Published in final edited form as:

Chemometr Intell Lab Syst. 2014 November 15; 138: 193–202. doi:10.1016/j.chemolab.2014.07.002.

Constructing Metabolic Association Networks Using High-dimensional Mass Spectrometry Data

Imhoi Koo¹, Xiaoli Wei¹, Xue Shi¹, Zhanxiang Zhou², Seongho Kim³, and Xiang Zhang¹

¹Department of Chemistry, University of Louisville, Louisville, KY 40292, USA

²Department of Nutrition, University of North Carolina at Greensboro, Greensboro, North Carolina 27412, USA

³Biostatistics Core, Karmanos Cancer Institute, Department of Oncology, Wayne State University School of Medicine, Detroit, MI, USA 48201

Abstract

The goal of metabolic association networks is to identify topology of a metabolic network for a better understanding of molecular mechanisms. An accurate metabolic association network enables investigation of the functional behavior of metabolites in a cell or tissue. Gaussian Graphical model (GGM)-based methods have been widely used in genomics to infer biological networks. However, the performance of various GGM-based methods for the construction of metabolic association networks remains unknown in metabolomics. The performance of principle component regression (PCR), independent component regression (ICR), shrinkage covariance estimate (SCE), partial least squares regression (PLSR), and extrinsic similarity (ES) methods in constructing metabolic association networks was compared by estimating partial correlation coefficient matrices when the number of variables is larger than the sample size. To do this, the sample size and the network density (complexity) were considered as variables for network construction. Simulation studies show that PCR and ICR are more stable to the sample size and the network density than SCE and PLSR in terms of F1 scores. These methods were further applied to analysis of experimental metabolomics data acquired from metabolite extract of mouse liver. For the simulated data, the proposed methods PCR and ICR outperform other methods when the network density is large, while PLSR and SCE perform better when the network density is small. As for experimental metabolomics data, PCR and ICR discover more significant edges and perform better than PLSR and SCE when the discovered edges are evaluated using KEGG

© 2014 Published by Elsevier B.V. All rights reserved

Correspondence to: Seongho Kim; Xiang Zhang.

IK: imhoi.koo@louisville.edu

XW: jujuxiao@gmail.com

XS: xueshisx@gmail.com

ZZ: z_zhou@uncg.edu

SK: kimse@karmanos.org

XZ: xiang.zhang@louisville.edu

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflict of Interest: Not conflicted.

pathway. These results suggest that the metabolic network is more complex than the genomic network and therefore, PCR and ICR have the advantage over PLSR and SCE in constructing the metabolic association networks.

Keywords

Metabolomics; Gaussian graphical model; partial correlation; independent component regression; principle component regression; partial least squares regression; extrinsic similarity

1. Introduction

Metabolomics is a rapidly emerging field to systemically analyze small-molecule metabolites, which are the end products of cellular processes in a biological organism [1]. Construction of metabolic association networks is a critical data analysis step in systems biology. The metabolic association network is a collection of metabolite relations during cellular processes. In this work, we focus on methods of constructing metabolic networks that represent biochemical transformations among metabolites.

A relatively smaller number of studies have been reported for metabolic network construction. Arkin et al. [2] predicted interactions within reaction networks over time for the glycolytic pathway. Steuer et al. [3] examined the relationship between data generated from networks and biochemical pathways using potato plant metabolism. Ursem et al. [4] constructed the metabolic networks from metabolite abundance in different tomato genotypes. All of these studies used the Pearson's correlation coefficients to construct the metabolic networks. A major drawback of Pearson's correlation-based networks is unable to distinguish between the direct and the indirect associations. On the other hand, Gaussian graphical models (GGMs) reveal direct associations with conditional independence/dependence among variables, using partial correlation coefficients that are calculated by the correlation of two variables after removing the effect of other variables [5, 6]. GGMs have been employed in metabolomics for several studies. Greenberg et al. [7] used the pseudo-inverse method to estimate the partial correlation for the study of the influence of enzyme evolution on *Drosophila* metabolic pathway. Chan et al. [8] also constructed the metabolic network to quantify metabolites present in *Arabidopsis thaliana* using the first-order correlation in which the effects of only one variable are removed. Theis et al. [9] used GGMs for reconstructing pathway reactions from human population cohort when the size of samples (experiments) was larger than the number of variables (metabolites). None of these studies, however, used dimension-reduced regression to construct the network.

Reconstructing GGMs using high-dimensional data remains as a difficult task, especially when the number of variables is larger than the sample size. The standard estimation of partial correlations includes either inversion of sample covariance matrices or estimation of p least squares regression problems, where p is the number of variables. If the number of samples (observations) n is much smaller than p , these approaches are inappropriate. One alternative is to use dimension-reduced regression such as the partial least squares regression (PLSR) [10–12]. Its goal is to discover orthogonal components (score matrix) to maximize the covariance of dependent (response) and independent (predictor) variables.

Independent component and principle component regression analyses (ICR and PCR, respectively) were considered in this study, and their performance for the construction of metabolic network was compared with the performance of PLSR, shrinkage covariance estimator (SCE) [13], and extrinsic similarity (ES) [14, 15]. Note that PLSR and SCE were included in this comparison based on the previous studies [10, 11]. Although some studies have been performed to compare the performance among different GGM-based methods including PLSR and SCE [10, 12], none of these studies included PCR and ICR for network construction. PCR finds a score matrix to maximize variance of independent variables, while ICR finds it to maximize independence. It is known that these two methods will produce the same results if a normal distribution is assumed [16]. The main difference between ICR/PCR and PLSR is that ICR and PCR reduce the dimensions of data without using dependent (response) variables.

Several studies compared the performances between PCR/ICR and PLSR but not in metabolomics. For example, Dupret et al. [17] showed that PLSR performs better than ICR, while Funatsu et al. [18] verified that ICR is superior to PLSR when it was applied to a quantitative structure-property relationship analysis. Also, Wentzull and Montoto [19] reported that no significant difference is shown between PLSR and PCR in terms of prediction errors although Yenaiy and Göktas [20] urged that PLSR outperforms PCR. It still remains unclear which of these methods provides the precise output for network construction in metabolomics.

2. METHODS

2.1. Partial Correlation

The partial correlation $\rho_{XY|Z}$ between X and Y given a set of n variables $\mathbf{Z} = \{Z_1, \dots, Z_n\}$ is the correlation between the residuals R_X and R_Y resulting from the linear regression of X and Y on $\mathbf{Z}$, respectively. Partial correlation (PC) can be interpreted as the association between two random variables after eliminating the effect of a set of random variable. Consider x_i, y_i and $\mathbf{z}_i = (z_1^i, \dots, z_n^i)$ as samples of a joint probability distribution over X and Y on $\mathbf{Z}$, and assume that the multiple regression problems are

$$x_i = w_0^x + w_1^x z_1^i + \dots + w_n^x z_n^i, \quad (1)$$

$$y_i = w_0^y + w_1^y z_1^i + \dots + w_n^y z_n^i, \quad (2)$$

where $i = 1, \dots, N$. Then the least square solutions $\hat{w}_X, \hat{w}_Y$ of the regressions find the vectors to minimize the mean squared error of estimators $\hat{x}_i$ and $\hat{y}_i$ with respect to x and y , respectively. The residuals then are

$$r_{x,i} = x_i - \hat{x}_i, \quad (3)$$

$$r_{y,i} = y_i - \hat{y}_i, \quad (4)$$

and the sample partial correlation is

$$\hat{\rho}_{XY \setminus Z} = \text{Corr}(R_X, R_Y), \quad (5)$$

where, $\text{Corr}(\cdot, \cdot)$ denotes the Pearson's correlation coefficient of two random variables, $R_X = (r_{X,1}, \dots, r_{X,N})$ and $R_Y = (r_{Y,1}, \dots, r_{Y,N})$.

The problem often is that $X^T X$ is singular or ill-posed because the sample size is smaller than the number of variables. An alternative solution of this problem is to use dimension reduction methods for linear regression, which transforms the high-dimensional space into a space spanned by fewer components. Also, those methods can be applied to linear regression and machine learning approaches to increase performance [21, 22]. It is desirable that the dimension-reduced data ($\hat{p} \leq n$) can make $X^T X$ well-posed as well as increase the performance.

As mentioned before, we employ five methods to resolve this difficulty in this study, which are shrinkage covariance estimation (SCE) [13], principle component regression (PCR), partial least squares regression (PLSR) [23], independent component regression (ICR) [24], and extrinsic similarity (ES) [15]. Here SCE is a regularized approach with shrinkage intensity, while ES uses mutual information. PCR, ICR, and PLSR are dimension reduction methods with feature extraction. The differences among PCR, ICR, and PLSR are as follows: PLSR uses both dependent and independent variables to reduce data dimension, while PCR/ICR uses only independent variables, and PCR/PLSR finds orthogonal features based on the normality assumption, while ICR finds independent features based on non-normality. Several studies considered PLSR to see the performance on biological network construction [10, 11]. However, there is no study to see the effect of differences among these three approaches on network construction. For this reason, we consider PCR, ICR, and PLSR in this comparison study. Furthermore, we employ SCE as a reference based on the previous comparison study [10], and ES is also included to see the effect of mutual information.

2.2. Shrinkage Covariance Estimation

Schäfer and Strimmer [13] proposed shrinkage covariance estimator (SCE) to estimate the partial correlation when the covariance matrix Σ is singular. Under singularity of covariance matrix, an alternative method is to trade off the unbiased sample covariance Σ and low dimensional shrinkage target matrix T ;

$$\hat{\Sigma} = \lambda T + (1 - \lambda) \Sigma, \quad (6)$$

where $\lambda \in (0,1)$ is shrinkage intensity. The optimal value of the tuning parameter λ is analytically determined and estimated from the data. For a more detailed description, refer to Schäfer and Strimmer [13].

2.3. Principle Component Regression and Partial Least Squares Regression

Principle component regression (PCR) and partial least squares regression (PLSR) [23] circumvent high-dimensional problem by decomposing a data matrix X into orthogonal scores T and loadings P

$$X = TP^T + X_R, \quad (7)$$

and regressing dependent variable Y on the first r important columns $\{t_1, t_2, \dots, t_r\}$ of the scores T , where X_R is the remains of decomposition. In PCR, the orthogonal scores $T(n \times r)$ and loadings $P(p \times r)$ matrices can be calculated by applying the singular value decomposition (SVD) method to a centered data matrix X as follows:

$$X = UDP^T, \quad (8)$$

where $U(n \times r)$ and $P(p \times r)$ are orthogonal matrices corresponding to r singular values. And the scores matrix T is defined by

$$T = UD. \quad (9)$$

After choosing the optimal or suitable number of components, the first r important components of X are preserved by T . Since the matrix T is orthogonal, $T^T T$ is diagonal and nonsingular matrix. Then the coefficient β_T for a linear regression Y on the score matrix T is estimated by

$$\hat{\beta}_T = (T^T T)^{-1} T^T Y, \quad (10)$$

Thus the coefficient of the original multiple linear regression $Y = X\beta + \varepsilon$ can be expressed by

$$\hat{\beta} = PD^{-1}U^T Y. \quad (11)$$

The residual of the model for a new sample z is then calculated by

$$y - \hat{y} = (y - \bar{y}) - (z - \bar{x}) PD^{-1}U^T Y \quad (12)$$

where $\bar{y}$ and $\bar{x}$ are the sample averages.

In case of PLSR, scores and loadings are chosen in such a way to describe the covariance between X and Y as much as possible. The standard algorithm for computing score and loading matrices T and P is nonlinear iterative partial least squares (NIPALS) as follows:

Let $X_1 = X$ and $y_1 = y$, for each $j = 1, \dots, r$, where r is a given number satisfying $r \ll n$,

1. Let $w_j = X_j^T y_j / \|X_j^T y_j\|$, and store w_j into W as a column.
2. Let $t_j = X_j w_j$, and store t_j into a matrix T as a column.
3. Let $\hat{\beta}_j = t_j^T y_j / t_j^T t_j$.
4. Let $p_j = X_j^T y_j / t_j^T t_j$, and store p_j into a matrix P as a column.
5. Let $X_{j+1} = X_j - t_j p_j^T$ and $y_{j+1} = y_j - t_j \hat{\beta}_j$.

Then we have

$$\hat{X}=TP^T \text{ and } \hat{y}-T\beta=XW\left(P^TW\right)^{-1}\beta, \quad (13)$$

and the residual is written by

$$y-\hat{y}=y-XW\left(P^TW\right)^{-1}\beta. \quad (14)$$

For both PLSR and PCR using p linear regression models,

$$X^{(i)}=X_{\setminus i}^i\beta_i+\varepsilon=\sum_{j\neq i}\beta_{ij}X_j+\varepsilon, \text{ for } i=1,\cdots,p \quad (15)$$

the method described in Pihur et al. [11] is used for the estimation of the partial correlation coefficients in this work, rather than the residual-based method as expressed in the equation (5).

2.4. Independent Component Regression

Independent component regression (ICR) finds linear regression coefficients using both linear least squares method and a linear transformation of latent variables to minimize the statistical dependence of components in representation data [24, 25]. In ICR, latent variables are assumed to be non-Gaussian or mutually independent and therefore, are called independent components of observed data.

The observation data matrix X is considered to be a linear combination of independent components, i.e. $X=AS$, where columns of S contain independent components and A is a linear mixing matrix. Several methods were introduced to estimate an un-mixing matrix W and independent source signal S satisfying an equation $S=WX$. In this study, we employed fastICA [25] to obtain un-mixing matrix W and independent component S .

To extract independent component S , the fastICA searches for an un-mixing matrix W that maximizes the non-Gaussian property of sources. Non-Gaussianity for the fastICA is measured using approximation to neg-entropy (J) which is more robust than kurtosis-based measures and fast to compute. The approximation of neg-entropy of $y=w^Tx$ takes the form

$$J(y)=[E[G(y)]-E[G(v)]]^2, \quad (16)$$

where v is a random variable following a normal distribution $N(0,1)$, and w and x are the column vectors of matrices W and X , respectively. The following two choices are available for the function G :

$$G_1(u)=-\left(\frac{1}{\alpha}\right)\log\cosh(\alpha u), \quad (17)$$

$$G_2(u)=-\exp\left(\frac{u^2}{2}\right), \quad (18)$$

where $\cosh(\cdot)$ denotes the hyperbolic cosine function, and $1 - \alpha - 2$ is a constant.

Consider the following fastICA algorithm, whose goal is to find a direction, i.e., a unit vector w that maximizes the neg-entropy $J(w^T x)$. The basic form of the fastICA algorithm is as follows:

1. Choose an initial weight vector w .
2. Calculate $w^+ = E[xg(w^T x)] - E[g'(w^T x)]w$, where the function g is a derivative of the function G in the equation either (17) or (18).
3. Calculate $w = w^+ / \|w^+\|$.
4. Go to Step 2 if not converged; otherwise, stop.

Note that if the dot-product of the old and the new values of w is (almost) equal to 1, it is considered as converged.

The fastICA algorithm corresponding to unit vectors $w_1, w_2, \dots, w_n$ is performed for each unit vector to estimate several independent components $w_1^T x, w_2^T x, \dots, w_n^T x$. Furthermore, to avoid different vectors to converge the same maxima of non-Gaussianity, Gram-Schmidt method for the unit vectors $w_1, w_2, \dots, w_n$ is performed for iteration as follows:

1. $w_{p+1} = w_{p+1} - \sum_{j=1}^p \langle w_{p+1}, w_j \rangle w_j$
2. $w_{p+1} = w_{p+1} / \|w_{p+1}\|$

To calculate partial correlation coefficients, the following linear regression models are considered:

$$\begin{aligned} x_i &= \sum_{j \neq i} \beta_{ij} x_j + \varepsilon_i = X^{\setminus i} (\beta_{i,1}, \dots, \beta_{i,i+1}, \dots, \beta_{i,n})^T + \varepsilon_i \\ x_k &= \sum_{j \neq k} \beta_{kj} x_j + \varepsilon_k = X^{\setminus k} (\beta_{k,1}, \dots, \beta_{k,k+1}, \dots, \beta_{k,n})^T + \varepsilon_k \end{aligned} \quad (19)$$

where $\varepsilon_i, \varepsilon_k$ are measurement errors and

$$X^{\setminus i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n), \quad (20)$$

Once the mixing matrices $\hat{A}_i, \hat{A}_k$ and the independent component matrices $\hat{S}_i, \hat{S}_k$ are estimated by the fastICA, i.e. $X^{\setminus i} = \hat{A}_i \hat{S}_i$ and $X^{\setminus k} = \hat{A}_k \hat{S}_k$, the regression coefficients $\beta_{S,i}, \beta_{S,k}$ from $\hat{S}_i, \hat{S}_k$ are calculated by the multiple linear regression (MLR) method as follows [18]:

$$\hat{\beta}_i = (\hat{S}_i^T \hat{S}_i)^{-1} \hat{S}_i^T x_i \text{ and } \hat{\beta}_k = (\hat{S}_k^T \hat{S}_k)^{-1} \hat{S}_k^T x_k, \quad (21)$$

Since $X^{\setminus i} = \hat{A}_i \hat{S}_i$ and $X^{\setminus k} = \hat{A}_k \hat{S}_k$, the coefficients $c_{lj}^{(i)}, c_{lj}^{(k)}$ are calculated by

$$C_i = (c_{ij}^{(i)}) = (\hat{A}_i^T \hat{A}_i)^{-1} \hat{A}_i^T \text{ and } C_k = (c_{ij}^{(k)}) = (\hat{A}_k^T \hat{A}_k)^{-1} \hat{A}_k^T, \quad (22)$$

where these values are contributions of variable j -th factor of X^i , X^k to the l -th factor of $\hat{S}_i$, $\hat{S}_k$.

Then, the partial correlation coefficients $\hat{\rho}_{ik}$ are estimated by method proposed by [11] using the estimated regression coefficients $\hat{\beta}_i$, $\hat{\beta}_k$ in equation (21) and the coefficients $c_{lk}^{(i)}$, $c_{li}^{(k)}$ in equation (22)

2.5. Extrinsic Similarity

The extrinsic similarity (ES) between two variables, i and j , can be defined as

$$ES_A(i, j) = \sum_{k \in A} f(i, j, k), \quad (23)$$

where $f(i, j, k)$ denotes a function that signifies the association between variables j and j with respect to the third variables k , a set A consists of variables that contribute to ES_A of variables i and j [14, 15]. To calculate ES_A , a variable set A and an association function f must be determined. In this study, a quantile approach was used instead of a constant cutoff value for a variable set A as used in Ucar et al. [15].

2.5.1. Choice of a variable set (A)—An effective variable set N_i is defined as the neighborhood list of variables i by

$$N_i = \{j | j \in G, |r_{ij}| > \kappa_\alpha\}, \quad (24)$$

where κ_α is a quantile value satisfying $\Pr(X > \kappa_\alpha) = \alpha$; a set G denotes the set of all variables in data; $|r_{ij}|$ refers to the absolute value of the Pearson's correlation coefficient between i and j . Note that the value α in this study was set to 0.05 based on our examination for finding an optimal value. The variable set A then is designated as the intersection of their neighborhood lists, i.e. $A = N_i \cap N_j$.

2.5.2. Choice of an association function(f)—We followed the specific mutual information suggested by Ucar et al. [15] and defined two probabilities $P(i, j)$ and $P(i)$ as

$$P(i, j) = \frac{\#(N_i \cap N_j)}{\sum_{i,j} \#(N_i \cap N_j)} \text{ and } P_i = \sum_j P(i, j) \quad (25)$$

where $\#(\cdot)$ is the number of variables.

The specific mutual information is a measure of association commonly used in the information theory to infer mutual dependency. Specific mutual information of two variables X and Y is defined as $P(X, Y) / P(X)P(Y)$, given their joint probability distribution $P(X, Y)$ and marginal distributions $P(X)$ and $P(Y)$. Based on this analysis, we used the following ES measure with the association function $f(i, j, k)$, which is the absolute

difference in specific mutual information of linked variables, to quantify dissimilarity of two variables (i and j):

$$ES_A(i, j) = \sum_{k \in A} f(i, j, k) = \frac{1}{\#(A)} \sum_{k \in A} \left| \frac{P(i, k)}{P(i)P(k)} - \frac{P(j, k)}{P(j)P(k)} \right|, \quad (26)$$

This definition ensures that two variables having the same co-occurrence relations with their common neighbors are closely related to each other (ES_A value close to 0).

2.6. Optimal Number of Components and False Discovery Rate

In order for PLSR, PCR, and ICR to select the optimal number $\hat{p}$ ($1 \leq \hat{p} \leq 15$) of components, 5-fold cross-validation for simulated data was used, while the leave-one-out was used for experimental data since the sample size of experimental data is too small as shown in Table 1. As for experimental data, the range of the optimal number of components was bounded by the sample size.

Once the numerical partial correlation coefficient scores $\hat{\rho}_{ij}$ for SCE, PLSR, PCR, and ICR are obtained to measure the strength of the relationship between any two variables i and j , we need to determine which of these are statistically significant. In other words, we are faced with testing multiple hypotheses [26] given by

$$H_0: \rho_{ij} = 0 \text{ versus } H_1: \rho_{ij} \neq 0 \quad (27)$$

where $\rho_{ij} = E(\hat{\rho}_{ij})$ are the population scores [11]. To do this, we used the q-values [27, 28] for all the methods except for ES method. The cutoff value of q-value was set to 0.2 as in Kramer et al. [10].

2.7. Data

Two data sets were used in this study to compare the performance of each method on metabolic network construction. One is simulated data and the other is a set of real experimental data.

2.7.1. Simulated data—The simulated data were generated by varying two parameters, sample size and network density. The number of variables p was always set to 100. We then used nine different densities, 2.5%, 4%, 5%, 6%, 7.5%, 10%, 12.5%, 15%, and 25%, to describe the network density. Given each density, the simulated data were generated using five different sample sizes, 25, 50, 100, 150, and 200. As a result, a total of 45 cases were considered in the simulation. For each case, we generated 100 data sets and then compared the performances of each method with the averages. The R package *GeneNet* was used to generate the simulated data [29]. Note that since *GeneNet* constructs the network under a normal distribution and therefore, the simulated networks become the random network. In fact, this was verified by fitting the histograms of the degrees (number of edges for each node) as shown in Supplementary Information Figure S-16 (A) and (B). We further considered two more packages, *iGraph* [30] and *Pajek* [31], to generate scale-free networks. *iGraph* requires users to set two parameters, the number of growth of the next step and the

power of Barabasi-Albert model. In this study, we used 3 and 5 for the number of growth and 0.5 and 1 for the power, respectively. To use *Pajek*, the user needs to determine two parameters, the maximum number of true edges and the average degree of vertices, to generate the scale-free networks. The maximum number of true edges was set as 200 and 500, and the average degree of vertices was assumed as 10 and 5.

2.7.2. Experimental data—Table 1 lists the experimental data of metabolites extracted from mouse liver. The experimental data consist of all compounds detected from 12 mouse samples on a linear trap quadrupole - Fourier transform ion cyclotron resonance mass spectrometer (LTQ-FTICR-MS) via direct infusion. For the association network study, we used 94 compound peaks that were detected in all 12 samples [32].

The log-transformation of peak intensities and normalization were conducted for data preprocessing. As for normalization, the L2-norm-based approach was applied [33, 34].

2.8. Evaluation measures

The true positive (TP) is the number of elements whose true value and predicted outcome are positive, the false negative (FN) is the number of elements whose true value is positive but predicted outcome is negative, and the false positive (FP) is the number of elements whose true value is negative but predicted outcome is positive. The performances of all methods were then evaluated using the following four criteria:

- (1) The true positive rate (TPR): TPR is the proportion of the actual positives which are correctly predicted

$$TPR = \frac{TP}{TP + FN}. \quad (28)$$

- (2) The positive predictive value (PPV): PPV is the proportion of subjects with positive output results which are correctly predicted

$$PPV = \frac{TP}{TP + FP}. \quad (29)$$

- (3) F1 score: It is a measure of accuracy, which is the harmonic average of TPR and PPV

$$F_1 = \frac{2 \cdot TPR \cdot PPV}{TPR + PPV}. \quad (30)$$

- (4) The number of significant edges.

3. Results

3.1. Simulated Data

The partial correlation matrix was estimated from PLSR, PCR, ICR, SCE and ES methods using the two types of simulated data for network construction. We generated simulated data as changing sample size from 25 to 200, and true network densities from 2.5% to 25%.

Figure 1 displays the F1 scores of each method according to the four selected network densities: 2.5%, 5%, 10%, and 25% and the five sample sizes: 25, 50, 100, 150, and 200 (see the Supplementary Information Figure S-1 for all other network densities). The comparative performance of each method is dependent on the density of the true network. In the sparse network of 2.5% density, SCE and PLSR perform better than others through all sample sizes. In the complex networks from 12.5% to 25% density, PCR and ICR perform better than SCE, PLSR and ES across all sample sizes. Interestingly, if the sample size is small (25 or 50) and the network is moderate complex (4% to 10%), PCR and ICR outperform SCE and PLSR; otherwise SCE and PLSR are better than PCR and ICR in terms of F1 score. Furthermore, F1 scores of SCE, PLSR, PCR and ICR increase as the samples size increases, as expected.

Supplementary Information Figures S-2, S-3, and S-4 give the number of significant edges, TPR and PPV related to all network densities, respectively. PCR and ICR tend to have the larger numbers of significant edges, while SCE and PLSR tend to have the least complex networks. For example, in case of 5% network density, ICR and PCR construct more complex networks than PLSR and SCE do. As the sample size increases, the number of significant edges of ICR and PCR decreases, but these of PLSR and SCE increase. Compared with PCR and ICR, SCE and PLSR have much wider variation in TPR as the sample size and the network density are varied (Supplementary Information Figure S-3). Furthermore, PCR and ICR have consistently much larger TPR than SCE and PLSR when the network density is larger than or equal to 10%. However, the PPV has the opposite trend compared to TPR, as shown in Supplementary Information Figure S-4. That is, both SCE and PLSR outperform PCR and ICR across all network densities. Interestingly, SCE performs the best in terms of PPV if SCE generates no empty network.

The precision-recall (PPV-TPR), receiver operating characteristic (ROC), and areas under the curve (AUC) plots are further investigated with network density of 2.5%, 5%, 10%, 12.5%, 15% and 25% as shown in Supplementary Information Figures S-5 (A) and (B) and S-6, respectively. In the PPV-TPR plot, PCR and ICR have better recall (TPR) values than these of other methods when the sample size is small, while SCE has the largest precision (PPV) values across all the densities, as can be seen in Figure S-5 (A). Figure S-5 (B) shows that PCR and ICR have larger FPR values than other methods. This might be because significant edges are more observed with PCR and ICR. In case of AUC as depicted in Figure S-6, its value becomes less than 0.6 when the network density is more than or equal to 10%, so the performance of all the methods become slightly better than the case corresponding to random chance (AUC = 0.5). Overall, it can be seen that AUC has the similar trends to F1 scores.

It is noteworthy that, in this study, we used the Pihur et al. [11] approach to estimate the partial correlation for PLSR, PCR and ICR while the equation (5) was used for SCE using the R package *GeneNet* [29]. Note that this method is not suitable for SCE and so SCE still uses the equation (5). Supplementary Information Table S-1 shows the number of empty networks generated from PLSR and SCE out of 100 replications for each case, respectively. SCE has much more empty networks especially when the sample size is small. As for Figure 1, the empty networks were excluded for the calculation of the averages of each of the four evaluating criteria. If all the 100 replications were resulted in empty networks, we obtained no value including F1 scores. Therefore, there are no point and line in Figure 1 for SCE and PLSR in some cases.

In Table 2, we calculated the mean squared error (MSE) between the true and the estimated partial correlation coefficients for SCE, PLSR, PCR, and ICR only when the sample size was 25 since it is more practical for real-world application. The MSE of PLSR increases as the network density increases and is worse than that of SCE and PCR/ICA. The MSEs of PCR and ICR are stable than these of PLSR, indicating that PCR and ICR are less sensitive to the network density. The MSE of SCE is much smaller than these of other methods. This is because the partial correlation of SCE was calculated based on the equation (5), while these of PLSR and PCR were computed by Pihur et al. [11] approach. We will discuss the more important details of this consideration in the Discussion and Conclusions section.

The structure of the constructed networks was further evaluated by fitting to the power-law functions. The residual sum of squares for each method is shown in Supplementary Information Table S-2. Interestingly, although the simulated networks belong to the random network, the estimated networks are well fit to the scale-free network. However, it is well known that the real biological networks follow a scale-free network other than a random network [35]. For this reason, *iGraph* and *Pajek* packages were also considered to generate the scale-free networks and then each method was implemented on these simulated scale-free networks. Supplementary Information Figure S-7 shows the F1 scores of each method for the reconstruction of the scale-free networks generated by *iGraph* and *Pajek*. Because each package has its own approach to generating the network, it was unmanageable to generate the networks that have the same conditions as these of *GeneNet*. Therefore, we first generated the networks using certain configurations of each package and then checked the network density. By doing so, we obtained four cases for *iGraph* and two cases for *Pajek*, respectively, which are similar to the networks with density of 5% or 10% generated by *GeneNet*. Supplementary Information Figure S-7 depicts the results of *iGraph* and *Pajek*. From this additional simulation, no difference between the random networks and the scale-free networks is observed in terms of F1 scores.

In summary, SCE and PLSR have the similar performance in reconstruction of networks, while PCR and ICR have almost identical performance with the simulation data. In terms of F1 scores, SCE and PLSR outperform others when the network is less complex, while PCR and ICR perform better than others as the network is more complex. PCR and ICR estimate the closer partial correlation coefficients to the true partial correlation coefficients than PLSR in regard to MSE.

3.2. Experimental Data

The true networks in the experimental data are unknown. For this reason, we compared the performance of each method in terms of the number of significant edges and the number of overlapped edges, followed by verifying the reconstructed networks using a public metabolic pathway database.

In Table 4, the values in diagonal are the numbers of significant edges for each method, and the values in the upper diagonal are the numbers of overlapped significant edges in two corresponding methods. The trend of the number of significant edges of each method is similar to the simulated data with a network density of 5%, where PCR and ICR construct the largest network but ES has the smallest one, although the densities of the constructed networks range from 0.47% to 2.8%. As expected, ES constructed the smallest network. There are 38 edges in common between SCE and PCR/ICR, 54 edges in common between PLSR and PCR/ICR, and 43 edges in common between SCE and PLSR. These indicate that each method constructed the different networks from each other except for the networks constructed by PCR and ICR.

Since the whole network is large, we considered the sub-networks of each method with all the nodes that have one or more edges for each constructed network. Note that PCR and ICR constructed the exactly identical networks. These sub-networks are displayed in Figure 2. Their whole networks can be found in Supplementary Information Figure S-8 for the experimental data sets identified by *MetSign* [32]. As observed in Figure 2, PCR and ICR constructed the most complex network, while ES constructed the least complex network, which is consistent with the simulated data. In addition, the network topology of the estimated networks for each method was investigated. Similar to the simulated data, the estimated networks are the scale-free network as depicted in Supplementary Information Figure S-9.

Each constructed network was further evaluated using Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (www.genome.jp/kegg). In this case, all the edges of each constructed network were verified to see if the discovered edge is present in the KEGG pathway database. SCE, PLSR, PCR/ICR and ES discovered 34, 39, 67 and 9 edges that present in the KEGG pathway database, respectively. Figure 3 shows the Venn diagram for the edges among the four constructed networks that present in the KEGG pathway database. Among all the edges discovered by the 5 algorithms (164 unique edges), 32 (47.8%) edges are discovered by only PCR/ICR, while 5 (14.7%) and 3 (7.7%) edges by SCE and PLSR, respectively. As a consequence, the PPVs for SCE, PLSR and PCR/ICR are 50.0% (= 34/68), 55.7% (= 39/70), and 54.6% (= 67/123), respectively, assuming that the edges found in the KEGG pathway database are true. The network constructed using the edges present in the KEGG pathway database is displayed in Figure 4. The compound names and edges discovered in the KEGG pathway database are listed in Supplementary Information Tables S-3 and S-4. Interestingly, most of the edges found by ES are related to fat digestion and absorption pathway.

4. DISCUSSION AND CONCLUSIONS

We compared the performance of PCA, ICA, PLSR, SCE, and ES for the construction of metabolic association networks using both the simulated and real biological data. As for the simulated data, the performance of each method depends on network density and sample size. In a sparse network, F1 scores of SCE and PLSR are better than these of PCR, ICR and ES. In a moderate network, PCR and ICR are superior to SCE and PLSR when the sample size is small while SCE and PLSR outperform PCR and ICR in case of large sample size. PCR and ICR perform the best in all sample sizes when the network is complex.

For PCR and ICR, all the performance measures are very similar to each other regardless of the sample size and the network density in the simulated data. This observation agrees with the literature reported results [16, 36]. The main difference between PCR and ICR is that PCR maximizes variance of data, while ICR maximizes the independence of latent variables. However, under the Gaussian normal distribution, the components of PCR are mutually uncorrelated so that these components become independent. For this reason, it is natural that these two methods have the same performance with the simulated data. One interesting trend is that PCR and ICR generate exactly the same network even for the experimental data. It might be because of the log-transformation of the data to make the data look like a symmetric distribution.

In this study, the estimation method of partial correlation coefficient for PLSR, PCR, and ICR was replaced with Pihur et al. [11] approach. The estimates using the Pihur's approach are two or more times larger than these using the equation (5), resulting in larger difference from the true coefficients. For this reason, the MSEs of PLSR and PCR/ICR in this study are much greater than that of SCE in Table 2.

Supplementary Information Figure S-10 depicts the variation between the true and the estimated partial correlations for a network with density of 5% and sample size of 25, which has a total of 248 true nonzero partial correlations. Compared with the other approaches, the estimated partial correlations with a true correlation value of zero (i.e., no correlation) have much narrower variation in case of SCE, indicating that the zero true partial correlations are better estimated by SCE (Figure S-10a). However, the nonzero true partial correlations are better estimated by the other approaches (Figure S-10b, S-10c, S-10d). 95% of the true partial correlations have true correlation values of zero, which generate a very small MSE value for the SCE approach. On the other hand, the F1 score of SCE becomes worse because the true partial correlations close to either -1 or 1 are much worse estimated by SCE. For these reasons, SCE has a much smaller value of MSE but a small value of F1 score.

The effect of FDR on the number of empty networks was investigated by employing the Benjamini-Hochberg FDR method (BH-FDR) [26], as shown in Supplementary Information Table S-3. We can see that the trend of BH-FDR method is generally similar to that of the q-value method used in this paper, meaning that the number of empty networks of SCE decreases as the sample size increases. It should be noted that the number of empty networks of PLSR is zero in case of BH-FDR method. The reason is that the BH-FDR method generally constructs about ten empty networks less than the q-value method, but the number

of empty networks of PLSR in Table S-1 is less than ten. Overall, there is no difference between the BH-FDR and q-value method in terms of the number of empty networks.

In analyzing the simulated data, PLSR and SCE outperform PCR and ICR when the network density is small ($d=2.5\%$), while PCR and ICR perform better when the network density is large ($d>5\%$) as shown in Figure 1. Such different performance might be that PCR and ICR use only predictor (independent) variables while PLSR use both predictor (independent) and response (dependent) variables for dimension reduction. Therefore, PLSR can be considered as supervised learning-based methods, while PCR and ICR as unsupervised learning-based method [37].

In this study, the random network was mainly used to compare the performance of each method for reconstructing the networks. However, since the real biological networks follow a scale-free network other than a random network, we further considered the scale-free networks using two existing packages, *iGraph* and *Pajek*, when the network density is near 5%. The results demonstrate that there is no difference in analyzing the performance of SCE, PLSR, PCR and ICR by either using a random network or a scale-free network. SCE and PLSR perform better than PCR and ICR when the sample size is large, while PCR and ICR outperform others when the sample size is small (Supplementary Information Figure S-7). The main reason why no difference was found in using a random network and a scale-free network is that the number of significant edges does not depend on the structure (or topology) of the network but on the strength of association. In other words, no matter what the topology of a network looks like, the FDR performs the calculation or estimation only based on the estimated partial correlation.

In order to detect the significant edges, FDR is employed to each method except for the ES method. FDR requires either a statistic, such as z-, t-statistic, or p-value. However, it is difficult for ES method to infer a statistic due to its complicated expression as described in the equation (26). Moreover, when comparing the distributions of the estimated partial correlations of SCE, PLSR, PCR, and ES methods as depicted in Supplementary Information Figures S-11 (a)–(c) and S-12, respectively, we found that the distribution of ES method is essentially different from these of the other methods. The majority of the estimated ES values are near to zero, while other methods have the majority of the estimated values near to zero. A zero value of the ES method represents a strong association while a zero value in other methods represents no association. Therefore, we implemented the ES method with two more thresholds, $ES = 0.001$ and $ES = 0.1$, along with $ES = 0$, to see the effect of the cut-off value on the performance. These results are displayed in Supplementary Information Figures S-1, S-2, S-3, S-4, and S-7. It can be seen that the overall F1 value increases as the threshold increases, but its performance is still worse than the other methods. On the other hand, it often happens that the denominator of the fractions in the equation (26) is zero, causing that the value of ES method becomes infinity. In this study, we considered the edges having the value of ES method of infinity as non-significant edges and removed them for further analysis. To circumvent this difficulty, we investigated the performance of ES method by adding a small number 10^{-10} into the probability of $P(i, j)$ in the equation (25) and the results with a small number are shown in Supplementary Information Figures S-13, S-14, and S-15. However, in terms of F1 value, we found that the

ES method without adding the small number is better than the method with the small number. The method of finding a statistic for FDR and how to deal with infinity are left as future work.

The area under the ROC curve (AUC) was further evaluated for each method as shown in Supplementary Information Figure S-6. Except for the density of 2.5%, the AUCs of all the methods become less than 60% when the sample size is less than or equal to 50. Moreover, when the density is 25%, the AUCs of the sample size of 25 and 50 are close to 50%, which is the case corresponding to random chance. In other words, this fact addresses that the current state of reverse engineering in network reconstruction is slightly better than the random chance in terms of AUC.

As for experimental data, PCR and ICR discovered more significant edges than SCE and PLSR so that PCR and ICR constructed more complex network than others. This is consistent with the results from the simulated data. In fact, although F1 score of PCR and ICR is larger than that of SCE and PLSR, the PPVs of SCE and PLSR are in general larger than that of PCR and ICR in the simulated data analysis as seen in Figure 1. Therefore, if the objective of the network construction is to discover individual pathways rather than to investigate the global systems of pathways, SCE and PLSR can be of a choice. However, the performance with experimental data shows that there are more edges present in the KEGG pathway database in the network constructed by PCR and ICR than SCE and PLSR, and the predicted PPV of PCR and ICR is very similar to that of SCE and PLSR (54.57% for PCR/ICR, 50.00% for SCE, and 55.71% for PLSR), assuming that the edges found in the KEGG pathway database are true. This indicates that the density of the true metabolic network of this experimental data might be suited to PCR and ICR considered as density of 5% with small sample size. However, further research is needed to better understand and fully answer how complex the metabolite network is.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by National Institute of Health (NIH) grant 1RO1GM087735 through the National Institute of General Medical Sciences (NIGMS) and DMS-1312603 through the National Science Foundation (NSF). The Biostatistics Core is supported, in part, by NIH Center grant P30 CA022453 to the Karmanos Cancer Institute at Wayne State University.

References

- [1]. Watkins SM, German JB, Hammock BD. Metabolomics: building on a century of biochemistry to guide human health. *Metabolomics*. 2005; 1:3–9. [PubMed: 16680201]
- [2]. Arkin A, Shen PD, Ross J. A test case of correlation metric construction of a reaction pathway from measurements. *Science*. 1997; 277:1275–1279.
- [3]. Steuer R, Kurths J, Fiehn O, Weckwerth W. Observing and interpreting correlations in metabolomic networks. *Bioinformatics*. 2003; 19:1019–1026. [PubMed: 12761066]
- [4]. Ursem R, Tikunov Y, Bovy A, van Berloo R, van Eeuwijk F. A correlation network approach to metabolic data analysis for tomato fruits. *Euphytica*. 2008; 161:181–193.

- [5]. Dobra A, Hans C, Jones B, Nevins JR, Yao GA, West M. Sparse graphical models for exploring gene expression data. *J Multivariate Anal.* 2004; 90:196–212.
- [6]. Whittaker, J. Graphical models in applied multivariate statistics. Wiley; Chichester: 1990.
- [7]. Greenberg AJ, Stockwell SR, Clark AG. Evolutionary Constraint and Adaptation in the Metabolic Network of *Drosophila*. *Mol Biol Evol.* 2008; 25:2537–2546. [PubMed: 18799713]
- [8]. Chan EKF, Rowe HC, Hansen BG, Kliebenstein DJ. The Complex Genetic Architecture of the Metabolome. *Plos Genet.* 2010; 6
- [9]. Theis FJ, Krumsiek J, Suhre K, Illig T, Adamski J. Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. *Bmc Syst Biol.* 2011; 5
- [10]. Kramer N, Schafer J, Boulesteix AL. Regularized estimation of large-scale gene association networks using graphical Gaussian models. *Bmc Bioinformatics.* 2009; 10
- [11]. Pihur V, Datta S, Datta S. Reconstruction of genetic association networks from microarray data: a partial least squares approach. *Bioinformatics.* 2008; 24:561–568. [PubMed: 18204062]
- [12]. Tenenhaus A, Guillemot V, Gidrol X, Frouin V. Gene Association Networks from Microarray Data Using a Regularized Estimation of Partial Correlation Based on PLS Regression. *Ieee Acm T Comput Bi.* 2010; 7:251–262.
- [13]. Schafer J, Strimmer K. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat Appl Genet Mo B.* 2005; 4
- [14]. Das, G.; Mannila, H.; Ronkainen, P. Knowledge Discovery and Data Mining. 1998. Similarity of attributes by external probes; p. 23-29.
- [15]. Ucar, D.; Altıparmak, F.; Ferhatosmanoglu, H.; Parthasarathy, S. Mutual Information Based Extrinsic Similarity for Microarray Analysis. In: Rajasekaran, S., editor. *Bioinformatics and Computational Biology*. Springer; Berlin / Heidelberg: 2009. p. 424-436.
- [16]. Chechile RA. Independent component analysis: A tutorial introduction. *J Math Psychol.* 2005; 49:426–426.
- [17]. Dupret, Y.; Perrin, E.; Grolier, JL.; Kielbasa, R. Comparison of three different methods to model the semiconductor manufacturing yield. *Advanced Semiconductor Manufacturing Conference and Workshop, 2005 IEEE/SEMI*; 2005. p. 118-123.
- [18]. Funatsu K, Kaneko H, Arakawa M. Development of a new regression analysis method using independent component analysis. *J Chem Inf Model.* 2008; 48:534–541. [PubMed: 18321087]
- [19]. Wentzell PD, Montoto LV. Comparison of principal components regression and partial least squares regression through generic simulations of complex mixtures. *Chemometr Intell Lab.* 2003; 65:257–279.
- [20]. Yeniay Ö, Göktas A. A comparison of partial least squares regression with other prediction methods. *Hacettepe Journal of Mathematics and Statistics.* 2002; 31:111.
- [21]. Antoniadis A, Lambert-Lacroix S, Leblanc F. Effective dimension reduction methods for tumor classification using gene expression data. *Bioinformatics.* 2003; 19:563–570. [PubMed: 12651713]
- [22]. Fukumizu K, Bach FR, Jordan MI. Dimensionality reduction for supervised learning with reproducing kernel Hilbert spaces. *J Mach Learn Res.* 2004; 5:73–99.
- [23]. Mevik BH, Wehrens R. The pls package: Principal component and partial least squares regression in R. *J Stat Softw.* 2007; 18
- [24]. Comon P. Independent Component Analysis, a New Concept. *Signal Process.* 1994; 36:287–314.
- [25]. Hyvarinen A, Oja E. Independent component analysis: algorithms and applications. *Neural Netw.* 2000; 13:411–430. [PubMed: 10946390]
- [26]. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met.* 1995; 57:289–300.
- [27]. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *P Natl Acad Sci USA.* 2003; 100:9440–9445.
- [28]. Strimmer K. A unified approach to false discovery rate estimation. *Bmc Bioinformatics.* 2008; 9
- [29]. Schafer J, Strimmer K. An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics.* 2005; 21:754–764. [PubMed: 15479708]

- [30]. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal Complex Systems*. 2006; 1695:1695.
- [31]. Batagelj V, Mrvar A. Pajek-program for large network analysis. *Connections*. 1998; 21:47–57.
- [32]. Wei X, Sun W, Shi X, Koo I, Wang B, Zhang J, Yin X, Tang Y, Bogdanov B, Kim S, Zhou Z, McClain C, Zhang X. MetSign: A Computational Platform for High-Resolution Mass Spectrometry-Based Metabolomics. *Anal Chem*. 2011; 83:7668–7675. [PubMed: 21932828]
- [33]. Crawford LR, Morrison JD. Computer Methods in Analytical Mass Spectrometry - Development of Programs for Analysis of Low Resolution Mass Spectra. *Anal Chem*. 1971; 43:1790.
- [34]. Oresic M, Sysi-Aho M, Katajamaa M, Yetukuri L. Normalization method for metabolomics data using optimal selection of multiple internal standards. *Bmc Bioinformatics*. 2007; 8
- [35]. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL. The large-scale organization of metabolic networks. *Nature*. 2000; 407:651–654. [PubMed: 11034217]
- [36]. Baek K, Draper BA, Beveridge JR, She K. PCA vs. ICA: A comparison on the FERET data set, in, *Citeseer*. 2002:824–827.
- [37]. Ramadan Z, Jacobs D, Grigorov M, Kochhar S. Metabolic profiling using principal component analysis, discriminant partial least squares, and genetic algorithms. *Talanta*. 2006; 68:1683–1691. [PubMed: 18970515]

Highlights

- We propose PCR and ICR-based methods for metabolic network construction.
- We examine the performance of various network construction methods in metabolomics.
- PCR and ICR outperform other methods when the network density is large
- PLSR and SCE perform better when the network density is small.
- PCR and ICR have the advantage over PLSR and SCE in metabolic network construction.

Figure 1. Plots of F1 score, TPR, PPV with respect to the sample size and the network densities of 2.5%, 5%, 10%, and 25%

Each column stands for a density of 2.5%, 5%, 10%, and 25%, respectively. Row plots are related to the performance measures of F1 score, TPR, and PPV. Each horizontal-axis represents the sample size of 25, 50, 100, 150, and 200.

Figure 2. Comparison of estimated metabolic association network using FT-MS data

Each plot just depicts the metabolites with significant edges constructed by each method using the 92 experimentally measured metabolites. Full name of each node is shown in Supplementary Information Table S-5. This plot was generated with the Cytoscape application.

Figure 3. Venn diagram of edges of SEC, PLSR, PCR/ICR, and ES using the experimental data
The edges used in the Venn diagram are all present in the KEGG pathway database.

Figure 4. The KEGG pathway of SCE, PLSR, PCR/ICR and ES for the experimental data
 The nodes used in the plot are selected from the KEGG pathway database. The number in circle is the index of metabolite appeared in Supplementary Information Table S-5. The detail pathways related to the edges are described in Supplementary Information Table S-6 (a) and (b). This plot was generated with the Cytoscape application.

Table 1

The information of experimental data. The value in parentheses describe for the number of nodes and full edges used in Figure 2.

Data set	Type of instrument	Samples	Metabolite peaks	Size of full graph
Mouse liver	LTQ-FTICR/MS	12	94 (76)	4371 (2850)

Table 2

Mean squared error (MSE) for SCE, PLSR and PCR with sample size of 25 and several network densities.

	MSE (mean $\pm$ standard deviation)			
	2.5%	5%	10%	25%
SCE	0.0066 $\pm$ 0.0003	0.0045 $\pm$ 0.0002	0.0026 $\pm$ 0.0001	0.0012 $\pm$ 0.0001
PLSR	0.0483 $\pm$ 0.0133	0.1132 $\pm$ 0.0200	0.1459 $\pm$ 0.0231	0.1499 $\pm$ 0.0207
PCR	0.0470 $\pm$ 0.0109	0.0721 $\pm$ 0.0150	0.0741 $\pm$ 0.0142	0.0720 $\pm$ 0.0124
ICR	0.0477 $\pm$ 0.0105	0.0730 $\pm$ 0.0140	0.0741 $\pm$ 0.0139	0.0720 $\pm$ 0.0132

Table 3

The number of significant edges and the intersection between two methods calculated using the experimental data.

	SCE (%)	PLSR (%)	PCR (%)	ICR (%)	ES (%)
SCE	68 (1.60)	43 (0.98)	38 (0.87)	38 (0.87)	4 (0.09)
PLSR		70 (1.60)	54 (1.24)	54 (1.24)	4 (0.09)
PCR			123 (2.81)	123 (2.81)	4 (0.09)
ICR				123 (2.81)	4 (0.09)
ES					12 (0.47)

The diagonal number in the table stands for the number of significant edges for each method, and the numbers in upper triangle stand for the number of intersection of their estimated networks. The values in parentheses are percentage of significant edges out of all possible connections.