BMC Systems Biology

Methodology article

From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data

Rainer Opgen-Rhein*[1] and Korbinian Strimmer[2]

Address: [1]Department of Statistics, Ludwig-Maximilians-Universität München, Ludwigstraße 33, D-80539 München, Germany and [2]Institute for Medical Informatics, Statistics and Epidemiology (IMISE), University of Leipzig, Härtelstr. 16-18, 04107 Leipzig, Germany

Email: Rainer Opgen-Rhein* - opgen-rhein@stat.uni-muenchen.de; Korbinian Strimmer - strimmer@uni-leipzig.de

* Corresponding author

Abstract

Background: The use of correlation networks is widespread in the analysis of gene expression and proteomics data, even though it is known that correlations not only confound direct and indirect associations but also provide no means to distinguish between cause and effect. For "causal" analysis typically the inference of a directed graphical model is required. However, this is rather difficult due to the curse of dimensionality.

Results: We propose a simple heuristic for the statistical learning of a high-dimensional "causal" network. The method first converts a correlation network into a partial correlation graph. Subsequently, a partial ordering of the nodes is established by multiple testing of the log-ratio of standardized partial variances. This allows identifying a directed acyclic causal network as a subgraph of the partial correlation network. We illustrate the approach by analyzing a large *Arabidopsis thaliana* expression data set.

Conclusion: The proposed approach is a heuristic algorithm that is based on a number of approximations, such as substituting lower order partial correlations by full order partial correlations. Nevertheless, for small samples and for sparse networks the algorithm not only yield sensible first order approximations of the causal structure in high-dimensional genomic data but is also computationally highly efficient.

Availability and Requirements: The method is implemented in the "GeneNet" R package (version 1.2.0), available from CRAN and from http://strimmerlab.org/software/genets/. The software includes an R script for reproducing the network analysis of the *Arabidopsis thaliana* data.

Background

Correlation networks are widely used to explore and visualize high-dimensional data, for instance in finance [1-3], ecology [4], gene expression analysis [5,6], or metabolomics [7]. Their popularity is owed to a large extent to the ease with which a correlation network can be constructed, as this requires only two simple steps: i) the computation of all pairwise correlations for the investigated variables, and ii) a thresholding or filtering procedure [8] to identify significant correlations, and hence edges, of the network.

However, for shedding light on the causal processes underlying the observed data, correlation networks are only of limited use. This is due to the fact that correlations not only confound direct and indirect associations but also provide no means to distinguish between response variables and covariates (and thus between cause and effect).

Therefore, causal analysis requires tools different from correlation networks: much of the work in this area has focused on Bayesian networks [9] or related regression models such as systems of recursive equations [10,11] or influence diagrams [12]. All of these models have in common that they describe causal relations by an underlying directed acyclic graph (DAG).

There already exist numerous methods for learning DAGs from observational data – see for instance the summarizing review in [13] and the references therein. However, with few exceptions [e.g., the PC algorithm, [14,15]] virtually all of these methods have been devised for comparatively small numbers of variables and with large sample size in mind. For instance, the numerical example of the recently proposed algorithm described in [16] uses n = 10,000 observations for p = 7 variables. Unfortunately, the data that would be most interesting to explore with causal methods, namely those commonly visualized by correlation networks (see above), have completely different characteristics, in particular they are likely of high dimension.

In this paper we follow [15] and focus on modeling large-scale linear recursive systems. Specifically, we present a simple discovery algorithm that enables the inference of causal relations from small sampled data and for large numbers of variables. It proceeds in two steps as follows:

• First, the correlation network is transformed into a partial correlation network, which is essentially an undirected graph that displays the direct linear associations only. This type of network model is also known under the names of graphical Gaussian model (GGM), concentration graph, covariance selection graph, conditional independence graph (CIG), or Markov random field. Note that there is a simple relationship between correlation and partial correlation. Moreover, in recent years there has been much progress with regard to statistical methodology for learning large-scale partial correlation graphs from small samples [e.g., [17-22]]. Here we employ the approach described in [20].

• Second, the undirected GGM is converted into a *partially* directed graph. This is done by estimating a pairwise ordering of the nodes from the data using multiple testing of the log-ratios of standardized partial variances, and by subsequent projection of this partial ordering onto the

GGM. The inferred causal network is the subgraph containing all the directed edges.

Note that this algorithm is similar to the PC algorithm in that edges are being removed from the independence graph to obtain the underlying DAG. However, our criterion for eliminating an edge is distinctly different from that of the PC algorithm.

The remainder of the paper is organized as follows. First, we describe the methodology. Second we consider its statistical interpretation and further properties. Subsequently, we illustrate the approach by analyzing an 800 gene data set from a large-scale *Arabidopsis thaliana* gene expression experiment. Finally, we conclude with some discussion of the method, commenting also on the limitations of the approach.

Methods
Theoretical basis
Consider a linear regression with Y as response and X_1, ..., X_k, ..., X_K as covariates. We assume that X_k and Y are random variables with known variances var(Y) and var(X_k) and with covariance cov(Y, X_k). The best linear predictor of Y in terms of the X_k that minimizes the MSE of $\sum_k \beta_k X_k$ - Y is given by [e.g. ref. [23], p. 206]

$$\beta_k^y = \tilde{\rho}_{yk} \sqrt{\frac{\tilde{\sigma}_y^2}{\tilde{\sigma}_k^2}}, \qquad (1)$$

where $\tilde{\rho}_{yk}$ and is the *partial* correlation between Y and X_k, and $\tilde{\sigma}_y^2$ and $\tilde{\sigma}_k^2$ are the respective *partial* variances.

The partial correlation is the correlation that remains between two variables if the effect of the other variables has been regressed away. Likewise, the partial variance is the variance that remains if the influences of all other variables are taken into account. Table 1 lists the definitions and formulas for the computation of these quantities (note that in our notation a tilde on top of a symbol indicates "partial").

From Equation 1 it is immediately clear that the complete linear system and thus all β_k^y are determined by the joint covariance matrix of Y and X_k [see also, e.g., [24,24]]. For only a single dependent variable Equation 1 reduces to the well-known relation $\beta_x^y = \rho_{yx} \sqrt{\sigma_y^2 / \sigma_x^2}$, which contains only the unconditioned correlation and variances (without the tilde).

Table 1: Formulas for computing partial variances and partial correlations

	Definition	True value	Estimate
Covariance matrix:	$\mathrm{cov}(X_k, X_l) = \sigma_{kl}$	$\Sigma = (\sigma_{kl})$	$\boldsymbol{S} = (s_{kl})$
Concentration matrix:	$\Omega = \Sigma^{-1}$	$\Omega = (\omega_{kl})$	
Variances:	$\mathrm{var}(X_k) = \sigma_{kk} = \sigma_k^2$	σ_{kk}	s_{kk}
Partial variances	$\mathrm{var}(X_k\|X_{\neq k}) = \tilde{\sigma}_{kk} = \tilde{\sigma}_k^2 = \omega_{kk}^{-1}$	$\tilde{\sigma}_{kk}$	$\tilde{s}_{kk}$
Correlations:	$\mathrm{corr}(X_k, X_l) = \rho_{kl} = \sigma_{kl}(\sigma_{kk}\,\sigma_{ll})^{-1/2}$	$\boldsymbol{P} = (\rho_{kl})$	$\boldsymbol{R} = (r_{kl})$
Partial correlations:	$\mathrm{corr}(X_k, X_l\|X_{\neq k, l}) =$		
	$\tilde{\rho}_{kl} = -\omega_{kl}(\omega_{kk}\omega_{ll})^{-1/2}$	$\tilde{P} = \left(\tilde{\rho}_{kl}\right)$	$\tilde{R} = \left(\tilde{r}_{kl}\right)$

Index i runs from 1 to n (sample size), and indices k and l run from 1 to p (dimension). A tilde denotes a "partial" quantity.

We emphasize that Equation 1 has a direct relation with the usual ordinary least squares (OLS) estimator for the regression coefficient. This is recovered if the empirical covariance matrix is plugged into Equation 1. However, note that Equation 1 also remains valid if other estimates of the covariance are used, such as penalized or shrinkage estimators (note that there is no hat on β_k^γ).

For the following it is important that Equation 1 can be further rewritten by introducing a scale factor. Specifically, by abbreviating the standardized partial variance $\tilde{\sigma}_k^2/\sigma_k^2$ by SPV$_k$, we can decompose the regression coefficient into the simple product

$$\beta_k^\gamma = \underbrace{\tilde{\rho}_{\gamma k}}_{\mathcal{A}} \underbrace{\sqrt{\frac{\mathrm{SPV}_\gamma}{\mathrm{SPV}_k}}}_{\mathcal{B}} \underbrace{\sqrt{\frac{\sigma_\gamma^2}{\sigma_k^2}}}_{\mathcal{C}} . \qquad (2)$$

Note that SPV$_\gamma$ and SPV$_k$ take on values from 0 to 1. All three factors have an immediate and intuitive interpretation:

$\mathcal{A}$: This factor determines whether there is a direct association between Y and the covariate X_k. If the partial correlation between X_k and Y vanishes, so will also the two corresponding regression coefficients β_k^γ and β_γ^k. In a partial correlation graph an edge is drawn between two nodes Y and X_k if $\mathcal{A} \neq 0$.

$\mathcal{B}$: This factor adjusts the regression coefficient for the relative reduction in variance of Y and X_k due to the respective other covariates. In the algorithm outlined below a

test of log($\mathcal{B}$) establishes the directionality of edges of a partially causal network.

$\mathcal{C}$: This is a scale factor correcting for different units in Y and X_k.

The product $\mathcal{AB} = \beta_k^\gamma \sqrt{\sigma_k^2/\sigma_\gamma^2}$ is also known as the standardized regression coefficient. Note that for computing both $\mathcal{A}$ and $\mathcal{B}$ only the correlation matrix is needed, as the variance information is already accounted for by the third factor $\mathcal{C}$.

In this context it is also helpful to recall the diverse statistical interpretations of SPV:

• SPV is the *proportion* of variance that remains (unexplained) after regressing against all other variables.

• For the OLS estimator SPV is equal to 1 - R^2, where R is the usual coefficient of determination.

• SPV is the inverse of the diagonal of the inverse of the *correlation* matrix. Thus, if there is no correlation (unit diagonal correlation matrix) the partial variance equals the variance, and hence SPV = 1.

• SPV may also be estimated by 1/VIF, where VIF is the usual variance inflation factor [cf. [26]].

Heuristic algorithm for discovering approximate causal networks

The above decomposition (Equation 2) suggests the following simple strategy for statistical learning of causal networks. First, by multiple testing of $\mathcal{A} = 0$ we determine the network topology, i.e. we identify those edges for which the corresponding partial correlation is not vanishing. Second, by subsequent multiple testing of log($\mathcal{B}$) = 0

we establish a partial ordering of the nodes, which in turn imposes a partial directionality upon the edges.

In more detail, we propose the following five-step algorithm:

1. First, it is essential to determine an accurate and positive definite estimate *R* of the correlation matrix. Only if the sample size is large with many more observations than variables ($n >> p$) the usual empirical correlation estimate will be suitable. In all other instances, the use of a regularized estimator is absolutely vital (e.g., the Stein-type shrinkage estimator of [20]) in order to improve efficiency and to guarantee positive definiteness. In addition, if the samples are longitudinal it may be necessary to adjust for autocorrelation [27].

2. From the estimated correlations we compute the partial variances and correlations (see Table 1), and from those in turn plug-in estimates of the factors $\mathcal{A}$ and $\mathcal{B}$ of Equation 2 for all possible edges. Note that in this calculation each variable assumes in turn the role of the response *Y* . An efficient way to calculate the various $\mathcal{B}$ is given by taking the square root of the diagonal of the inverse of the estimated correlation matrix, and computing the corresponding pairwise ratios.

3. Subsequently, we infer the partial correlation graph following the algorithm described in [19]. Essentially, we perform multiple testing of all partial correlation coefficients $\mathcal{A}$. Note that for high dimensions (large *p*) the null distribution of partial correlations across edges can be determined from the data, which in turn allows the adaptive computation of corresponding false discovery rates [28].

4. In a similar fashion we then conduct multiple testing of all log($\mathcal{B}$). As $\mathcal{B}$ is the ratio of two variances with the same degrees of freedom, it is implicit that log($\mathcal{B}$) is approximately normally distributed [29], with an unknown variance parameter θ. Thus, the observed $z = $ log($\mathcal{B}$) across all edges follow a mixture distribution

$$f(z) = \eta_0 N(0, \theta) + (1 - \eta_0) f_A(z). \qquad (3)$$

Assuming that most *z* belong to the null model, i.e. that most edges are undirected, it is possible to infer non-parametrically the alternative distribution $f_A(z)$, the proportion η_0, as well as the variance parameter θ – for an algorithm see [28]. From the resulting densities and distribution functions local and tail-area-based false discovery rates for the test log($\mathcal{B}$) = 0 are computed. Note that in this procedure we include all edges, regardless of the corresponding value of $\mathcal{A}$ or the outcome of the test $\mathcal{A}$ = 0.

5. Finally, a partially directed network is constructed as follows. All edges in the correlation graph with significant log($\mathcal{B}$) ≠ 0 are directed in such a fashion that the direction of the arrow points from the node with the larger standardized partial variance (the more "exogenous" variable) to the node with the smaller standardized partial variance (the more "endogenous" variable). The other edges with log($\mathcal{B}$) ≈ 0 remain undirected. The subgraph consisting of all directed edges constitutes the inferred causal network. Note that this does not necessarily include all nodes that are contained in the GGM network.

Results and discussion
Interpretation of the resulting graph
The above algorithm returns a partially directed partial correlation graph, whose directed edges form a causal network.

This procedure can be motivated by the following connection between partial correlation graph and a system of linear equations, where each node is in turn taken as a response variable and regressed against all other remaining nodes. In this setting the partial correlation coefficient is the geometric mean of β_k^γ and the corresponding reciprocal coefficient β_γ^k , i.e.

$$\sqrt{\beta_\gamma^k \beta_k^\gamma} = \left| \tilde{\rho}_{\gamma k} \right| \qquad (4)$$

[see also equation 16 of ref. [20]]. In this light, an undirected edge between two nodes A and B in a partial correlation graph may also be interpreted as bidirected edge, in the sense that A influences B and vice versa in the underlying system of regression. Therefore, the test $\mathcal{B}$ = 1 can be understood as *removing* one of these two directions, where Equation 2 suggests that only the relative variance reduction between the two involved nodes needs to be considered for establishing the final direction.

Reconstruction efficiency and approximations underlying the algorithm
Topology of the network
The proposed algorithm is an extension of the GGM inference approach of [19,20]. Its accuracy of correctly recovering the *topology* of the partial correlation graph has been established, e.g., in [30].

However, it is well known that a directed Bayesian network and the corresponding undirected graph are not necessarily topologically identical: in the undirected graph for computing the partial correlations one conditions on all other nodes whereas in the directed graph one conditions only on a subset of nodes, in order to avoid conditioning "on the future" (i.e. on the dependent nodes). Therefore, it is critical to evaluate to what extent full order partial correlations are reasonable approximations for lower order partial correlations. This has already been investigated intensively by [31] who showed that in certain situations (sparse graphs, faithfulness assumption etc.) lower order partial correlations may be used as approximate substitute of full conditional correlations. Therefore, in the proposed algorithm we adopt the very same argument but apply it in the different direction, i.e. we approximate lower order partial correlation by full order partial correlation.

Node ordering

A second approximation implicit in our algorithm concerns the determination of the ordering of the nodes, which is done by multiple testing of pairwise ratios of standardized partial variances. We have conducted a number of numerical simulations (data not shown) that indicate that for randomly simulated DAGs the ordering of the nodes is indeed well reflected in the partial variances, as expected.

However, from variable selection in linear models it is also known that the partial variance (or the related R^2) may not always be a reliable indicator for variable importance. Nevertheless, the partial ordering of nodes according to SPV and the implicit model selection in the underlying regressions is a very different procedure in comparison to the standard variable selection approaches, in which the increase or decrease of the R^2 is taken as indicator of whether or not a variable is to be included, or a decomposition of R^2 is sought [for a review see, e.g., [32]]. The distinctive feature of our procedure is that by performing all tests $\log(\mathcal{B}) \neq 0$ simultaneously we consider all p regression equations at once, even if the final feature selection occurs only locally on the level of an individual regression.

It is also noteworthy that, as we impose directionality from the less well explained variable (large SPV, "exogenous", "independent") to the one with relatively lower SPV (well explained, "endogenous", "dependent" variable), we effectively choose the direction with the relatively *smaller* regression coefficient (conditional that the corresponding partial correlation is also significant).

Further properties of the heuristic algorithm and of the resulting graphs

The simple heuristic network discovery algorithm exhibits a number of further properties worth noting:

1. The estimated partially directed network cannot contain any (partially) directed cycles. For instance, it is not possible for a graph to contain a pattern such as $A \rightarrow B \rightarrow A$. This example would imply $\text{SPV}_A > \text{SPV}_B > \text{SPV}_A$, which is a contradiction. As a consequence, the subgraph containing the directed edges only is also acyclic (and hence a DAG).

2. The assignment of directionality is transitive. If there is a directed edge from *A* to *B* and from *B* to *C* then there must also be a directed edge from *A* to *C*. Note however, that actual inclusion of a directed edge into the causal network is conditional on a non-zero partial correlation coefficient.

3. As the algorithm relies on correlations as input, causal processes that produce the same correlation matrix lead to the same inferred graph, and hence are indistinguishable. The existence of such equivalence classes is well known for SEMs [33] and also for Bayesian belief networks [34].

4. The proposed algorithm is scale-invariant by construction. Hence, a (linear) change in any of units of the data has no effect on the overall estimated partially directed network, and the implied causal relations.

5. We emphasize that the partially directed network is *not* the chain graph representing the equivalence class of the causal network that is obtained by considering only its directed edges – see [34].

6. The computational complexity of the algorithm is $O(p^3)$. Hence, it is no more expensive than computing the partial correlation graph, and thus allows for estimation of networks containing in the order of thousands and more nodes.

Analysis of a plant expression data set

To illustrate our algorithm for discovering causal structure, we applied the approach to a real world data example. Specifically, we reanalyzed expression time series resulting from an experiment investigating the impact of the diurnal cycle on the starch metabolism of *Arabidopsis thaliana* [35]. This is the same data set we used in a sister paper concerning the estimation of a vector autoregressive model [36].

The data are gene expression time series measurements collected at 11 different time points (0, 1, 2, 4, 8, 12, 13, 14, 16, 20, and 24 hours after the start of the experiment).

The corresponding calibrated signal intensities for 22,814 genes/probe sets and for two biological replicates are available from the NASCArrays repository, experiment no. 60 [37]. After log-transforming the data we filtered out all genes containing missing values and whose maximum signal intensity value was lower than 5 on a log-base 2 scale. Subsequently, we applied the periodicity test of [38] to identify the probes associated with the day-night cycle. As a result, a subset of 800 genes remained for further analysis.

In order to estimate the correlation matrix for the 800 genes described by the data set we employed the dynamical correlation shrinkage estimator of [39] as this takes account of the autocorrelation. The corresponding corre-

lation graph is displayed in Figure 1. It shows the 150 edges with the largest absolute values of correlation. This graph is very hard to interpret, the branches do not have any immediate or intuitive meaning (a complete annotation of the nodes can be found along with the dataset itself in the R package "GeneNet" [40]). For instance, there are no hubs as typically observed in biological networks [41,42].

This is in great contrast to the partially directed partial correlation graph. For this specific data set, by multiple testing of the factor $\mathcal{A}$ we identified 6, 102 significant edges connecting 669 nodes. For the second factor $\mathcal{B}$, determined whether edges are directed, the distribution of
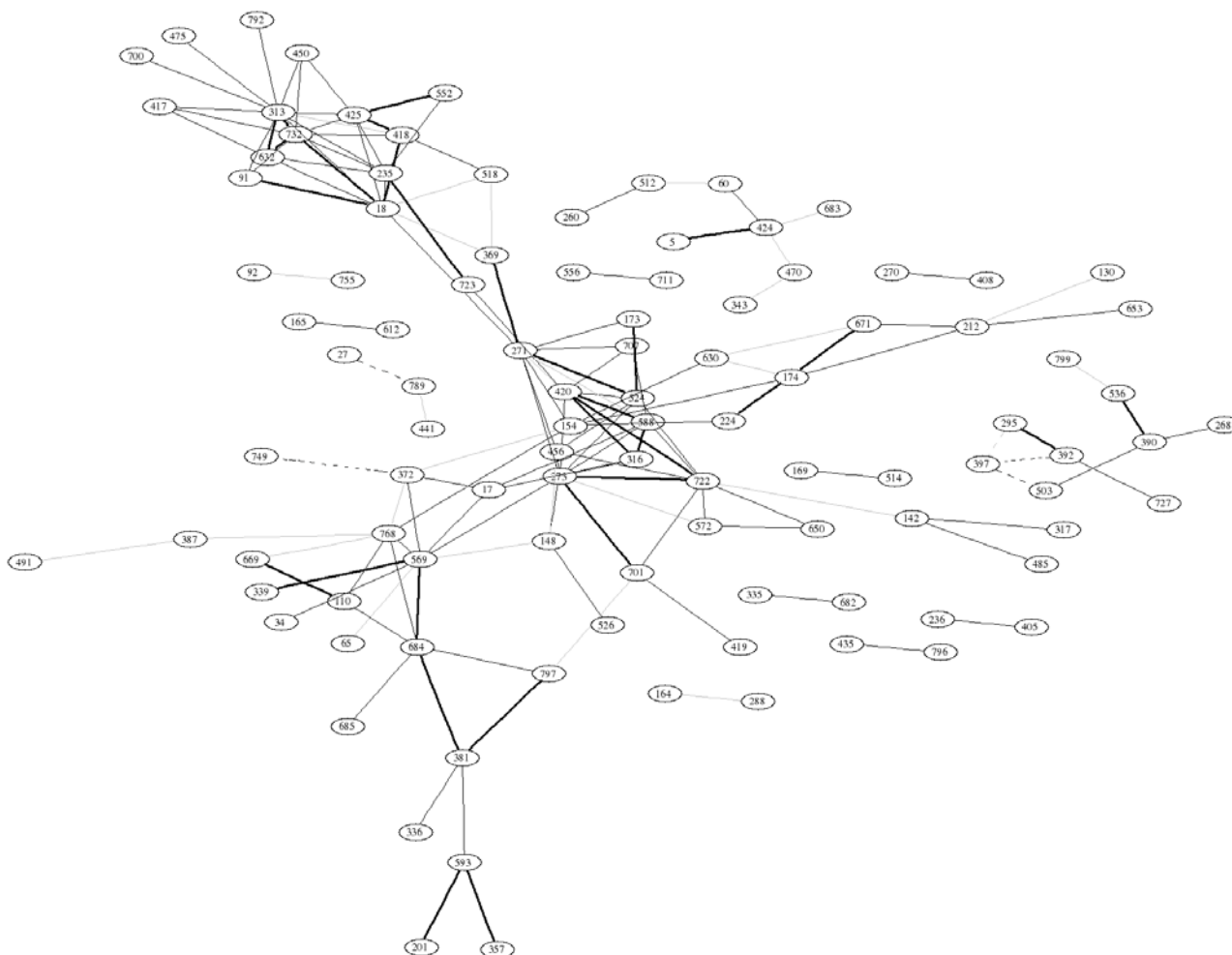

Figure 1
Correlation network inferred from the *Arabidopsis thaliana* data. The solid and dotted lines indicate positive and negative correlation coefficients, respectively, and the line intensity denotes their strength. The network displays the 150 edges with the largest absolute correlation. For annotation of the nodes in this graph see the electronic information contained in the R package "GeneNet" [40] and the original data paper [35].
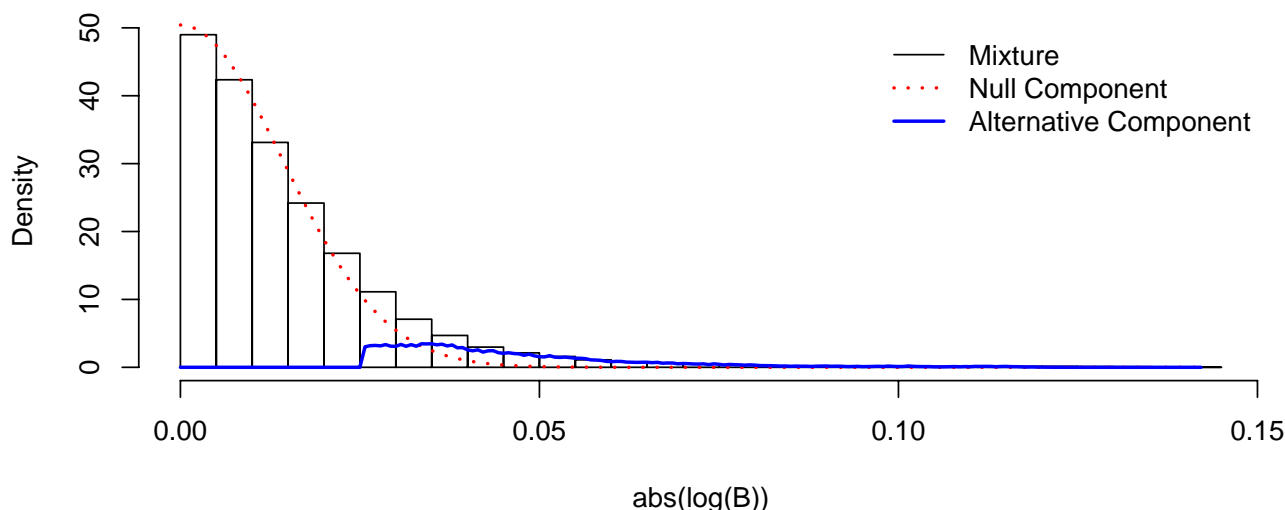
Empirical Distribution of log B

Figure 2 $\mathcal{B}$

Distribution of log $\mathcal{B}$ for the *Arabidopsis thaliana* data. The null distribution is depicted by the dashed line; it follows a normal distribution with zero mean and a standard deviation of 0.014. The solid line signifies the alternative distribution. The empirical distribution (indicated by the histogram) is composed of the null distribution (η_0 = 0.8995) and of the alternative distribution (η_A = 0.1005).

log($\mathcal{B}$) is displayed in Figure 2. The null distribution (dashed line) follows a normal distribution and characterizes the edges that cannot be directed. The alternative distribution (solid line) coincides with the directed edges. In total, we found 15, 928 significant directions.

To construct the network, we projected upon the significant edges (factor $\mathcal{A}$) the significant directions (factor $\mathcal{B}$). In the network of significant associations, 1,216 directions were significant. Note that the fraction of significant directions is by far greater in the subset of the significant partial correlations than in the complete set of all partial correlations. This agrees with the intuitive notion, that causal influences can only be attributed to existing connections between variables.

The resulting partially causal network is shown in Figure 3. For reasons of clarity we show only the subnetwork containing the 150 most significant edges, which connect 107 nodes. This graph exhibits a clear "hub" connectivity structure (nodes filled with red color). A prominent example for this is node 570, others are 81, 558, 783 and a few more genes. We see that many of the hub nodes have mostly outgoing arcs, which is indicative for key regula-tory genes. This applies, e.g., to node 570, an AP2 transcription factor, or to node 81, a gene involved in DNA-directed RNA polymerase. An interesting aspect of the partially causal network is the web of highly connected genes (colored yellow in the lower right corner of Figure 3), which we hypothesize to constitute some form of a functional module. In this module, it is not possible to determine any directions, which could be due to complex interactions among the nodes of the module. Node 627 is another hub in the network that connects the functional module with the rest of the network and which according to the annotation of [35] encodes a protein of unknown function.

We also see that the partially directed network contains both directed and undirected nodes. This is a distinct advantage of the present approach. Unlike, e.g., a vector autoregressive model [36], it does not *force* directions onto the edges.

Finally, in order to investigate the stability of the inferred partial causal network, we randomly removed data points from the sample, and repeatedly reconstructed the network from the reduced data set. In all cases the general topological structure of the network remained intact,
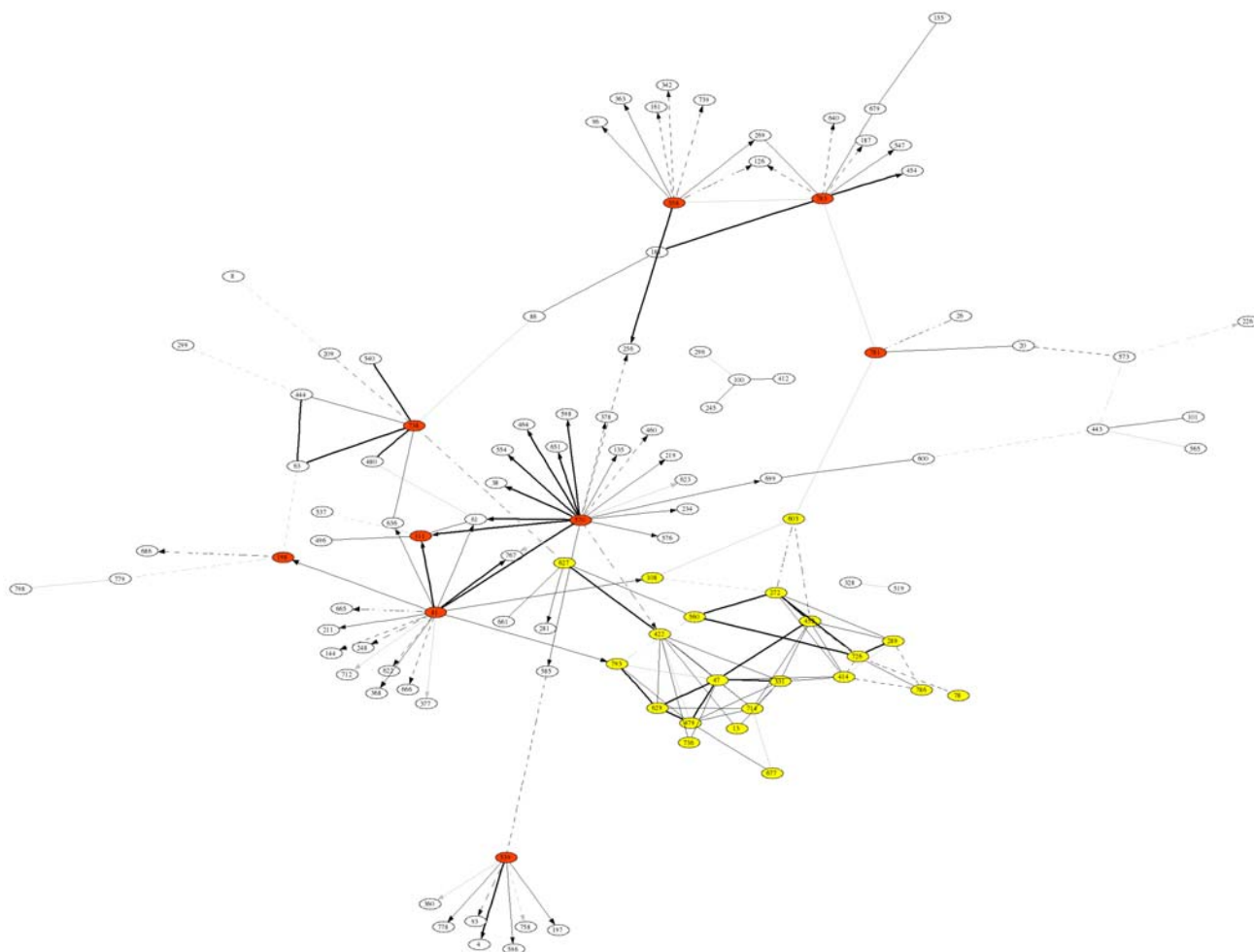
Figure 3
Partially causal network inferred from the *Arabidopsis thaliana* data by the method introduced in this paper – note the difference to the correlation network of Figure 1. The topology of the partially causal network is identical to that of a partial correlation graph (GGM, CIG). However, edges with significant directionality (as indicated by a factor $\mathcal{B}$ that is significantly smaller or larger than one) are oriented.

which indicates that this is a signal inherent in the data. This is also confirmed by the analysis using vector autoregressions [36].

Conclusion
Methods for exploring causal structures in high-dimensional data are growing in importance, particularly in the study of complex biological, medical and financial systems. As a first (and often only) analysis step these data are explored using correlation networks.

Here we have suggested a simple heuristic algorithm that, starting from a (positive definite) correlation matrix, infers a partially directed network that in turn allows gen-

erating causal hypotheses of how the data were generated. Our approach is approximate, but it allows analysis of high-dimensional small sampled data, and its computational complexity is very modest. Thus, our heuristic is likely to be applicable whenever a correlation network is computed, and therefore is suitable for screening large-scale data set for causal structure.

Nevertheless, there a several lines along which this method could be extended. For instance, non-linear effects could be accounted for by employing entropy criteria, or by using higher order moments [16]. Furthermore, more sophisticated algorithms may be used to enhance the approximation of lower order partial correlations or

the inference of the ordering of the nodes. However, ultimately this would lead to a method similar to the PC algorithm [14,15].

Note that the PC algorithm is more refined than our algorithm, primarily due to additional steps that aim at removing spurious edges (i.e. those edges that are induced between otherwise uncorrelated parent nodes by conditioning on a common child node). However, these iterative refinements may be very time consuming, in particular for high-dimensional graphs.

In contrast, our procedure is non-iterative and therefore both computationally and algorithmically (nearly) as simple as a correlation network. Nevertheless, it still enables the discovery of partially directed processes underlying the data.

In summary, we recommend our approach as a procedure for exploratory screening for causal mechanisms. Subsequently, the resulting hypotheses may then form the basis for more refined analyzes, such as full Bayesian network modeling.

Authors' contributions
Both authors participated in the development of the methodology and wrote the manuscript. RO carried out all analyzes. All authors approved of the final version of the manuscript.

Availability and requirements
The method is implemented in the "GeneNet" R package (version 1.2.0), available from CRAN and from http://strimmerlab.org/software/genets/. The software includes an R script for reproducing the network analysis of the *Arabidopsis thaliana* data.

Acknowledgements

References
1. Mantegna RN, Stanley HE: *An Introduction to Econophysics: Correlations and Complexity in Finance* Cambridge, UK: Cambridge University Press; 2000.
2. Onnela JP, Kaski K, Kertész J: **Clustering and information in correlation based financial networks.** *Eur Phys J B* 2004, **38**:353-362.
3. Boginski V, Butenko S, Pardalos PM: **Statistical analysis of financial networks.** *Comp Stat Data Anal* 2005, **48**:431-443.
4. Shipley B: *Cause and Correlation in Biology* Cambridge University Press; 2000.
5. Butte AJ, Tamayo P, Slonim D, Golub TR, Kohane IS: **Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks.** *Proc Natl Acad Sci USA* 2000, **97**:12182-12186.
6. Oldham M, Horvath S, Geschwind D: **Conservation and evolution of gene coexpression networks in human and chimpanzee brains.** *Proc Natl Acad Sci USA* 2006:17973-17978.
7. Steuer R: **On the analysis and interpretation of correlations in metabolomic data.** *Brief Bioinform* 2006, **151**:151-158.
8. Tumminello M, Aste T, Di Matteo T, Mantegna RN: **A tool for filtering information in complex systems.** *Proc Natl Acad Sc USA* 2005, **102**:10421-10426.
9. Pearl J: *Causality: Models, Reasoning, and Inference* Cambridge, UK: Cambridge University Press; 2000.
10. Freedman DA: *Statistical Models: Theory and Practice* Cambridge, UK: Cambridge University Press; 2005.
11. Wermuth N: **Linear recursive equations, covariance selection, and path analysis.** *J Amer Statist Assoc* 1980, **75**:963-972.
12. Schachter RD, Kenley CR: **Gaussian influence diagrams.** *Management Sci* 1989, **35**:527-550.
13. Tsamardinos I, Brown LE, Aliferis CF: **The max-min hill-climbing Bayesian network structure learning algorithm.** *Machine Learning* 2006, **65**:31-78.
14. Spirtes P, Glymour C, Scheines R: *Causation, Prediction, and Search* 2nd edition. MIT Press; 2000.
15. Kalisch M, Bühlmann P: **Estimating high-dimensional directed acyclic graphs with the PC-algorithm.** *J Machine Learn Res* 2007, **8**:613-636.
16. Shimizu S, Hoyer PO, Hyvärinen A, Kerminen A: **A linear non-Gaussian acyclic model for causal discovery.** *J Machine Learn Res* 2006, **7**:2003-2030.
17. de la Fuente A, Bing N, Hoeschele I, Mendes P: **Discovery of meaningful associations in genomic data using partial correlation coefficients.** *Bioinformatics* 2004, **20**:3565-3574.
18. Dobra A, Hans C, Jones B, Nevins JR, Yao G, West M: **Sparse graphical models for exploring gene expression data.** *J Multiv Anal* 2004, **90**:196-212.
19. Schäfer J, Strimmer K: **An empirical Bayes approach to inferring large-scale gene association networks.** *Bioinformatics* 2005, **21**:754-764.
20. Schäfer J, Strimmer K: **A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics.** *Statist Appl Genet Mol Biol* 2005, **4**:32.
21. Wille A, Bühlmann P: **Low-order conditional independence graphs for inferring genetic networks.** *Statist Appl Genet Mol Biol* 2006, **5**:1.
22. Li H, Gui J: **Gradient directed regularization for sparse Gaussian concentration graphs, with applications to inference of genetic networks.** *Biostatistics* 2006, **7**:302-317.
23. Cox DR, Wermuth N: **Linear dependencies represented by chain graphs.** *Statistical Science* 1993, **8**:204-218.
24. Whittaker J: *Graphical Models in Applied Multivariate Statistics* New York: Wiley; 1990.
25. Studený M: *Probabilistic Conditional Independence Structures* Springer; 2005.
26. Stewart GW: **Collinearity and least squares regression (with discussion).** *Statist Sci* 1987, **2**:68-100.
27. Opgen-Rhein R, Strimmer K: **Inferring gene dependency networks from genomic longitudinal data: a functional data approach.** *REVSTAT* 2006, **4**:53-65.
28. Efron B: **Large-scale simultaneous hypothesis testing: the choice of a null hypothesis.** *J Amer Statist Assoc* 2004, **99**:96-104.
29. Fisher RA: **On a distribution yielding the error functions of several well known statistics.** *Proc Intl Congr Math* 1924, **2**:805-813.
30. Werhli AV, Grzegorczyk M, Husmeier D: **Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical Gaussian models and Bayesian networks.** *Bioinformatics* 2006, **22**:2523-2531.
31. Castelo R, Roverato A: **A robust procedure for Gaussian graphical model search from microarray data with *p* larger than *n*.** *J Machine Learn Res* 2006, **7**:.
32. Grömping U: **Relative importance in linear regression in R: the package relaimpo.** *J Statist Soft* 2006, **17**:1.
33. Bollen KA: *Structural Equations With Latent Variables* John Wiley & Sons; 1989.
34. Chickering DM: **Learning equivalence classes of Bayesian-network structures.** *J Machine Learn Res* 2002, **2**:445-498.
35. Smith SM, Fulton DC, Chia T, Thorneycroft D, Chapple A, Dunstan H, Hylton C, Smith SCZAM: **Diurnal changes in the transcriptom encoding enzymes of starch metabolism provide evidence for both transcriptionaland posttranscriptional regulation of starch metabolism inArabidopsis leaves.** *Plant Physiol* 2004, **136**:2687-2699.

36. 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, **8(Suppl 2):**S3.
37. **NASCArrays: the Nottingham Arabidopsis Stock Centre's microarray database** [http://affymetrix.arabidopsis.info/narrays/experimentbrowse.pl]
38. Wichert S, Fokianos K, Strimmer K: **Identifying periodically expressed transcripts in microarray time series data.** *Bioinformatics* 2004, **20:**5-20.
39. Opgen-Rhein R, Strimmer K: **Using regularized dynamic correlation to infer gene dependency networks from time-series microarray data.** *Proceedings of the 4th International Workshop on Computational Systems Biology (WCSB 2006), Tampere* 2006, **4:**73-76.
40. Schäfer J, Opgen-Rhein R, Strimmer K: **Reverse engineering genetic networks using the "GeneNet" package.** *R News* 2006, **6/5:**50-53.
41. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási A-L: **Hierarchical organsation of modularity in metabolic networks.** *Science* 2002, **297:**1551-1555.
42. Barabási AL, Oltvai ZN: **Network biology: understanding the cell's functional organization.** *Nature Rev Genetics* 2004, **5:**101-113.