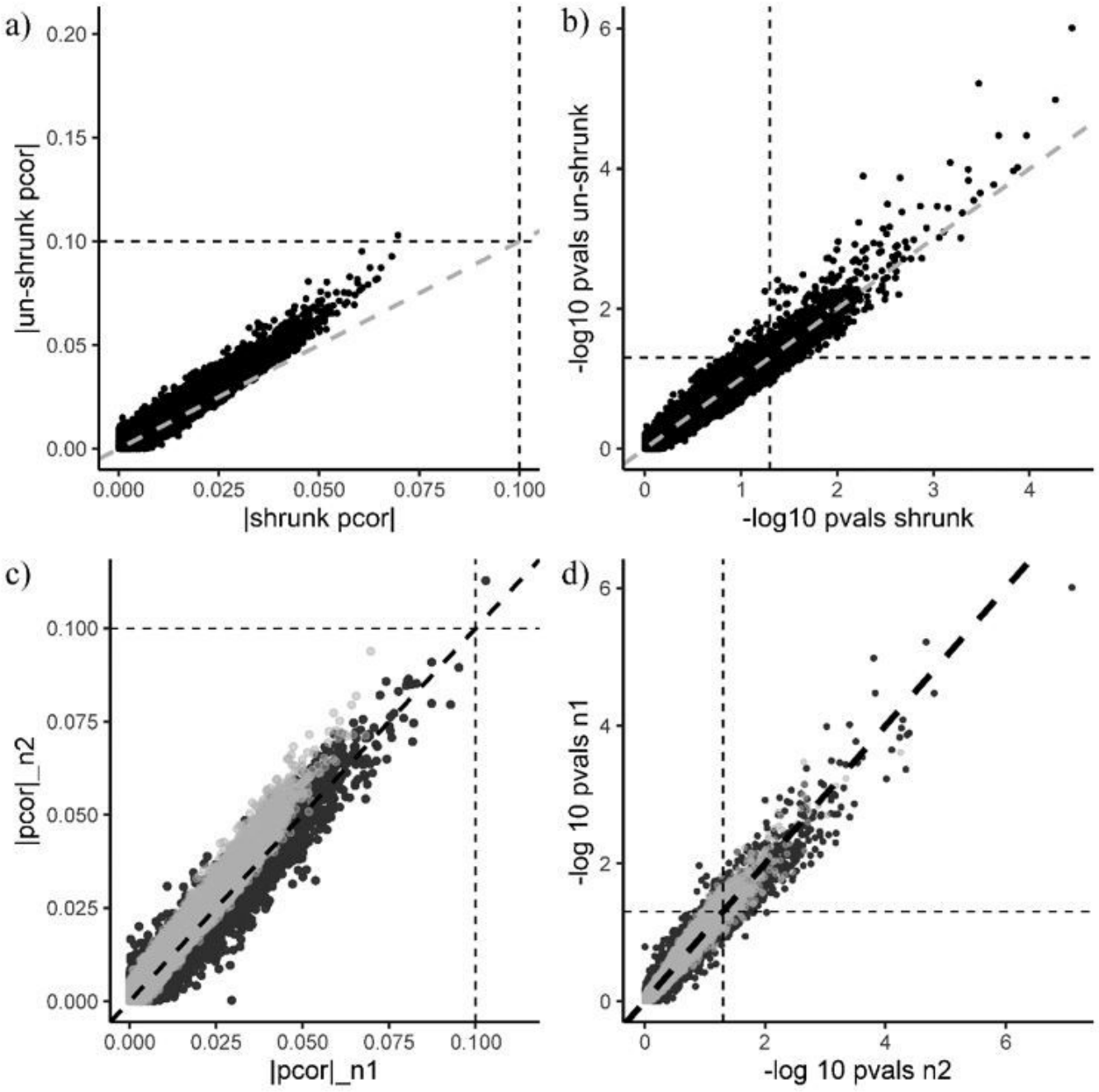

Figure 5

Comparison of partial correlations from Mus musculus RNA-seq data. Panel a) Scatter plot of the partial correlation in absolute value |pcorr|. Panel b) scatter plot of −log10 (p-values). Panel c, d) scatter plots of |pcorr|, and of −log10 (p-values) for different sample sizes (n1 = 21, n2 = 31). To generate an additional dataset (n2 = 31) from the original one (n1 = 21), we use a sampling approach. The dashed lines show

the threshold of |pcorr| = 0.1 and p-values = 0.05, respectively. P-values were computed with the shrunk approach (Bernal et al. 2019). In light grey: the ‘shrunk’ method with λ = 0.69 and 0.42. In dark grey: the ‘un- shrunk’ method. The ‘shrunk’ partial correla-tions suffer a slope shift for different sample sizes and are not comparable. The ‘un-shrunk’ partial correlations are concentrated around the diagonal, what reflects that they are comparable.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SuppUnShrinkageBMC28082020.pdf