He *et al. BMC Bioinformatics* 2018, **19**(Suppl 17):493
https://doi.org/10.1186/s12859-018-2461-2

BMC Bioinformatics

RESEARCH Open Access

A new insight into underlying disease mechanism through semi-parametric latent differential network model

Yong He[1], Jiadong Ji[1*], Lei Xie[2,3], Xinsheng Zhang[4] and Fuzhong Xue[5]

From The International Conference on Intelligent Biology and Medicine (ICIBM) 2018
Los Angeles, CA, USA. 10-12 June 2018

Abstract

Background: In genomic studies, to investigate how the structure of a genetic network differs between two experiment conditions is a very interesting but challenging problem, especially in high-dimensional setting. Existing literatures mostly focus on differential network modelling for continuous data. However, in real application, we may encounter discrete data or mixed data, which urges us to propose a unified differential network modelling for various data types.

Results: We propose a unified latent Gaussian copula differential network model which provides deeper understanding of the unknown mechanism than that among the observed variables. Adaptive rank-based estimation approaches are proposed with the assumption that the true differential network is sparse. The adaptive estimation approaches do not require precision matrices to be sparse, and thus can allow the individual networks to contain hub nodes. Theoretical analysis shows that the proposed methods achieve the same parametric convergence rate for both the difference of the precision matrices estimation and differential structure recovery, which means that the extra modeling flexibility comes at almost no cost of statistical efficiency. Besides theoretical analysis, thorough numerical simulations are conducted to compare the empirical performance of the proposed methods with some other state-of-the-art methods. The result shows that the proposed methods work quite well for various data types . The proposed method is then applied on gene expression data associated with lung cancer to illustrate its empirical usefulness.

Conclusions: The proposed latent variable differential network models allows for various data-types and thus are more flexible, which also provide deeper understanding of the unknown mechanism than that among the observed variables. Theoretical analysis, numerical simulation and real application all demonstrate the great advantages of the latent differential network modelling and thus are highly recommended.

Keywords: Adaptive estimation, Gaussian copula, Differential graphical model, Latent variable, Rank-based approach

Background

In genomic studies, graphical model has been an important tool to capture dependence among different genes. Particularly, Gaussian graphical model has been widely applied to infer the relationship between genes at the transcriptional level [1–4]. Under the Gaussian assumption, estimating the structure of the graphical model is equivalent to recover the support of precision matrix which is defined to be the inverse of the covariance matrix. However, in some cases, compared to focusing on a particular network, it is of greater interest to investigate how the network of connected gene pairs change from one experimental condition to another, which provides deeper insights on an underlying biological process such as identification of pathways that correspond to such a change. For instance, medical experiment usually involves two groups: the patient group and the control group.

*Correspondence: jiadong@sdufe.edu.cn
[1]School of Statistics, Shandong University of Finance and Economics, 250014 Jinan, China
Full list of author information is available at the end of the article

The analysis of group difference in biological networks or pathways may offer us a new insight into the underlying disease mechanism, which have extensive biomedical and clinical applications, such as identifying effective targets for drug development in a cost-effective and timely manner. Indeed, differential networking modelling has recently emerged as an important tool to analyze a set of changes in graph structure between two conditions (see, for example; [5–17]). In the context of genomic analysis, it is reasonable to assume that two genes are defined to be connected in the differential network if the magnitude of their conditional dependency relationship changes between two conditions. The precision matrix which is defined as the inverse of covariance matrix can capture the conditional dependency relationship. Thus the differential network is typically modelled as the difference of two precision matrices and this type of modelling has been widely used [7–9, 14, 15]. Figure 1a, b, c illustrate the definition of differential network. Each node represents a gene. For two groups depicted in (a) and (b), there is an edge between genes (i, j) if and only if (i, j)-th element of $\mathbf{\Omega}$ is nonzero. For each edge, there exists a weight which is the magnitude of (i, j)-th element of $\mathbf{\Omega}$. Gene pair (i, j) is defined to be connected in the differential network in (c) if the magnitudes of (i, j)-th elements of two precision matrices change between two groups.

One straightforward approach to estimate the difference of two precision matrices is to separately estimate the precision matrices and then subtract the estimates. In the high dimensional setting where the dimension p is much larger than the sample size n, which is often the case for genomic study, many estimation approaches for the precision matrix have been proposed and proved to enjoy nice theoretical properties and computation advantage under the key assumption of sparsity. And this topic has been an active area of research in recent years [18–22].

Another type of approach to estimate the difference of two precision matrices is to jointly estimate the precision matrices. Guo et al. [23] penalized the joint loglikelihood with a hierarchical penalty that targets the removal of common zeros in the inverse covariance matrices across categories. Danaher et al. [24] proposed the joint graphical Lasso, which is based upon maximizing a penalized log-likelihood with generalized fused Lasso or group Lasso penalty. Motivated by the constrained ℓ_1 minimization approach to precision matrix estimation of [22], Zhao et al. [7] proposed an estimation approach to directly estimate the difference of the precision matrices.

For the separately estimating methods, Liu et al. [25] proposed the nonparanormal family to relax the Gaussian assumption. While the nonparanormal family is much larger than the standard parametric Gaussian family, the independence relations among the variables are still encoded in the precision matrix. In addition, Liu et al. [26] proposed a semiparametric approach called nonparanormal SKEPTIC to estimate high dimensional undirected graphical models efficiently and robustly and proved that the nonparanormal SKEPTIC achieves the optimal parametric rates of convergency in terms of precision matrix estimation and graph recovery. Xue and Zou [27] proposed a similar regularized rank-based estimation idea for estimating nonparanormal graphical models and analyzed adaptive versions of rank-based Dantzig selector and CLIME estimators. He et al. [28] proposed a multiple testing procedure to estimate high-dimensional nonparanormal graphical model and proved that the proposed

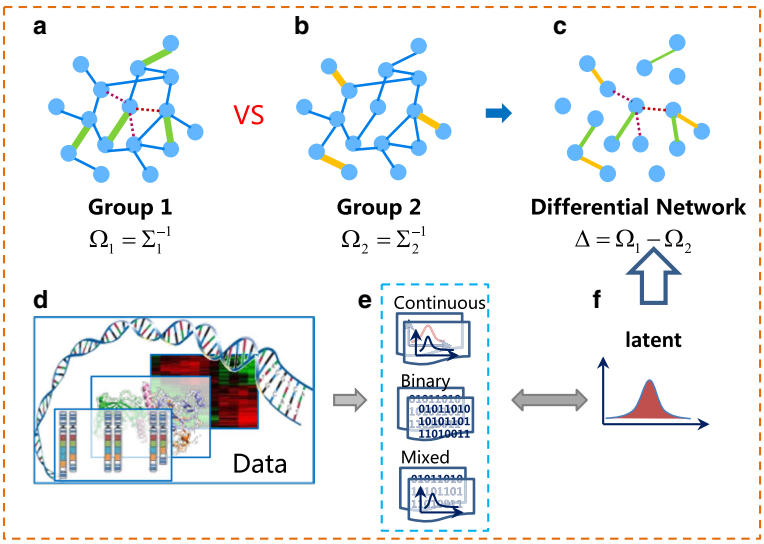

Fig. 1 Illustration of latent differential network. **a** Network in group 1. **b** Network in group 2. **c** Differential network. **d** Data sources. **e** Data type. **f** Latent distribution

procedure can control the false discovery rate (FDR) asymptotically.

The disadvantage of Gaussian or nonparanormal graphical models lies in that they are only tailored for modeling continuous data. However, in genomic studies, we may encounter discrete data (e.g. CNV data and SNP data), continuous data (e.g. gene expression and methylation data) or data of hybrid types with both discrete and continuous variables. Besides, in some circumstances, even if the data are continuous, we still need to transform the data into discrete data to remove the heterogeneity (e.g. batch effect, outliers and population stratification). For instance, in the analysis of gene expression data collected from different platforms, to remove the unwanted variation among different experiments known as the batch effects, numerical expression data are often transformed into 0/1 binary data, where lower expression values are encoded as 0 and higher expression values are encoded as 1. In this setting, it is reasonable to assume that the discrete variable is obtained by discretizing a latent variable. Fan et al. [29] proposed a general model named the latent Gaussian copula graphical model, assuming that the observed discrete data are generated by discretizing a latent continuous variable at some unknown cutoff.

In this paper, we consider estimating differential network for various types of biological data in a joint way. We propose a unified semi-parametric latent variable differential network model. The latent differential network model is illustrated in Fig. 1e-f. For biological data, there exist continuous data, discrete data or data of hybrid types with both continuous and discrete data. It is assumed that these data are collected by transforming latent continuous variables which are unobservable. We are interested in the differential network of the latent variables, which provide deeper understanding of the unknown mechanism than that among the observed variables. To the best of our knowledge, our work provides the first method for differential network estimation for binary or mixed data with theoretical guarantees under the high dimensional scaling. The advantages of the proposed methods lie in the following aspects: (I) Our method provides a way to infer the differential network structure among latent variables, which provides deeper understanding of the unknown mechanism than that among the observed variables. (II) Theoretical analysis shows that the proposed methods achieve the same parametric rates of convergence for both difference matrix estimation and differential graph recovery, as if the latent variables were observed. (III) The proposed methods are much more robust to outliers due to the rank-based correlation matrix estimator. (IV) The proposed approaches do not require precision matrices to be sparse, and thus can allow the individual networks to contain hub nodes. Simulation result shows that the proposed method performs much better and more robustly than several state-of-the-art methods. The proposed methods are applied on a gene expression data set associated with lung cancer. A target gene WIF1 stands out by the proposed method, which indeed is verified as a frequent target for epigenetic silencing in various human cancers [30]. The real data example illustrates the great usefulness of the current work.

Methods

In this part, we propose novel definitions of latent differential network model for various types of data. In essence, we define the differential network as the difference of two precision matrices of the latent variables, which greatly generalizes the applicability in areas such as bioinformatics, medical research and so on.

Gaussian copula differential graphical model

We first review the definition of the Gaussian copula distribution. Let $f = \{f_1, \ldots, f_p\}$ be a set of strictly increasing univariate functions. A p dimensional random variable $\boldsymbol{X} = (X_1, \ldots, X_p)^\top$ is said to follow the Gaussian copula distribution if and only if $f(\boldsymbol{X}) := \left(f_1(X_1), \ldots, f_p\left(X_p\right)\right)^\top := \boldsymbol{Z} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and is noted as $\boldsymbol{X} \sim \text{NPN}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, f\right)$, where $\boldsymbol{\mu} = \left(\mu_1, \ldots, \mu_p\right), \boldsymbol{\Sigma} = \left[\Sigma_{jk}\right]$ are respectively the mean vector and the correlation matrix of the Gaussian variate $\boldsymbol{Z}$. The conditional independence structure of $\boldsymbol{X}$ is encoded by the sparsity pattern of $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$. Specifically, it can be shown that X_i is conditionally independent of X_j given all other variables if and only if $\omega_{ij} = 0$, where ω_{ij} is the (i, j)-th element of $\boldsymbol{\Omega}$. Therefore, the differential network of the Gaussian copula variables can be defined to be the difference between the two precision matrices, just the same as for the parametric Gaussian case.

Assume $\boldsymbol{X}_i = \left(X_{i1}, \ldots, X_{ip}\right)^\top$ for $i = 1, \ldots, n_X$ are independent observations of the expression levels of p genes from one group denoted by $\boldsymbol{X}$ and $\boldsymbol{Y}_i = \left(Y_{i1}, \ldots, Y_{ip}\right)^\top$ for $i = 1, \ldots, n_Y$ from the other denoted by $\boldsymbol{Y}$, $\boldsymbol{X} \sim \text{NPN}\left(\boldsymbol{\mu}^X, \boldsymbol{\Sigma}^X, f^X\right)$ and $\boldsymbol{Y} \sim \text{NPN}\left(\boldsymbol{\mu}^Y, \boldsymbol{\Sigma}^Y, f^Y\right)$. The differential network is defined to be the difference between the two precision matrices, denoted by $\boldsymbol{\Delta}_0 = \boldsymbol{\Omega}^Y - \boldsymbol{\Omega}^X$, where $\boldsymbol{\Omega}^Y$ and $\boldsymbol{\Omega}^X$ are the inverse matrices of $\boldsymbol{\Sigma}^Y$ and $\boldsymbol{\Sigma}^X$ separately.

We propose a rank-based estimator of $\boldsymbol{\Sigma}^X$. It is known that if $\boldsymbol{Z} \sim \text{NPN}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, f\right)$, then we have $\Sigma_{jk} = \sin\left(\frac{\pi}{2}\tau_{jk}\right)$, where τ_{jk} is Kendall's tau correlation between Z_j and Z_k. Thus we can estimate the unknown correlation matrix $\boldsymbol{\Sigma}^X$ by:

$$\hat{S}^X_{jk} = \begin{cases} \sin\left(\frac{\pi}{2}\hat{\tau}^X_{jk}\right) & j \neq k \\ 1 & j = k \end{cases}, \tag{1}$$

where $\hat{\tau}^X_{jk}$ is the sample Kendall's tau correlation between X_j and X_k. Similarly, we can estimate $\boldsymbol{\Sigma}^Y$ in the same way

and obtain the estimator $\hat{\boldsymbol{S}}^Y$. Motivated by the direct estimation method of the difference of two precision matrices proposed by [7], one can obtain the estimator of $\boldsymbol{\Delta}_0$ by solving

$$\arg\min|\boldsymbol{\Delta}|_1, \text{ subject to } |\hat{\boldsymbol{S}}^X\boldsymbol{\Delta}\hat{\boldsymbol{S}}^Y - \hat{\boldsymbol{S}}^X + \hat{\boldsymbol{S}}^Y|_\infty \le \lambda_n,$$

which is equivalent to the optimization problem:

$$\begin{aligned} &\arg\min|\boldsymbol{\Delta}|_1, \quad \text{subject to} \\ &\left|\left(\hat{\boldsymbol{S}}^X \otimes \hat{\boldsymbol{S}}^Y\right)\mathrm{Vec}(\boldsymbol{\Delta}) - \mathrm{Vec}\left(\hat{\boldsymbol{S}}^X - \hat{\boldsymbol{S}}^Y\right)\right|_\infty \le \lambda_n, \end{aligned} \tag{2}$$

where $\otimes$ denotes the Kronecker product, $|\boldsymbol{\Delta}|_1 = \sum_{jk}\delta_{jk}$ is the element-wise ℓ_1 norm of the matrix $\boldsymbol{\Delta}$. Here, for a matrix $\boldsymbol{A} = [A_{jk}]$, $|\boldsymbol{A}|_\infty = \max_{jk}|A_{jk}|$ and for a vector $\boldsymbol{a} = (a_j)$, $|\boldsymbol{a}|_\infty = \max_j|a_j|$.

As seen from Eq. (2), the proposed approach can directly estimate the difference matrix without implicitly estimating the individual precision matrices. Thus there is no need to assume the sparsity of $\left(\boldsymbol{\Sigma}^Y\right)^{-1}$ and $\left(\boldsymbol{\Sigma}^X\right)^{-1}$. We only need to assume that $\boldsymbol{\Delta}_0$ is sparse. Besides, compared to the sample covariance matrix, the rank-based estimators here can enjoy modelling flexibility and estimation robustness, especially when outliers exist.

Latent Gaussian copula differential graphical model for binary data

In the analysis of gene expression data, to remove the batch effects, numerical expression data are often transformed into 0/1 binary data, where lower expression values are encoded as 0 and higher expression values are encoded as 1. To estimate the underlying differential network for the binary data from two different groups, we assume that the observed discrete data are generated by discretizing a latent continuous variable at some unknown cutoff. To make the model more flexible, we assume the latent continuous variable is Gaussian copula distributed instead of Gaussian. Let $\boldsymbol{B} = (B_1, B_2, \ldots, B_p)^\top \in \{0,1\}^p$ be a p-dimensional 0/1-random vector. The 0/1-random vector $\boldsymbol{B}$ satisfies the latent Gaussian copula model (LGCM) for binary data, if there exists a p dimensional random vector $\boldsymbol{X} \sim \mathrm{NPN}(\mathbf{0}, \boldsymbol{\Sigma}, f)$ such that

$$B_j = I\left(X_j > C_j\right), \text{ for all } j = 1, \ldots, p,$$

where $I(\cdot)$ is the indicator function and the cutoff $\boldsymbol{C} = \left(C_1, \ldots, C_p\right)$ is a vector of constants. Then we denote $\boldsymbol{B} \sim \mathrm{LGCM}(\boldsymbol{\Sigma}, f, \boldsymbol{C})$. We call $\boldsymbol{\Sigma}$ the latent correlation matrix. The latent Gaussian copula model involves parameters $(\boldsymbol{\Sigma}, f, \boldsymbol{C})$. Merely based on the binary random vector $\boldsymbol{B}$, only $f_j(C_j), j = 1, \ldots, p$ are identifiable. Denote $\boldsymbol{\Lambda} = \left(\Lambda_1, \ldots, \Lambda_p\right)$, where $\Lambda_j = f_j(C_j)$. For notational simplicity, we write $\mathrm{LGCM}(\boldsymbol{\Sigma}, \boldsymbol{\Lambda})$ for $\mathrm{LGCM}(\boldsymbol{\Sigma}, f, \boldsymbol{C})$.

Assume $\boldsymbol{B}_i^1 = \left(B_{i1}^1, \ldots, B_{ip}^1\right)^\top$ for $i = 1, \ldots, n_1$ are independent observations of the binary expression levels of p genes from one group denoted by $\boldsymbol{B}^1$ and $\boldsymbol{B}_i^2 = \left(B_{i1}^2, \ldots, B_{ip}^2\right)^\top$ for $i = 1, \ldots, n_2$ from the other denoted by $\boldsymbol{B}^2$, where $\boldsymbol{B}^1 \sim \mathrm{LGCM}\left(\boldsymbol{\Sigma}^1, \boldsymbol{\Lambda}^1\right)$ and $\boldsymbol{B}^2 \sim \mathrm{LGCM}\left(\boldsymbol{\Sigma}^2, \boldsymbol{\Lambda}^2\right)$. The differential network is defined to be the difference between the two precision matrices, denoted by $\boldsymbol{\Delta}_0^{\mathrm{B}} = \left(\boldsymbol{\Sigma}^2\right)^{-1} - \left(\boldsymbol{\Sigma}^1\right)^{-1}$. Motivated by Eq. (2), we should first derive estimators for $\boldsymbol{\Sigma}^1$ and $\boldsymbol{\Sigma}^2$. For ease of presentation, we only present the procedure to construct the estimator for $\boldsymbol{\Sigma}^1$, estimator for $\boldsymbol{\Sigma}^2$ can be obtained similarly. Denote the Kendall's tau correlation between B_j^1 and B_k^1 by τ_{jk}^1, it can be shown that τ_{jk}^1 satisfies:

$$\tau_{jk}^1 = 2\left\{\Phi_2\left(\Lambda_j^1, \Lambda_k^1, \Sigma_{jk}^1\right) - \Phi\left(\Lambda_j^1\right)\Phi\left(\Lambda_k^1\right)\right\},$$

where

$$\Phi_2(u, v, t) = \int_{x_1<u}\int_{x_2<v}\phi_2(x_1, x_2; t)dx_1dx_2,$$

is the cumulative distribution function of the standard bivariate normal distribution, $\phi_2(x_1, x_2; t)$ is the probability density function of the standard bivariate normal distribution with correlation t. Denote by

$$F\left(t; \Lambda_j^1, \Lambda_k^1\right) = 2\left\{\Phi_2\left(\Lambda_j^1, \Lambda_k^1, t\right) - \Phi\left(\Lambda_j^1\right)\Phi\left(\Lambda_k^1\right)\right\}.$$

For any fixed Λ_j^1 and Λ_k^1, it can be shown that $F\left(t; \Lambda_j^1, \Lambda_k^1\right)$ is a strictly monotonic increasing function on $t \in (-1, 1)$ and thus is invertible. Given Λ_j^1 and Λ_k^1, one can estimate Σ_{jk}^1 by $F^{-1}\left(\hat{\tau}_{jk}^1; \Lambda_j^1, \Lambda_k^1\right)$. However, the cutoff values are unknown in practice. As $E\left(B_{ij}^1\right) = 1 - \Phi\left(\Lambda_j^1\right)$, we can estimate Λ_j^1 by $\hat{\Lambda}_j^1 = \Phi^{-1}\left(1 - \bar{B}_j^1\right)$, where $\bar{B}_j^1 = \sum_{i=1}^n B_{ij}^1/n$. Thus the Kendall's tau rank-based correlation matrix estimator $\hat{\boldsymbol{b}}^1 = \left[\hat{\mathrm{R}}_{jk}^1\right]$ for $\boldsymbol{\Sigma}^1$ is a $p \times p$ matrix with element entry given by

$$\hat{\mathrm{R}}_{jk}^1 = \begin{cases} F^{-1}\left(\hat{\tau}_{jk}^1; \hat{\Lambda}_j^1, \hat{\Lambda}_k^1\right) & j \ne k, \\ 1, & j = k. \end{cases} \tag{3}$$

Similarly, the Kendall's tau rank-based correlation matrix estimator $\hat{\boldsymbol{b}}^2 = \left[\hat{\mathrm{R}}_{jk}^2\right]$ for $\boldsymbol{\Sigma}^2$ is a $p \times p$ matrix with element entry given by

$$\hat{\mathrm{R}}_{jk}^2 = \begin{cases} F^{-1}\left(\hat{\tau}_{jk}^2; \hat{\Lambda}_j^2, \hat{\Lambda}_k^2\right) & j \ne k, \\ 1, & j = k. \end{cases} \tag{4}$$

Motivated by Eq. (2), we can obtain an estimator of $\mathbf{\Delta}_0^{\mathrm{B}}$ by solving the following optimization problem:

$$\begin{aligned} &\arg\min|\mathbf{\Delta}|_1, \ \text{ subject to} \\ &\left|\left(\hat{\boldsymbol{b}}^1 \otimes \hat{\boldsymbol{b}}^2\right)\mathrm{Vec}(\mathbf{\Delta}) - \mathrm{Vec}\left(\hat{\boldsymbol{b}}^1 - \hat{\boldsymbol{b}}^2\right)\right|_\infty \leq \lambda_n. \end{aligned} \tag{5}$$

For the binary data, we aim to infer the differential network among latent variables, which provides deeper understanding of the unknown mechanism than that among the observed binary variables. Thus, our model complements the existing work on high dimensional differential network estimation, which mostly focused on learning differential network among observed variables including, for example, the Ising model.

Latent Gaussian copula differential graphical model for mixed data

In the analysis of biological data, there also exists the case where some biological data are discrete while some others are continuous. For instance, multi-level omics data integrative analysis involves gene mutation, expression, methylation, metabolome and phenome data. In this case, mixed data appear naturally. We start with the definition of the latent Gaussian copula model for mixed data. Assume that $\boldsymbol{M} = (\boldsymbol{M}_1, \boldsymbol{M}_2)$, where $\boldsymbol{M}_1$ represents the p_1-dimensional binary variables and $\boldsymbol{M}_2$ represents the p_2-dimensional continuous variables. The random vector $\boldsymbol{M}$ satisfies the latent Gaussian copula model for mixed data, if there exists a p_1 dimensional random vector $\boldsymbol{X}_1$ such that $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{M}_2) \sim \mathrm{NPN}(0, \mathbf{\Sigma}, f)$ and

$$M_j = I\left(X_j > C_j\right) \ \text{ for all } j = 1, \ldots, p_1,$$

where $\boldsymbol{C} = \left(C_1, \ldots, C_{p_1}\right)$ is a vector of constants. Then we denote $\boldsymbol{M} \sim \mathrm{LGCM}(0, \mathbf{\Sigma}, f, \boldsymbol{C})$, and call $\mathbf{\Sigma}$ the latent correlation matrix. In the latent Gaussian copula regression model, the binary components $\boldsymbol{M}_1$ are generated by a latent continuous random vector $\boldsymbol{X}_1$ truncated at $\boldsymbol{C}$, and combining with the continuous components $\boldsymbol{M}_2$, $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{M}_2)$ satisfies the Gaussian copula model. For the binary data $\boldsymbol{M}_1$, only $\Lambda_j = f_j\left(C_j\right), j = 1, \ldots, p_1$ are identifiable. For the continuous components $\boldsymbol{M}_2$, the marginal transformations $f_j(\cdot), j = p_1 + 1, \ldots, p$ are identifiable.

Assume $\boldsymbol{M}_i^1 = \left(M_{i1}^1, \ldots, M_{ip}^1\right)^\top$ for $i = 1, \ldots, n_1$ are independent observations of the expression levels of p genes from one group denoted by $\boldsymbol{M}^1$ and $\boldsymbol{M}_i^2 = \left(M_{i1}^2, \ldots, M_{ip}^2\right)^\top$ for $i = 1, \ldots, n_2$ from the other denoted by $\boldsymbol{M}^2$, where $\boldsymbol{M}^1 \sim \mathrm{LGCM}\left(\mathbf{\Sigma}^1, \mathbf{\Lambda}^1\right)$ and $\boldsymbol{M}^2 \sim \mathrm{LGCM}\left(\mathbf{\Sigma}^2, \mathbf{\Lambda}^2\right)$. The differential network is defined to be the difference between the two precision matrices, denoted by $\mathbf{\Delta}_0^{\mathrm{M}} = \left(\mathbf{\Sigma}^2\right)^{-1} - \left(\mathbf{\Sigma}^1\right)^{-1}$. Similar to the discussions in the last sections, we first need to construct estimators for $\mathbf{\Sigma}^1$ and $\mathbf{\Sigma}^2$. For ease of presentation, we only present the procedure to construct the estimator for $\mathbf{\Sigma}^1$, estimator for $\mathbf{\Sigma}^2$ can be obtained similarly. For discrete components $M_{ij}^1, M_{ik}^1 (1 \leq j, k \leq p_1)$, as what we have discussed in the last subsection with a slight change of notation, we can estimate Σ_{jk}^1 by:

$$\hat{\mathrm{T}}_{jk}^1 = \begin{cases} F^{-1}\left(\hat{\tau}_{jk}^1; \hat{\Lambda}_j^1, \hat{\Lambda}_k^1\right) & 1 \leq j \neq k \leq p_1, \\ 1, & 1 \leq j = k \leq p_1. \end{cases} \tag{6}$$

For continuous components M_{ij}^1, M_{ik}^1, as what we have discussed, we can estimate Σ_{jk}^1 by:

$$\hat{\mathrm{T}}_{jk}^1 = \begin{cases} \sin\left(\frac{\pi}{2}\hat{\tau}_{jk}\right) & p_1 + 1 \leq j \neq k \leq p, \\ 1, & p_1 + 1 \leq j = k \leq p. \end{cases} \tag{7}$$

where $\hat{\tau}_{jk}$ is defined as follows:

$$\hat{\tau}_{jk}^1 = \frac{2}{n_1(n_1-1)} \sum_{1 \leq i < i' \leq n_1} \mathrm{sign}\left(M_{ij}^1 - M_{i'j}^1\right) \cdot \mathrm{sign}\left(M_{ik}^1 - M_{i'k}^1\right).$$

We still need to consider the mixed case. Without loss of generality, we assume that M_{ij}^1 is binary and M_{ik}^1 is continuous. In this case, the Kendall's tau correlation can be expressed by

$$\hat{\tau}_{jk}^1 = \frac{2}{n_1(n_1-1)} \sum_{1 \leq i < i' \leq n_1} \left(M_{ij}^1 - M_{i'j}^1\right) \cdot \mathrm{sign}\left(M_{ik}^1 - M_{i'k}^1\right).$$

The population version of Kendall's tau correlation $\tau_{jk}^1 = E\left(\hat{\tau}_{jk}^1\right)$ can be expressed by $\tau_{jk}^1 = H\left(\Sigma_{jk}^1; \Lambda_j^1\right)$, where

$$H\left(t; \Lambda_j^1\right) = 4\Phi_2\left(\Lambda_j^1, 0, t/\sqrt{2}\right) - 2\Phi\left(\Lambda_j^1\right).$$

Moreover, for fixed Λ_j^1, $H\left(t; \Lambda_j^1\right)$ is an invertible function of t. The parameter Λ_j^1 could be estimated by $\Lambda_j^1 = \Phi^{-1}\left(1 - \bar{M}_j^1\right)$, where $\bar{M}_j^1 = \sum_{i=1}^n M_{ij}^1/n$. Thus when M_{ij}^1 is binary and M_{ik}^1 is continuous, Σ_{jk}^1 can be estimated by the Kendall' tau rank-based estimator:

$$\hat{\mathrm{T}}_{jk}^1 = H^{-1}\left(\hat{\tau}_{jk}^1; \hat{\Lambda}_j^1\right), \ \ 1 \leq j \leq p_1 < k \leq p, \tag{8}$$

where $H^{-1}\left(\tau, \Lambda_j^1\right)$ is the inverse function of $H\left(t, \Lambda_j^1\right)$ for fixed Λ_j^1. Thus the Kendall's tau rank-based correlation matrix estimator $\hat{\mathbf{T}}^1 = \left[\hat{\mathrm{T}}_{jk}^1\right]$ for $\mathbf{\Sigma}^1$ is a $p \times p$ matrix with corresponding element entry given by Eqs. (6), (7), and (8) respectively. Similarly, we can obtain estimator $\hat{\mathbf{T}}^2$ for $\mathbf{\Sigma}^2$. Motivated by Eq. (2), we can obtain an estimator of $\mathbf{\Delta}_0$ by solving the following optimization problem:

$$\begin{aligned} &\arg\min|\mathbf{\Delta}|_1, \ \text{ subject to} \\ &\left|\left(\hat{\mathbf{T}}^1 \otimes \hat{\mathbf{T}}^2\right)\mathrm{Vec}(\mathbf{\Delta}) - \mathrm{Vec}\left(\hat{\mathbf{T}}^1 - \hat{\mathbf{T}}^2\right)\right|_\infty \leq \lambda_n. \end{aligned} \tag{9}$$

We show that the rank-based covariance matrix estimators achieve the same parametric rate of convergence for

both difference matrix estimation and differential graph recovery in the Additional file 1. Thus the extra modelling flexibility comes at almost no cost of statistical efficiency. Besides, for the binary data or data of hybrid types with both binary and continuous variables, the differential network among latent variables can be well estimated, which provides deeper understanding of the unknown mechanism than that among the observed variables.

Implementation

In this section we will present how to solve the optimization problems in Eqs. (2), (5), and (9). For ease of presentation, we only present the procedure to obtain the solution to optimization problem in Eq. (2) and solutions to optimization problems in Eqs. (5) and (9) can be obtained in the similar way.

Recall that in Eq. (2), the optimization problem is

$$\arg\min|\boldsymbol{\Delta}|_1, \quad \text{subject to}$$

$$\left|\left(\hat{\boldsymbol{S}}^X \otimes \hat{\boldsymbol{S}}^Y\right)\mathrm{Vec}(\boldsymbol{\Delta}) - \mathrm{Vec}\left(\hat{\boldsymbol{S}}^X - \hat{\boldsymbol{S}}^Y\right)\right|_\infty \le \lambda_n.$$

Let $\boldsymbol{\Delta} = [\delta_{jk}]_{1\le j,k\le p}$ and define $\boldsymbol{\theta}$ to be the $p(p+1)/2 \times 1$ vector with $\boldsymbol{\theta} = (\delta_{jk})_{1\le j\le k\le p}$. Estimating a symmetric $\boldsymbol{\Delta}$ is thus equivalent to estimating $\boldsymbol{\theta}$, which alleviates the computation burden especially when p is large. Define the $p^2 \times p(p+1)/2$ matrix $\boldsymbol{\Gamma}$ with columns indexed by $1 \le j \le k \le p$ and with rows indexed by $l = 1,\ldots,p$ and $m = 1,\ldots,p$, so that each entry is labeled by $\Gamma_{lm,jk}$. For $j \le k$, let $\Gamma_{jk,jk} = \Gamma_{kj,jk} = 1$ and set all other entries of $\boldsymbol{\Gamma}$ equal to zero. With these notations, one may consider the following optimization problem:

$$\hat{\boldsymbol{\theta}} = \arg\min|\boldsymbol{\theta}|_1 \ \text{ subject to}$$

$$\begin{cases} \left|\boldsymbol{\Gamma}^\top\hat{\boldsymbol{S}}\boldsymbol{\Gamma}\boldsymbol{\theta} - \boldsymbol{\Gamma}^\top\hat{\boldsymbol{s}}\right|_{O\infty} \le \lambda_n, \\ \left|\boldsymbol{\Gamma}^\top\hat{\boldsymbol{S}}\boldsymbol{\Gamma}\boldsymbol{\theta} - \boldsymbol{\Gamma}^\top\hat{\boldsymbol{s}}\right|_{D\infty} \le \lambda_n/2, \end{cases} \tag{10}$$

where $\hat{\boldsymbol{S}} = \hat{\boldsymbol{S}}^X \otimes \hat{\boldsymbol{S}}^Y$, $\hat{\boldsymbol{s}} = \mathrm{Vec}\left(\hat{\boldsymbol{S}}^X - \hat{\boldsymbol{S}}^Y\right)$ and for a $p(p+1)/2 \times 1$ vector $\boldsymbol{c}$, $|\boldsymbol{c}|_{O\infty}$ denotes the sup-norm of the entries of $\boldsymbol{c}$ corresponding to the off diagonal elements of its matrix form, while $|\boldsymbol{c}|_{D\infty}$ denotes the sup-norm of the entries of $\boldsymbol{c}$ corresponding to the diagonal elements. The matrix form of $\hat{\boldsymbol{\theta}}$ will be denoted by $\hat{\boldsymbol{\Delta}}$ in the following sections. The optimization problem in Eq. (10) can be solved by the alternating direction method of multipliers (ADMM), for a thorough discussion, we refer to [31]. For the optimization problem in Eq. (10), to apply the ADMM algorithm, we rewrite it as:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta},\boldsymbol{z}}\left\{|\boldsymbol{\theta}|_1 + g(\boldsymbol{z})\right\}$$

$$\text{subject to} \ \ \boldsymbol{\Gamma}^\top\hat{\boldsymbol{S}}\boldsymbol{\Gamma}\boldsymbol{\theta} + \boldsymbol{z} = \boldsymbol{\Gamma}^\top\hat{\boldsymbol{s}},$$

where the function $g(\cdot)$ is defined by

$$g(\boldsymbol{z}) = \begin{cases} \infty & |\boldsymbol{z}_{O\infty}| > \lambda_n \ \text{ or } \ |\boldsymbol{z}_{D\infty}| > \lambda_n/2. \\ 0, & \text{otherwise.} \end{cases}$$

The augmented Lagrangian can be written as

$$\begin{aligned} L_\rho(\boldsymbol{\theta},\boldsymbol{z},\boldsymbol{u}) &= \boldsymbol{u}^\top\left(\boldsymbol{\Gamma}^\top\hat{\boldsymbol{S}}\boldsymbol{\Gamma}\boldsymbol{\theta} + \boldsymbol{z} - \boldsymbol{\Gamma}^\top\hat{\boldsymbol{s}}\right) + |\boldsymbol{\theta}|_1 \\ &\quad + \tfrac{\rho}{2}\left|\boldsymbol{\Gamma}^\top\hat{\boldsymbol{S}}\boldsymbol{\Gamma}\boldsymbol{\theta} + \boldsymbol{z} - \boldsymbol{\Gamma}^\top\hat{\boldsymbol{s}}\right|_2^2 + g(\boldsymbol{z}), \end{aligned} \tag{11}$$

where $\boldsymbol{u}$ is the Lagrange multiplier and ρ is a positive penalty parameter which can be specified by users. The ADMM algorithm is based on minimizing the augmented Lagrangian in (11) over $\boldsymbol{\theta}$ and $\boldsymbol{z}$ and then applying a dual variable update to the Lagrange multiplier $\boldsymbol{u}$, which yields the updates

$$\begin{aligned} \boldsymbol{z}^{(t+1)} &= \arg\min_{\boldsymbol{z}}\left|\boldsymbol{u}^{(t)}/\rho + \boldsymbol{\Gamma}^\top\hat{\boldsymbol{s}} - \boldsymbol{\Gamma}^\top\hat{\boldsymbol{S}}\boldsymbol{\Gamma}\boldsymbol{\theta}^{(t)} - \boldsymbol{z}\right|_2^2 \\ &\quad + 2g(\boldsymbol{z})/\rho \\ \boldsymbol{\theta}^{(t+1)} &= \arg\min_{\boldsymbol{\theta}}\left|\tfrac{\boldsymbol{u}^{(t)}}{\rho} + \boldsymbol{\Gamma}^\top\hat{\boldsymbol{s}} - \boldsymbol{\Gamma}^\top\hat{\boldsymbol{S}}\boldsymbol{\Gamma}\boldsymbol{\theta} - \boldsymbol{z}^{(t+1)}\right|_2^2 \\ &\quad + 2|\boldsymbol{\theta}|_1/\rho \\ \boldsymbol{u}^{(t+1)} &= \boldsymbol{u}^{(t)} + \rho\left(\boldsymbol{\Gamma}^\top\hat{\boldsymbol{s}} - \boldsymbol{\Gamma}^\top\hat{\boldsymbol{S}}\boldsymbol{\Gamma}\boldsymbol{\theta}^{(t+1)} - \boldsymbol{z}^{(t+1)}\right) \end{aligned}$$

for iterations $t = 0, 1, 2\ldots$. As for the tuning parameter λ_n in (10), it can be chosen by an approximate Akaike information criterion (AIC). λ_n is chosen to minimize

$$(n_X + n_Y)L(\lambda_n) + 2k,$$

where k is the effective degrees of freedom that can be approximated by $|\hat{\boldsymbol{\theta}}|_0$ and $L(\lambda_n)$ represents the loss function either L_∞ or L_F which are defined by

$$L_\infty(\lambda_n) = \left|\hat{\boldsymbol{S}}^X\hat{\boldsymbol{\Delta}}(\lambda_n)\hat{\boldsymbol{S}}^Y - \hat{\boldsymbol{S}}^X + \hat{\boldsymbol{S}}^Y\right|_\infty,$$

$$L_\mathrm{F}(\lambda_n) = \left\|\hat{\boldsymbol{S}}^X\hat{\boldsymbol{\Delta}}(\lambda_n)\hat{\boldsymbol{S}}^Y - \hat{\boldsymbol{S}}^X + \hat{\boldsymbol{S}}^Y\right\|_\mathrm{F}.$$

In this paper we focus on the loss functions with the supremum and Frobenius norms for further theoretical development. One may also use other matrix norms, such as spectral norm:

$$L_\mathrm{sp}(\lambda_n) = \left\|\hat{\boldsymbol{S}}^X\hat{\boldsymbol{\Delta}}(\lambda_n)\hat{\boldsymbol{S}}^Y - \hat{\boldsymbol{S}}^X + \hat{\boldsymbol{S}}^Y\right\|_2.$$

Similarly, for the latent Gaussian copula model for binary data, one can solve the following optimization problem:

$$\hat{\boldsymbol{\theta}}^\mathrm{B} = \arg\min|\boldsymbol{\theta}|_1 \ \text{ subject to}$$

$$\begin{cases} \left|\boldsymbol{\Gamma}^\top\hat{\boldsymbol{b}}\boldsymbol{\Gamma}\boldsymbol{\theta} - \boldsymbol{\Gamma}^\top\hat{\boldsymbol{r}}\right|_{O\infty} \le \lambda_n, \\ \left|\boldsymbol{\Gamma}^\top\hat{\boldsymbol{b}}\boldsymbol{\Gamma}\boldsymbol{\theta} - \boldsymbol{\Gamma}^\top\hat{\boldsymbol{r}}\right|_{D\infty} \le \lambda_n/2, \end{cases} \tag{12}$$

where $\hat{\boldsymbol{b}} = \hat{\boldsymbol{b}}^1 \otimes \hat{\boldsymbol{b}}^2$, $\hat{\boldsymbol{r}} = \text{Vec}\left(\hat{\boldsymbol{b}}^1 - \hat{\boldsymbol{b}}^2\right)$. The matrix form of $\hat{\boldsymbol{\theta}}^{\text{B}}$ will be denoted by $\hat{\boldsymbol{\Delta}}^{\text{B}}$ in the following sections. For the latent Gaussian copula model for mixed data, one can solve the following optimization problem:

$$\hat{\boldsymbol{\theta}}^{\text{M}} = \arg\min |\boldsymbol{\theta}|_1 \text{ subject to}$$

$$\begin{cases} \left|\boldsymbol{\Gamma}^\top \hat{\boldsymbol{\Gamma}} \hat{\boldsymbol{\Gamma}} \boldsymbol{\theta} - \boldsymbol{\Gamma}^\top \hat{\boldsymbol{t}}\right|_{O\infty} \le \lambda_n, \\ \left|\boldsymbol{\Gamma}^\top \hat{\boldsymbol{\Gamma}} \hat{\boldsymbol{\Gamma}} \boldsymbol{\theta} - \boldsymbol{\Gamma}^\top \hat{\boldsymbol{t}}\right|_{D\infty} \le \lambda_n/2, \end{cases} \tag{13}$$

where $\hat{\boldsymbol{\Gamma}} = \hat{\boldsymbol{\Gamma}}^1 \otimes \hat{\boldsymbol{\Gamma}}^2$, $\hat{\boldsymbol{t}} = \text{Vec}\left(\hat{\boldsymbol{\Gamma}}^1 - \hat{\boldsymbol{\Gamma}}^2\right)$. The matrix form of $\hat{\boldsymbol{\theta}}^{\text{M}}$ will be denoted by $\hat{\boldsymbol{\Delta}}^{\text{M}}$ in the following sections. Besides, corresponding Akaike information criterion can be proposed to choose the tuning parameter λ_n.

Simulation

Simulation for Gaussian copula differential graphical model In this part, we conduct simulation study for differential network estimation under Gaussian copula model. We mainly focus on the graphs that contain hub nodes. First we generate the edge set E^X for the group $\boldsymbol{X}$. We partition p features into 5 equally-sized and non-overlapping sets: $C_1 \cup C_2 \cdots \cup C_5 = \{1, \ldots, p\}$, $|C_k| = p/5$, $C_i \cap C_j = \emptyset$. For the smallest $i \in C_k$, we set $(i,j) \in C_k$ for all $\{j \neq i : j \in C_k\}$. The non-zero entries of $\boldsymbol{\Omega}^X$ is then determined by the edge set E^X, where $\boldsymbol{\Omega}^X = \left(\boldsymbol{\Sigma}^X\right)^{-1}$. Next, the value of each nonzero entry of $\boldsymbol{\Omega}^X$ was generated from a uniform distribution with support $[-0.75, -0.25] \cup [0.25, 0.75]$. To ensure positive definiteness of $\boldsymbol{\Omega}^X$, let $\boldsymbol{\Omega}^X = \boldsymbol{\Omega}^X + (0.2 + |\lambda_{\min}\left(\boldsymbol{\Omega}^X\right)|)\mathbf{I}$. At last the $\boldsymbol{\Omega}^X$ is rescaled such that $\boldsymbol{\Sigma}^X$ is a correlation matrix. Then we proceed to generate the differential network $\boldsymbol{\Delta}_0$. We randomly select two hub nodes from the 5 equally-sized and non-overlapping sets. The differential network $\boldsymbol{\Delta}_0$ is generated such that the connections of these two hub nodes change sign between $\boldsymbol{\Omega}^X$ and $\boldsymbol{\Omega}^Y$. The correlation matrix $\boldsymbol{\Sigma}^X$ and $\boldsymbol{\Sigma}^Y$ are generated by $\left(\boldsymbol{\Omega}^X\right)^{-1}$ and $\left(\boldsymbol{\Omega}^Y\right)^{-1}$ respectively. Finally we generate n_X i.i.d observations of $\boldsymbol{Z}^X$ from the $N\left(\mathbf{0}, \boldsymbol{\Sigma}^X\right)$ distribution and n_Y i.i.d observations of $\boldsymbol{Z}^Y$ from the $N\left(\mathbf{0}, \boldsymbol{\Sigma}^Y\right)$ distribution. Next we sample n_X i.i.d samples from the nonparanormal distribution $\text{NPN}\left(\mathbf{0}, \boldsymbol{\Sigma}^X, f^X\right)$ and n_Y i.i.d samples from the nonparanormal distribution $\text{NPN}\left(\mathbf{0}, \boldsymbol{\Sigma}^Y, f^Y\right)$. For simplicity, we use the same univariate transformations on each dimension: $f_1^X = f_2^X = \cdots = f_p^X = f$ and $f^X = f^Y$. To sample data from the nonparanormal distribution, we also need $g := f^{-1}$. We consider the Gaussian CDF Transformation of g which is used in [26].

In the simulation study, we let p = 50,80,100,120 and $n_X = n_Y = 100$. The simulation result is based on 100 replications. For each simulated data set, we apply three estimation methods. That is , the direct differential network estimator (DDN) in [7], the rank-based differential network estimator (RDN) and the direct differential network estimator based on the latent variable $\boldsymbol{Z}$ and Pearson correlation (ZP-DDN). In ZP-DDN, we assume that $\boldsymbol{Z}^X$ and $\boldsymbol{Z}^Y$ are observed and the Pearson correlation estimator of $\text{cov}(\boldsymbol{Z}^X)$ and $\text{cov}(\boldsymbol{Z}^Y)$ are plugged into the direct estimation procedure. While ZP-DDN are often not available in real applications, we use ZP-DDN as benchmarks for quantifying the information loss of the remaining estimators.

We evaluate the performance of the estimation methods from two aspects: support recovery and estimation error. The support recovery results are evaluated by true positive rate (TPR) and true negative rate (TNR) along a range of tuning parameter λ. Suppose the true difference matrix $\boldsymbol{\Delta}_0$ has the support $\mathcal{S}_0 = \{(j,k) : \delta_{jk}^0 \neq 0, \text{ and } j \neq k\}$ and its estimator $\hat{\boldsymbol{\Delta}}$ has the support set $\hat{\mathcal{S}}$. TPR and TNR are defined as follows:

$$\text{TPR} = \frac{\text{TP}}{|\mathcal{S}_0|}, \quad \text{TNR} = \frac{\text{TN}}{p(p-1) - |\mathcal{S}_0|},$$

where TP and TN are the numbers of true positives and true negatives respectively, which are defined as

$$\text{TP} = \#\left\{(i,j) : (i,j) \in \mathcal{S}_0 \cap \hat{\mathcal{S}}\right\},$$
$$\text{TN} = \#\left\{(i,j) : (i,j) \in \mathcal{S}_0^c \cap \hat{\mathcal{S}}^c\right\}.$$

To evaluate the support recovery performance, we use the true discovery rate, which is defined as $\text{TD} = \text{TP}/|\hat{\mathcal{S}}_0|$. As for the estimation error, we calculate the element-wise L_∞ norm and Frobenius norm of $\hat{\boldsymbol{\Delta}} - \boldsymbol{\Delta}_0$.

Simulation for latent Gaussian copula differential graphical model In this part, we conduct simulation study for differential network estimation under Latent Gaussian copula model. We assume that the cutoff vector $\boldsymbol{C} \sim \text{Unif}\,[0,1]$ and let $\boldsymbol{\Sigma}^1$ and $\boldsymbol{\Sigma}^2$ be generated in the same way as $\boldsymbol{\Sigma}^X$ and $\boldsymbol{\Sigma}^Y$ described in the last subsection. We consider the following three Scenarios:

- **Scenario 1** Generate data $\boldsymbol{B}^1 = \left(B_1^1, \ldots, B_p^1\right)^\top$, where $B_j^1 = I(X_j > C_j), j = 1, \ldots, p$ and $\boldsymbol{X} \sim \text{NPN}\left(\mathbf{0}, \boldsymbol{\Sigma}^1, f^1\right)$; Generate data $\boldsymbol{B}^2 = \left(B_1^2, \ldots, B_p^2\right)^\top$, where $B_j^2 = I(Y_j > C_j), j = 1, \ldots, p$ and $\boldsymbol{Y} \sim \text{NPN}\left(\mathbf{0}, \boldsymbol{\Sigma}^2, f^2\right)$. The transformation functions f^1 and f^2 are Gaussian CDF transformation.
- **Scenario 2** Generate data $\boldsymbol{M}^1 = \left(M_1^1, \ldots, M_p^1\right)^\top$, where $M_j^1 = I\left(X_j > C_j\right), j = (p/2+1), \ldots, p$, $\boldsymbol{X} \sim \text{NPN}\left(0, \boldsymbol{\Sigma}^1, f^1\right)$ and $M_j^1 = X_j, j = 1, \ldots, p/2$;

Generate data $\boldsymbol{M}^2 = \left(M_1^2, \ldots, M_p^2\right)^\top$, where $M_j^2 = I\left(Y_j > C_j\right), j = p/2+1, \ldots, p$, and $\boldsymbol{Y} \sim \text{NPN}\left(0, \boldsymbol{\Sigma}^2, f^2\right)$ and $M_j^2 = Y_j, j = 1, \ldots, p/2$. The transformation functions f^1 and f^2 are Gaussian CDF transformation.

• **Scenario 3** Generate data $\boldsymbol{B}^1 = \left(B_1^1, \ldots, B_p^1\right)^\top$, where $B_j^1 = I\left(Z_j^1 > C_j\right), j = 1, \ldots, p$ and $\boldsymbol{Z}^1 \sim N\left(\boldsymbol{0}, \boldsymbol{\Sigma}^1\right)$, where 10 entries in each $\boldsymbol{Z}^1$ is randomly sampled and replaced by -5 or 5;

Generate data $\boldsymbol{B}^2 = \left(B_1^2, \ldots, B_p^2\right)^\top$, where $B_j^2 = I\left(Z_j^2 > C_j\right), j = 1, \ldots, p$ and $\boldsymbol{Z}^2 \sim N\left(\boldsymbol{0}, \boldsymbol{\Sigma}^2\right)$, where 10 entries in each $\boldsymbol{Z}^2$ is randomly sampled and replaced by -5 or 5.

In Scenario 1 and Scenario 3, we generate binary data. Scenario 1 corresponds to the latent Gaussian copula model and Scenario 3 corresponds to the setting where the binary data can be misclassified due to the outliers of the latent Gaussian variable. Scenario 3 is designed to investigate the robustness of the proposed approach. Scenario 2 corresponds to the mixed data generated from the latent Gaussian copula model.

Application to gene expression data sets related to lung cancer

In this section we consider the differential network estimation for a gene expression data set related to lung cancer. The data set is publicly available from the Gene Expression Omnibus at accession number GDS2771 and was studied in [24]. It includes 22,283 microarray-derived gene expression measurements from large airway epithelial cells sampled from 97 patients with lung cancer and 90 controls in the data set. It is of interest to investigate how the structure of the gene co-expression network differs between the group of patients with lung cancer and the control group. It may shed light on underlying lung cancer mechanisms. In this real example study, we limited our analysis to the 122 genes in the Wnt signaling pathway. The Wnt signaling pathway has recently emerged as a critical pathway in lung carcinogenesis as already demonstrated in many cancers and particularly in colorectal cancer [32]. The Gene expression levels were analyzed on a logarithmic scale. Each gene feature was standardized to have mean zero and standard deviation 1 within the cancer samples and the controls separately.

Results

Simulation results for Gaussian copula differential graphical model

The receiver operating characteristic (ROC) curves of the three estimation methods are depicted in Fig. 2. It shows that the proposed method RDN compares favourably with the benchmark method ZP-DDN, which means that the information loss is negligible. Besides, Fig. 2 also shows that DDN performs pretty bad in the non-Gaussian case.

Table 1 gives the true discovery rates with different loss functions. The results also show the method RDN compares favourably with the benchmark method ZP-DDN. For all the methods, tuning using the L_{F} gives better true discovery rates than tuning using the L_∞. Table 1 depicts the elementwise L_∞ norm estimation accuracies of the thresholded estimators tuned using the loss functions L_∞ and L_{F} . From Table 1, we can see that the L_{F} loss function gives slightly better results than the L_∞ loss function. For all the methods, the elementwise L_∞ norm estimation accuracy are comparable. We point out that it is possible for RDN to simultaneously give better support recovery but similar estimation than DDN. The reason is that estimation error depends on the magnitudes of the estimated entries, while support recovery depends only on whether the entries are nonzero. Besides, RDN has comparable performance with the benchmark method ZP-DDN in terms of both support recovery and estimation accuracy, which indicates that the information loss of the estimator RDN is negligible.

Simulation results for Latent Gaussian copula differential graphical model

The ROC curves for Scenario 1 and Scenario 2 with different dimensionality p (varying from 50 to 120) is presented in Fig. 3. Table 2 give the true discovery rates with different loss functions and the elementwise L_∞ norm estimation accuracies of the thresholded estimators tuned using the loss functions L_∞ and L_{F}, respectively. For method ZR-RDN, we assume that the latent Gaussian copula variables are observed. In particular, the rank-based correlation matrix estimator of the latent Gaussian copula variables are plugged into the direct estimation procedure. With a slight abuse of notation, the RDN method here refers to either the rank-based method for binary data or for mixed data. The ROC curves in Fig. 3 show that the rank-based methods RDN proposed for latent Gaussian copula model (binary and mixed) perform pretty well even when the dimensionality is larger than the sample size.

By the ROC curves in Fig. 4, we can find that RDN is more robust to the data misclassification than the benchmark estimator ZP-DDN. The robustness of RDN to outliers illustrates the advantage of the dichotomization method. In the absence of misclassification, it is seen that the ROC curves of RDN and ZR-RDN are similar, which indicates little information loss for differential network recovery due to the dichotomization procedure. Table 3 gives the true discovery rates with different different loss functions for Scenario 3 and presents the elementwise L_∞ norm estimation accuracies of the thresholded estimators tuned using the loss functions L_∞ and L_{F} for Scenario 3. From Table 3, we can see that the L_{F} loss function gives

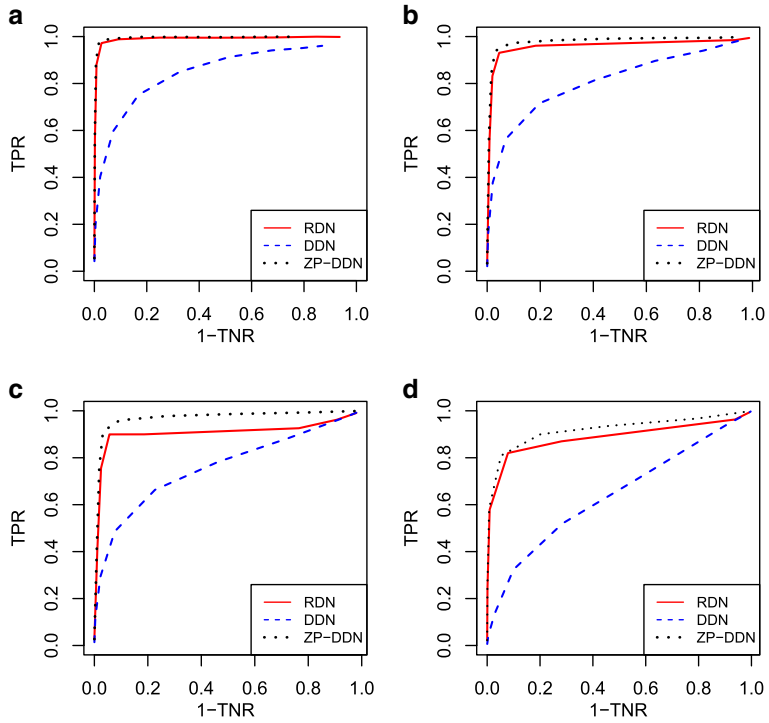

Fig. 2 Receiver operating characteristic curves under Gaussian copula model with dimensionality varying from 50 to 120. The red line represents the proposed RDN method, the black dotted represents the benchmark method ZP-DDN, the blue dotted line represents DDN method. **a** Scenario 3, $p = 50$. **b** Scenario 3, $p = 80$. **c** Scenario 3, $p = 100$. **d** Scenario 3, $p = 120$

slightly better results than the L_∞ loss function. Besides, we can see that the elementwise L_∞ norm estimation accuracy are comparable. This is also true for Scenario 1 and Scenario 2.

Theoretical results

The estimators $\hat{\mathbf{\Delta}}$, $\hat{\mathbf{\Delta}}^{\mathrm{B}}$ and $\hat{\mathbf{\Delta}}^{\mathrm{M}}$, after an additional threshold step, are shown to be able to recover not only the support of the true $\mathbf{\Delta}_0$ but also the signs of its nonzero entries as long as those entries are sufficiently large. Besides, under mild conditions, the estimation errors bounds in terms of matrix Frobenius norm and elementwise ℓ_∞ norm both achieve the parametric rate $\sqrt{\log p/\min(n_X, n_Y)}$, see details in Additional file 1. It indicates that the extra modeling flexibility and robustness come at almost no cost of statistical efficiency and it seems

Table 1 Average true discovery rates (%) and average estimation errors over 100 simulations

	ZP-DDN		RDN		DDN	
p	L_∞	L_F	L_∞	L_F	L_∞	L_F
Average true discovery rates						
50	74.0 (13.6)	83.2 (10.9)	75.6 (14.0)	89.1 (11.3)	45.9 (24.7)	27.8 (17.3)
80	91.4 (16.4)	99.6 (4.3)	95.2 (14.2)	100.0 (0.0)	44.9 (34.6)	51.0 (42.8)
100	96.3 (14.1)	100.0(0.0)	99.5 (5.2)	100.0 (0.0)	39.3 (40.3)	50.0 (49.1)
120	78.8 (16.8)	100.0(0.0)	79.0 (18.2)	100.0 (0.0)	23.4 (41.3)	30.0 (46.3)
Average estimation errors in the elementwise L_∞ norm						
50	3.26 (0.41)	2.91 (0.33)	3.08 (0.32)	2.59 (0.35)	2.27 (0.12)	2.41 (0.21)
80	2.06 (0.28)	1.92 (0.06)	1.98 (0.21)	1.91 (0.00)	1.97 (0.09)	1.94 (0.08)
100	1.86 (0.15)	1.82 (0.00)	1.82 (0.04)	1.82 (0.00)	1.87 (0.10)	1.83 (0.04)
120	1.12 (0.17)	0.87 (0.00)	1.12 (0.18)	0.87 (0.00)	0.89 (0.07)	0.87 (0.00)

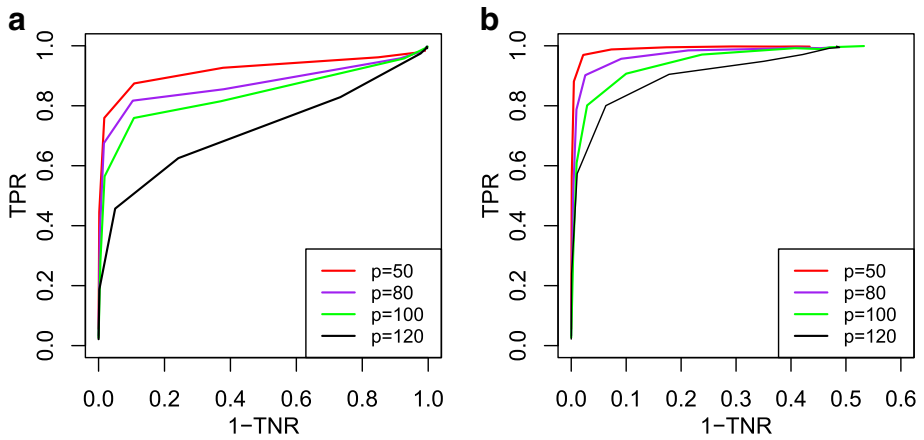

Fig. 3 Receiver operating characteristic curves for Scenario 1 and Scenario 2 under latent Gaussian copula model, with dimensionality varying from 50 to 120. **a** Scenario 1. **b** Scenario 2

as if the latent variable can be observed. Thus these new estimators can be used as a safe replacement of Gaussian estimators even when the data are truly Gaussian. Compared to the separate and joint approaches to estimating differential networks (e.g. [22, 23],) which require sparsity on each $\mathbf{\Sigma}^{-1}$, the proposed direction estimation methods for different types of data only require the sparsity of the difference matrix $\mathbf{\Delta}_0$. The detailed theorems and proofs are in the Additional file 1 available online.

Results of application

In the real application part, we compare three estimation methods. The first method is the Gaussian copula RDN method, which we denote as C-RDN. The second method is the latent Gaussian copula RDN method, which we denote as B-RDN. In specific, we first apply the adaptive dichotomization method implemented by the ArrayBin package in R to remove the batch effect in the gene expression data. The adaptive dichotomization method transforms the numerical gene expression data into 0/1 binary data. The genes with high expression level are encoded as 1 and the genes with lower expression level are encoded as 0. Then we apply the B-RDN to the 0/1 binary data. The third method is the direct differential network estimation method proposed by [7] with Gaussian assumption, which we denote as DDN.

We conduct Shapiro-Wilk test on the gene data set and 63% of the genes reject the normality null hypothesis. Therefore, the Gaussian assumption of DDN method is violated in this real data example. Thus we expect that C-RDN which relaxes the Gaussian assumption may provide a more reliable result. The deficiency of the C-RDN method lies in that it does not take the batch effect of the genes expression data from different platforms into consideration. For the B-RDN method, it removes the batch effect.

Figure 5 depicts the differential network estimated by the three methods. Table 4 gives the hub genes selected out by different estimation methods. For method C-RDN, the tuning parameter λ is selected by the AIC criterion with the elementwise ℓ_1 norm loss function. To ensure a fair comparison, the tuning parameter λ for method B-RDN and DDN are selected such that the number of edges in the estimated differential graphs by all three methods are almost the same. The number of edges selected by the three methods are 56, 59 and 52, respectively. From Fig. 5, we can see that B-RDN identifies an obvious hub gene WIF1 that is an extracellular antagonist of WNT. WIF1 is a frequent target for epigenetic silencing in various human cancers [30]. WIF1 promoter is frequently methylated in non-small cell lung cancer (NSCLC) cells to down-regulate its mRNA expression [33]. Both C-RDN and B-RDN select out a common hub gene APC. APC expression in lung cancer are associated

Table 2 Simulation results over 100 replications for Scenario 1 and Scenario 2

	Scenario 1		Scenario 2	
p	L_∞	L_F	L_∞	L_F
Average true discovery rates(%)				
50	78.8 (15.2)	98.4 (5.9)	79.6 (13.8)	40.8 (25.6)
80	76.4 (23.1)	100.0(0.0)	83.4 (17.0)	88.2 (17.6)
100	89.5 (22.1)	100.0(0.0)	84.8 (20.0)	99.3 (3.9)
120	76.5 (31.0)	94.0(24.0)	82.4 (15.2)	100.0 (0.0)
Average estimation errors in the elementwise L_∞ norm				
50	2.66 (0.26)	2.21 (0.15)	3.23 (0.40)	3.85 (0.55)
80	2.10 (0.20)	1.91 (0.00)	2.29 (0.35)	2.14 (0.32)
100	1.88 (0.13)	1.82 (0.00)	2.03 (0.28)	1.83 (0.08)
120	1.00 (0.16)	0.87 (0.00)	1.17 (0.16)	0.88 (0.07)

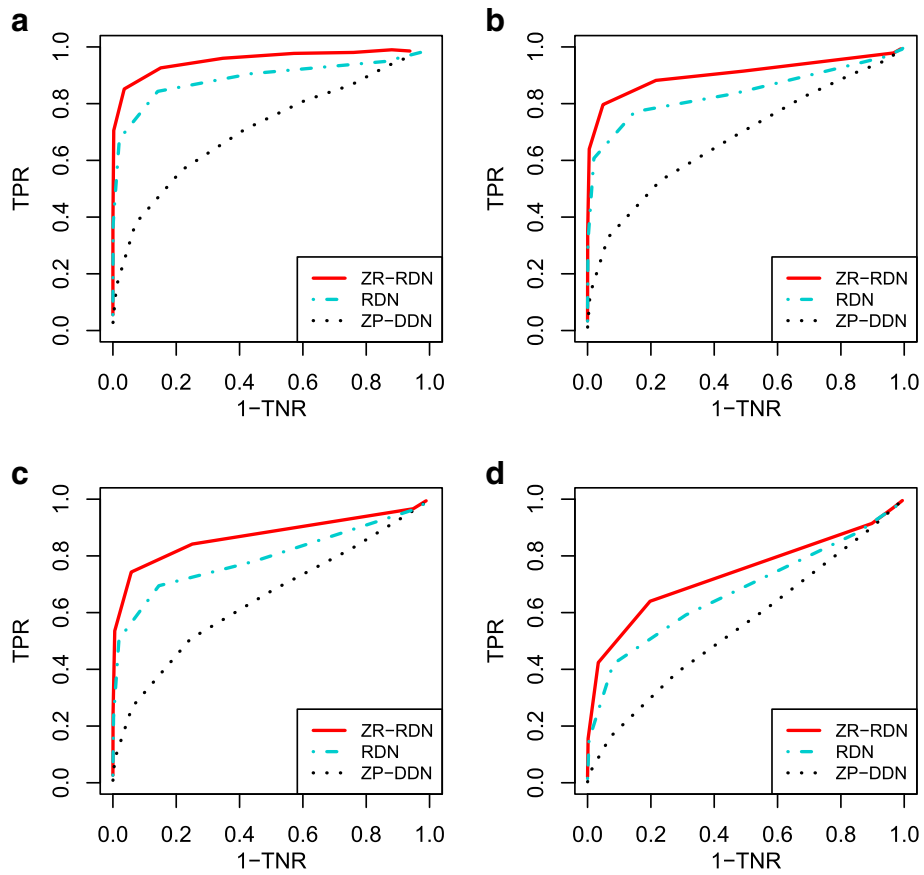

Fig. 4 Receiver operating characteristic curves for Scenario 3 under latent Gaussian copula model, with dimensionality varying from 50 to 120. The red line represents the proposed RDN method, the black dotted represents the benchmark method ZP-DDN, the blue dotted line represents DDN method. **a** Scenario 3, $p = 50$. **b** Scenario 3, $p = 80$. **c** Scenario 3, $p = 100$. **d** Scenario 3, $p = 120$

with survival time and is also related to cancer metastasis [34]. Both C-RDN and DDN select out a common hub gene, MAPK8, which plays a significant role in the promotion of lung inflammation and tumorigenesis subsequent to tobacco smoke exposure [35]. The expression level of DVL2 was reported significantly higher in lung adenocarcinomas than in squamous carcinomas, and was associated with poor tumor differentiation [36]. Winn et al. [37] reported that the restoration of FZD9 signaling inhibited both cell proliferation and anchorage-

Table 3 Simulation results over 100 replications for Scenario 3

	ZP-DDN		RDN		ZR-RDN	
p	L_∞	L_F	L_∞	L_F	L_∞	L_F
Average true discovery rates(%)						
50	39.8 (39.5)	46.1 (47.3)	87.6 (14.7)	97.3 (7.4)	88.0 (11.0)	90.0 (12.4)
80	32.4 (41.9)	35.5 (47.9)	80.7 (14.8)	99.8 (2.5)	89.5 (8.7)	95.4 (7.2)
100	23.5 (40.2)	31.7 (46.9)	75.6 (20.3)	100.0(0.0)	84.0 (12.0)	99.1 (4.2)
120	16.0 (37.0)	16.0 (37.0)	52.9 (44.6)	68.0(47.1)	70.4 (26.8)	93.0 (24.8)
Average estimation errors in the elementwise L_∞ norm						
50	2.15 (0.03)	2.16 (0.01)	2.05 (0.15)	2.12 (0.08)	2.05 (0.17)	2.02 (0.15)
80	1.91 (0.02)	1.91 (0.01)	1.91 (0.12)	1.92 (0.04)	1.92 (0.12)	1.91 (0.08)
100	1.82 (0.02)	1.82 (0.00)	1.88 (0.12)	1.82 (0.00)	1.90 (0.12)	1.83 (0.04)
120	0.87 (0.00)	0.87 (0.00)	0.91 (0.09)	0.87 (0.00)	0.97 (0.11)	0.88 (0.05)

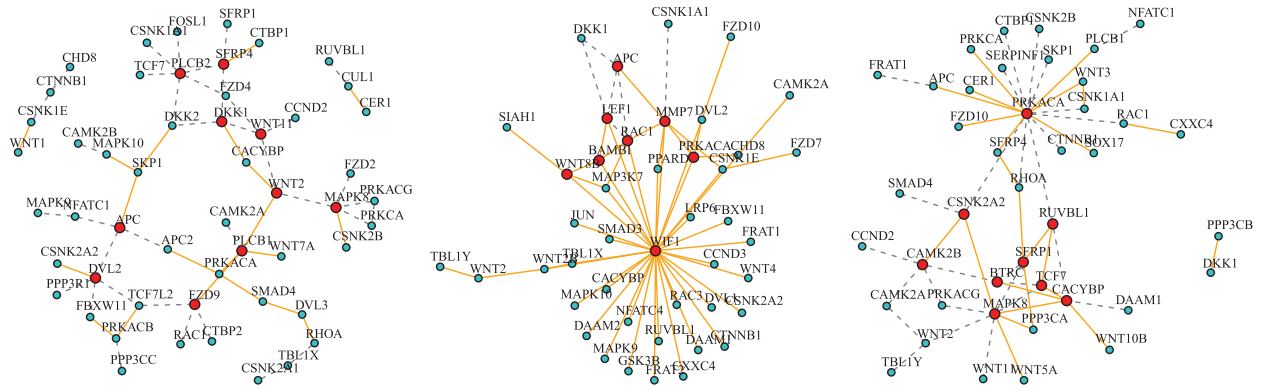

Fig. 5 Differential network estimated by different methods. Orange edges show an increase in conditional dependency from control group to lung cancer patient group; grey edges show a decrease. Red points stand for hub genes which have edges with more than 3 other genes. **a** C-RDN. **b** B-RDN. **c** DDN

independent growth, promoted cellular differentiation, and reversed the transformed phenotype in NSCLC. The overexpression of MMP7 was associated with tumor proliferation, and a poor prognosis in NSCLC [38]. RAC1 generally plays an important role in cancer progression and metastasis [39].

By comparing (a) and (b) in Fig. 5, we can see that the estimated differential network can be very different with/without considering the batch effect. Although it is inevitable to result in information loss in the discretization procedure for method B-RDN, [40] argued that this procedure can potentially improve the accuracy of the statistical analysis. In real data example, we recommend to use the B-RDN method to remove the batch effect despite the little information loss. At last we argue that statistical comparison of group difference in this biological network or pathway can provide new insight into the underlying lung cancer mechanism, which may further offer more effective targets for drug development.

To further interpret the underlying biological implications of the identified hub genes, we conducted Gene Ontology (GO) enrichment analysis. Table 5 shows the common GO terms enriched by C-RDN, B-RDN and DDN. The GO enrichment analysis is performed using R package "clusterProfiler" with the P-value adjusted by Benjamini-Hochberg method. It shows that our methods (C-RDN, B-RDN) have smaller P-value than DDN. The common molecular function and cellular component suggest that the change of frizzled binding, Wnt-protein binding and beta-catenin destruction complex are important in the etiology of lung cancer. These predictions are supported by the literatures [41–43] , which indicates that the proposed differential network model can provide biological meaningful underlying signals.

Table 4 Hub genes selected by different methods

DDN	PRKACA	MAPK8	CACYBP	CAMK2B	SFRP1	CSNK2A2	TCF7
	BTRC	RUVBL1					
C-RDN	PLCB2	DVL2	MAPK8	PLCB1	APC	WNT2	FZD9
	WNT11	DKK1	SFRP4				
B-RDN	WIF1	MMP7	RAC1	LEF1	APC	PRKACA	WNT8B
	BAMBI						

Discussion

A complex disease phenotype (e.g. diabetes, cancer) often reflects various pathobiological processes that interact in a network rather than the abnormality of a single gene. Such interactions are not static processes, instead they are dynamic in response to changing genetic, epigenetic and environmental factors, which further entails the analysis of differential network. In this paper, we propose adaptive estimation approaches for latent variable differential network model with the assumption that the true differential network is sparse, which do not require precision matrices to be sparse. The latent variable differential network model is fundamentally different from the existing ones in the literature in the sense that the differential structure in the unobserved latent variables are of primary interest. Theoretical analysis shows that the proposed methods achieve the same parametric convergence rate for both the difference of the precision matrices estimation and differential structure recovery, which means that the extra modelling flexibility comes at almost no cost of statistical efficiency. The unified latent variable differential network model provides deeper understanding of the unknown genomic mechanism than that among the observed variables.

Table 5 Gene Ontology (GO) enrichment analysis result

ID	Functional term	Ontology	Adjust *P*-value		
			C-RDN	B-RDN	DDN
GO:0016055	Wnt signaling pathway	BP	1.69×10^{-11}	2.96×10^{-6}	0.0022
GO:0198738	cell-cell signaling by wnt	BP	1.69×10^{-11}	2.96×10^{-6}	0.0022
GO:0060828	regulation of canonical Wnt signaling pathway	BP	1.49×10^{-9}	0.0012	0.0027
GO:0060070	canonical Wnt signaling pathway	BP	4.78×10^{-9}	0.0012	0.0027
GO:0030111	regulation of Wnt signaling pathway	BP	6.67×10^{-9}	0.0058	0.0091
GO:0005109	frizzled binding	MF	5.28×10^{-5}	0.0058	0.0091
GO:0007369	gastrulation	BP	0.0024	0.0058	0.0276
GO:0017147	Wnt-protein binding	MF	0.0025	0.0073	0.0286
GO:0060562	epithelial tube morphogenesis	BP	0.0068	0.0073	0.0290
GO:0003002	regionalization	BP	0.0074	0.0080	0.0331
GO:0035239	tube morphogenesis	BP	0.0082	0.0090	0.0332
GO:0001503	ossification	BP	0.0093	0.0131	0.0341
GO:0007389	pattern specification process	BP	0.0113	0.0131	0.0357
GO:0043393	regulation of protein binding	BP	0.0202	0.0175	0.0377
GO:0034329	cell junction assembly	BP	0.0205	0.0178	0.0382
GO:0030877	beta-catenin destruction complex	CC	0.0223	0.0377	0.0382
GO:0045216	cell-cell junction organization	BP	0.0229	0.0409	0.0402
GO:0034330	cell junction organization	BP	0.0259	0.0411	0.0408
GO:0071496	cellular response to external stimulus	BP	0.0281	0.0411	0.0418
GO:0071214	cellular response to abiotic stimulus	BP	0.0290	0.0421	0.0448
GO:0104004	cellular response to environmental stimulus	BP	0.0290	0.0450	0.0453
GO:0051098	regulation of binding	BP	0.0330	0.0474	0.0478
GO:0045992	negative regulation of embryonic development	BP	0.0341	0.0479	0.0495
GO:1903829	positive regulation of cellular protein localization	BP	0.0397	0.0489	0.0495
GO:1901990	regulation of mitotic cell cycle phase transition	BP	0.0409	0.0489	0.0498

BP: biological process; MF: molecular function; CC: cellular component

The current work could be extended in the following two aspects. First, in this paper, we consider the following optimization problem to directly estimate the difference matrix $\mathbf{\Delta}$:

$$\arg\min |\mathbf{\Delta}|_1, \quad \text{subject to}$$

$$\left| \hat{\boldsymbol{S}}^X \mathbf{\Delta} \hat{\boldsymbol{S}}^Y - \hat{\boldsymbol{S}}^X + \hat{\boldsymbol{S}}^Y \right|_\infty \leq \lambda_n,$$

where $\hat{\boldsymbol{S}}^X$ and $\hat{\boldsymbol{S}}^Y$ denote the rank-based estimators of the covariance matrices. The D-trace loss function [15, 44] can also be applied to to directly estimate the precision matrix difference. Thus, we may also consider the D-trace loss function to estimate the Gaussian copula and latent Gaussian copula differential graphical models. In specific, the difference matrix $\mathbf{\Delta}$ could be eatimated by:

$$\arg\min_{\mathbf{\Delta}} \frac{1}{2} \mathrm{Tr}\left(\mathbf{\Delta} \hat{\boldsymbol{S}}^X \mathbf{\Delta} \hat{\boldsymbol{S}}^Y\right) - \mathrm{Tr}\left(\mathbf{\Delta}\left(\hat{\boldsymbol{S}}^X - \hat{\boldsymbol{S}}^Y\right)\right) + \mathcal{G}_\lambda(\mathbf{\Delta}),$$

where $\lambda > 0$ is a regularization parameter and $\mathcal{G}_\lambda$ is a decomposable non-convex penalty function which has the form $\mathcal{G}_\lambda = \sum_{j,k} g_\lambda\left(\Delta_{jk}\right)$, such as smoothly clipped absolute deviation (SCAD) penalty [45]. The theoretical guarantees are still needed to be investigated, but we expect that the empirical performance could be comparable.

Second, for the latent Gaussian copula differential graphical model, we focus on the binary data. In fact, the methods can be extended to the discrete data with more than two categories. The properties of this procedure are left for future investigation as there are a lot of work still needed to be done.

Conclusions

The proposed latent variable differential network models are very flexible and provide deeper understanding of the unknown biological mechanism. It is demonstrated latent differential network models enjoy great advantages over existing models and thus are highly recommended in real application.

Additional file

Additional file 1: Contains the theoretical guarantee of of the proposed methods and proofs. (PDF 284 kb)

Abbreviations

CNV: Copy Number Variation; FDR: False Discovery Rate; GO: Gene Ontology; ROC: Receiver Operating Characteristic; SNP: Single Nucleotide Polymorphisms; TDR: True Discovery Rate; TNR: True Negative Rate; TPR: True Positive Rate

Funding

This work was supported by grants from the National Natural Science Foundation of China (grant number 81803336, 11801316, 11571080 and 81573259) and Natural Science Foundation of Shandong Province (ZR2018BH033). Publication of this article was sponsored by 81803336 grant. The funding body played no role in the design, writing or decision to publish this manuscript.

Availability of data and materials

The gene expression data set related to lung cancer is publicly available from the Gene Expression Omnibus at accession number GDS2771.

About this supplement

This article has been published as part of *BMC Bioinformatics Volume 19 Supplement 17, 2018: Selected articles from the International Conference on Intelligent Biology and Medicine (ICIBM) 2018: bioinformatics*. The full contents of the supplement are available online at https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-19-supplement-17.

Authors' contributions

YH and JJ contributed to the study design, analytical preparation and the writing of the manuscript. YH and JJ performed the simulation studies. JJ analyzed the data, LX, XZ and FX wrote and revised the manuscript. All authors read and approved this version of the manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Author details

[1]School of Statistics, Shandong University of Finance and Economics, 250014 Jinan, China. [2]Department of Computer Science, Hunter College, The City University of New York, 10065 New York, USA. [3]Ph.D. Program in Computer Science, The Graduate Center, The City University of New York, 10016 New York, USA. [4]School of Management, Fudan University, 200433 Shanghai, China. [5]School of Public Health, Shandong University, 250012 Jinan, China.

Published: 28 December 2018

References

1. Li H, Gui J. Gradient directed regularization for sparse gaussian concentration graphs, with applications to inference of genetic networks. Biostatistics. 2006;7(2):302–17.
2. Segal E, Friedman N, Kaminski N, Regev A, Koller D. From signatures to models: understanding cancer using microarrays. Nat Genet. 2005;37: 38–45.
3. Peng J, Wang P, Zhou N, Zhu J. Partial correlation estimation by joint sparse regression models. J Am Stat Assoc. 2009;104(486):735.
4. Cai T, Li H, Liu W, Xie J. Covariate-adjusted precision matrix estimation with an application in genetical genomics. Biometrika. 2013;100(1): 139–56.
5. Fuente ADL. From differential expression to differential networking–identification of dysfunctional regulatory networks in diseases. Trends Genet. 2010;26(7):326–33.
6. Ideker T, Krogan NJ. Differential network biology. Mol Syst Biol. 2012;8(1): 565.
7. Zhao SD, Cai T, Li H. Direct estimation of differential networks. Biometrika. 2014;101(2):253–68.
8. Tian D, Gu Q, Jian M. Identifying gene regulatory network rewiring using latent differential graphical models. Nucleic Acids Res. 2016;44(17):140.
9. Xia Y, Cai T, Cai T. Testing differential networks with applications to detecting gene-by-gene interactions. Biometrika. 2015;102(2):247–66.
10. Ji J, Yuan Z, Zhang X, Li F, Xu J, Liu Y, Li H, Wang J, Xue F. Detection for pathway effect contributing to disease in systems epidemiology with a case-control design. Bmj Open. 2015;5(1):006721.
11. Ji J, Yuan Z, Zhang X, Xue F. A powerful score-based statistical test for group difference in weighted biological networks. BMC Bioinformatics. 2016;17(1):86.
12. Yuan Z, Ji J, Zhang T, Liu Y, Zhang X, Chen W, Xue F. A novel chi-square statistic for detecting group differences between pathways in systems epidemiology. Stat Med. 2016;35(29):5512–24.
13. Yuan Z, Ji J, Zhang X, Xu J, Ma D, Xue F. A powerful weighted statistic for detecting group differences of directed biological networks. Sci Rep. 2016;6:34159.
14. Liu W. Structural similarity and difference testing on multiple sparse gaussian graphical models. Ann Stat. 2017;45(6):2680–2707.
15. Yuan H, Xi R, Chen C, Deng M. Differential network analysis via the lasso penalized d-trace loss. Biometrika. 2017;104:755–70.
16. He Y, Zhang X, Ji J, Liu B. Joint estimation of multiple high-dimensional gaussian copula graphical models. Aust N Z J Stat. 2017;59:289–310.
17. Ji J, He D, Feng Y, He Y, Xue F, Xie L. Jdinac: joint density-based non-parametric differential interaction network analysis and classification using high-dimensional sparse omics data. Bioinformatics. 2017;33(19): 3080–87.
18. Meinshausen N, Bühlmann P. High-dimensional graphs and variable selection with the lasso. Ann Stat. 2006;34:1436–62.
19. Yuan M, Lin Y. Model selection and estimation in the gaussian graphical model. Biometrika. 2007;94(1):19–35.
20. Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. Biostatistics. 2008;9(3):432–41.
21. Yuan M. High dimensional inverse covariance matrix estimation via linear programming. J Mach Learn Res. 2010;11(12):2261–86.
22. Cai T, Liu W, Luo X. A constrained ℓ_1 minimization approach to sparse precision matrix estimation. J Am Stat Assoc. 2011;106(494):594–607.
23. Guo J, Levina E, Michailidis G, Zhu J. Joint estimation of multiple graphical models. Biometrika. 2011;98(1):1–15.
24. Danaher P, Wang P, Witten DM. J R Stat Soc Ser B (Stat Methodol). 2014;76(2):373–97.
25. Liu H, Lafferty J, Wasserman L. The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. J Mach Learn Res. 2009;10(3):2295–328.
26. Liu H, Han F, Yuan M, Lafferty J, Wasserman L. High-dimensional semiparametric gaussian copula graphical models. Ann Stat. 2012;40(4): 2293–326.
27. Xue L, Zou H. Regularized rank-based estimation of high-dimensional nonparanormal graphical models. Ann Stat. 2012;40(5):2541–71.
28. He Y, Zhang X, Wang P, Zhang L. High dimensional Gaussian copula graphical model with FDR control. Comput Stat Data Anal. 2017;113: 457–74.
29. Fan J, Liu H, Ning Y, Zou H. High dimensional semiparametric latent graphical model for mixed data. J R Stat Soc. 2017;79(2):405–21.
30. Ying Y, Tao Q. Epigenetic disruption of the wnt/ß-catenin signaling pathway in human cancers. Epigenetics. 2009;4(5):307–12.
31. Boyd S, Parikh N, Chu E, Peleato B, Eckstein J. Distributed optimization and statistical learning via the alternating direction method of multipliers. Found Trends Mach Learn. 2011;3(1):1–122.
32. Mazieres J, He B, You L, Xu Z, Jablons DM. Wnt signaling in lung cancer. Cancer Lett. 2005;222(1):1–10.
33. Lee SM, Park J, Kim DS. Wif1 hypermethylation as unfavorable prognosis of non-small cell lung cancers with egfr mutation. Mol Cells. 2013;36(1): 69–73.

34. Brabender J, Usadel H, Danenberg KD, Metzger R, Schneider PM, Lord RV, Wickramasinghe K, Lum CE, Park J, Salonga D, et al. Adenomatous polyposis coli gene promoter hypermethylation in non-small cell lung cancer is associated with survival. Oncogene. 2001;20(27):3528–32.
35. Takahashi H, Ogata H, Nishigaki R, Broide DH, Karin M. Tobacco smoke promotes lung tumorigenesis by triggering ikkbeta- and jnk1-dependent inflammation. Cancer Cell. 2010;17(1):89.
36. Wei Q, Zhao Y, Yang ZQ, Dong QZ, Dong XJ, Han Y, Zhao C, Wang EH. Dishevelled family proteins are expressed in non-small cell lung cancer and function differentially on tumor progression. Lung Cancer. 2008;62(2):181–92.
37. Winn RA, Marek L, Han SY, Rodriguez K, Rodriguez N, Hammond M, Scoyk MV, Acosta H, Mirus J, Barry N. Restoration of wnt-7a expression reverses non-small cell lung cancer cellular transformation through frizzled-9-mediated growth inhibition and promotion of cell differentiation. J Biol Chem. 2005;280(20):19625.
38. Liu D, Nakano J, Ishikawa S, Yokomise H, Ueno M, Kadota K, Urushihara M, Huang CL. Overexpression of matrix metalloproteinase-7 (mmp-7) correlates with tumor proliferation, and a poor prognosis in non-small cell lung cancer. Lung Cancer. 2007;58(3):384–91.
39. Kaneto N, Yokoyama S, Hayakawa Y, Kato S, Sakurai H, Saiki I. Rac1 inhibition as a therapeutic target for gefitinib-resistant non-small-cell lung cancer. Cancer Sci. 2014;105(7):788–94.
40. McCall MN, Irizarry RA. Thawing frozen robust multi-array analysis (frma). BMC Bioinformatics. 2011;12(1):1.
41. Stewart DJ. Wnt signaling pathway in non-small cell lung cancer. J Natl Cancer Inst. 2014;106(1):356.
42. Nakayama S, Sng N, Carretero J, Welner R, Hayashi Y, Yamamoto M, Tan AJ, Yamaguchi N, Yasuda H, Li D. β-catenin contributes to lung tumor development induced by egfr mutations. Cancer Res. 2014;74(20):5891–902.
43. Rapp J, Jaromi L, Kvell K, Miskei G, Pongracz JE. Wnt signaling-lung cancer is no exception. Respir Res. 2017;18(1):167.
44. Wadsworth JL, Tawn JA. Sparse precision matrix estimation via lasso penalized d-trace loss. Biometrika. 2014;1(1):103–20.
45. Fan J, Li R. Variable selection via nonconvave penalized likelihood and its oracle properties. J Am Stat Assoc. 2001;96(456):1348–60.