**S1: detailed description of existing GGM related methods: GLASSO, JGL and HGLASSO**

Gaussian graphical models (GGMs) are widely used to reconstruct gene networks using gene expression data (Kumari, et al., 2016). The models assume that gene expression data on genes from each sample follows a multivariate normal distribution with mean and covariance matrix , where is a vector with elements and is a positive definite matrix. The conditional independence of two genes given other genes corresponds to a zero entry in the inverse covariance matrix (also called the precision or concentration matrix) (Lauritzen, 1996). Therefore, Gaussian graphical models have the advantage of reconstructing direct dependencies between genes that represent edges in the reconstructed network: an edge corresponds a non-zero entry in . A natural way to estimate is by maximizing the log-likelihood of the data. Under the Gaussian distribution, the log-likelihood function can be expressed as the following:

where is the sample covariance matrix, det is the determinant of a matrix, tr is the trace of a matrix. Define , then maximizing (1.1) with respect to leads to the maximum likelihood estimate of precision matrix . However, directly applying GGM to reconstruct GRN is not applicable due to two problems. First, since the number of samples () is generally much less than the number of genes () from gene expression data, the sample covariance matrix becomes singular and thus it is impossible to computing the likelihood function. Second, even if the sample covariance matrix is not singular, the elements (supposed to be zero) in the estimated precision matrix are in general not exactly equal to zero. For these reasons, Yuan and Lin (Lin, et al., 2007) proposed to maximize the regularized log-likelihood

where is a nonnegative tuning parameter, is the sum of absolute value of each element in . This log-likelihood (1.2) can be solved by the graphical lasso algorithm (GLASSO) (Friedman, et al., 2008). When is sufficiently large, we can get a sparse and positive definite estimate of from (1.2).

When applying GGMs to reconstruct gene regulatory networks, the underlying assumption is that each observation is drawn from the same distribution. However, when the gene expression data come from different tissues or under different treatments, this assumption is inappropriate. In this case, if one insists on modeling the gene expression data by one GRN, the results would be dubious and we cannot obtain the differential network we are interested in. A straightforward method to obtain the differential network is to reconstruct the network of each condition separately and then find the difference between them. However, this procedure ignores the similarity shared between GRNs across different tissues/treatments, which is critically important to reconstruct the GRNs, especially when the sample size is small. To reconstruct these dependent GRNs, Guo et al. (Guo, et al., 2011) proposed a joint penalized model using a hierarchical penalty and derived the convergence rate and sparsity properties of the resulting estimators. Danaher et al. (Danaher, et al., 2014) proposed a joint graphical lasso model (JGL) to estimate multiple GRNs simultaneously based on the fused graphical lasso or the group graphical lasso. In their models, the penalized log likelihood function is:

where is the number of conditions; , and are the sample size, the precision matrix and the sample covariance matrix for condition respectively; is the penalty function. For the fused graphical lasso, is:

where and are nonnegative tuning parameters. The first item encourages the off-diagonal elements of to be sparse, the second item encourages to be the same. For the group graphical lasso, is:

Here, the second item encourages to have similar sparsity pattern.

The above-mentioned methods do not impose any structural information of gene networks. That is, each gene has approximately the same number of interactions within the network, and each pair of nodes has equal probability to be an edge and all edges are independent. However, recent evidence points to scale-free properties in biological networks (Han, et al., 2004; van den Heuvel and Sporns, 2013), in which most genes interact with a few partners whereas a small proportion of genes, called hub genes, are densely-connected to many other genes (high connectivity). To incorporate hub genes in GRNs, Liu and Ihler (Liu and Ihler, 2011) replaced the regularization in (1.2) with a power law regularization and optimized the objective function by solving a sequence of iteratively reweighted regularization problems, where the regularization coefficients of nodes with high degree were reduced, which encouraged the appearance of hub genes. Tan et al. (Tan, et al., 2014) introduced a row-column overlap norm penalty to incorporate hub genes explicitly. In their model, called hub graphical lasso (HGLASSO), the penalized log likelihood function is:

In this formulation, the precision matrix was decomposed into two parts, namely and , where is a symmetric matrix that is encouraged to be sparse, is a matrix whose columns are encouraged to be either entirely zero or almost entirely non-zero through the norm penalization. The non-zero columns of correspond to hub genes.

**S2. Details for solving (2.9).**

Since the objective function is completely separable with respect to the variables , updating can be achieved by updating each variable and is given by:

Let denote the eigen decomposition of , the solution is given by , where is the diagonal matrix with diagonal element as follows: (see (Witten and Tibshirani, 2009) for the derivation).

The objective function is completely separable with respect to each pair of matrix elements . Therefore, it is equivalent to solve optimization problem.

This form belongs to a class of fused lasso problems. It can be solved in operations, see (Liu, et al., 2010) for more details. When , there is a very simple closed form solution for the case , see (Danaher, et al., 2014).

where , when , the solution to (2.13) can be obtained through soft-thresholding by .

The objective function is column separable, Let be the column of matrix , we solve

This problem looks like an elastic net problem (Zou and Hastie, 2005), the difference between this problem and elastic net is that the l2 norm is not squared. The solution is given in (Danaher, et al., 2014):

Let

where is the soft-thresholding operation. We give a derivation of (a.7) in S2.1.

Although the objective function is not separable with respect to the variables , updating , can be achieved as follows:

The derivation of (a.8~a.11) are given in S2.2.

**S2.1**. **Derivation of (a.7)**

Let . If , take the sub gradient of the objective function of (2.16), and set it to 0.

where is the sub gradient of with respect to .

It is easy to check when , it also satisfy the above formula.

where is the sub gradient of

It is easy to check when , it also satisfy the above formula.

**S2.2**. **Derivation of (a.8~a.11)**

where .

the Lagrange form of the above formula is

for each , subtract the last three equalities from the first equality, we get

add up the above equalities for all

Therefore

**S3.1. Proof of theorem 1.**

Let , let denote the restriction of the matrix to the set , that is

Assume is a feasible solution of (2.1), then is also a feasible solution to (2.1). We want to show that if (2.11) holds, then the objective value of (2.1) evaluated at is smaller than the objective value of (2.1) evaluated at . By Fischer’s inequality,

We need only to prove

As and , we need only to prove

where the last inequality follows from the sufficient condition.

**S3.2. Proof of theorem 2.**

Let , let , then is a feasible solution for (2.1), we want to show that if (2.12) holds, then the objective value of (2.1) evaluated at is smaller than the objective value of (2.1) evaluated at . As

We need only to prove,

 (3.3)

in order to prove above inequality, we need only to prove

That is

According to Cauchy inequality,

we need only to prove . This is true when (2.12) holds.

**S4. Comparison of JRmGRN and other GGM based methods under different network setting.**

To compare the performance JRmGRN and other GGM based methods (glasso, JGL and hglasso) under different network setting, we simulated the networks with varying sparsity and similarity in two tissues/conditions. The results are shown in following figure (Fig. S4). It is noticeable that our method performed very well persistently under different Sparsity and networks similarity. JGL, which is capable of constructing two networks corresponding to two conditions/tissues jointly without hubs, was the second best method.



Fig. S4. Precision-recall curve of JRmGRN and other GGM based methods (glasso, JGL and hglasso) for different networks setting. ‘E\_sparsity’ is the sparsity of elementary network; ‘H\_sparsity’ is the sparsity of Hub network, the last parameter shown in title is the difference of two elementary networks.

**S5. Comparison of the capability of JRmGRN and other GGM based methods for identifying tissue/condition specific edges**

Table S5. Comparison of the capability of JRmGRN and other GGM based methods for identifying tissue/condition specific edges

|  |  |  |  |
| --- | --- | --- | --- |
|  | 108 True Unique edges | 275 True Unique edges |  496 True Unique edges |
|  | TP | FP | Precision | Recall | F1 score | TP | FP | Precision | Recall | F1 score | TP | FP | Precision | Recall | F1 score |
| JRmGRN | 17 | 36 | 0.321 | 0.157 | 0.211 | 38 | 65 | 0.369 | 0.138 | 0.201 | 81 | 187 | 0.302 | 0.163 | 0.212 |
| GLASSO | 19 | 75 | 0.202 | 0.176 | 0.188 | 50 | 195 | 0.204 | 0.182 | 0.192 | 76 | 388 | 0.164 | 0.153 | 0.158 |
| JGL | 18 | 51 | 0.261 | 0.167 | 0.204 | 35 | 79 | 0.307 | 0.127 | 0.180 | 69 | 176 | 0.282 | 0.139 | 0.186 |
| HGLASSO | 13 | 42 | 0.236 | 0.120 | 0.159 | 47 | 154 | 0.235 | 0.170 | 0.197 | 55 | 412 | 0.118 | 0.110 | 0.114 |

**S6. Comparison of the capability of JRmGRN and other HGLASSO for identifying Hub genes**

Table S6. Comparison of the capability of JRmGRN and other HGLASSO for identifying Hub genes

|  |  |  |  |
| --- | --- | --- | --- |
|  | 80 genes: 5 hubs | 150 genes: 8 hubs  | 300 genes: 12 hubs |
|  | TP | FP | Precision | Recall | F1 score | TP | FP | Precision | Recall | F1 score | TP | FP | Precision | Recall | F1 score |
| JRmGRN |  4 | 2 | 0.667 | 0.80 | 0.727 | 6 | 0 | 1.00 | 0.75 | 0.857 | 12 | 2 | 0.857 | 1.00 | 0.923 |
| HGLASSO |  5 | 4 | 0.556 | 1.00 | 0.715 | 7 | 5 | 0.583 | 0.875 | 0.699 | 12 | 7 | 0.632 | 1.00 | 0.775 |

**S7. Comparison of the performance of JRmGRN and other GGM based methods under four conditions.**

****

Fig. S7. Comparison of the performance of JRmGRN and other GGM based methods under four conditions.

**S8. The number of different types of edge identified by four methods**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Common edges | Light only edges | Shade only edges |
| JRmGRN | 3702  | 150  | 52 |
| GLASSO | 192 | 1813  | 2304  |
| JGL | 3495  | 300  | 299 |
| HGLASSO | 35  | 902  | 2545 |

**S9. Comparison of JRmGRN and HGLASSO on hub gene identification (positive means related to far-red light and shade)**

|  |  |
| --- | --- |
| JRmGRN | HGLASSO |
| Known positive gene | Currently unknown  | Known positive gene | Currently unknown  |
| "PD1" "ELIP1" "POP12" "BLH10" "PLPC" "WAV2" "CCL" "PEX11B" "TCP11"  | "AT1G62310""PLIM2a" "AT3G45260""BZO2H3""NAC102""AT3G15570" | "SOB7""EIP6""ELIP2""TED\_5""TCP11" | "ERF11" "ETC2""WRKY33" "GATA8" "PFG2" "AT5G61590""ERF104""MYBR1" "GA4H" |

**References:**

Danaher, P., Wang, P. and Witten, D.M. The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2014;76(2):373-397.

Friedman, J., Hastie, T. and Tibshirani, R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 2008;9(3):432-441.

Guo, J.*, et al.* Joint estimation of multiple graphical models. *Biometrika* 2011:asq060.

Han, J.-D.J.*, et al.* Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature* 2004;430(6995):88-93.

Kumari, S.*, et al.* Bottom-up GGM algorithm for constructing multilayered hierarchical gene regulatory networks that govern biological pathways or processes. *BMC Bioinformatics* 2016;17:132.

Lauritzen, S.L. Graphical models. Clarendon Press; 1996.

Lin, Y.-H.*, et al.* Enhancement of ferromagnetic properties in Bi Fe O 3 polycrystalline ceramic by La doping. *Applied physics letters* 2007;90(17):172507.

Liu, J., Yuan, L. and Ye, J. An efficient algorithm for a class of fused lasso problems. In, *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM; 2010. p. 323-332.

Liu, Q. and Ihler, A. Learning scale free networks by reweighted l1 regularization. In, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. 2011. p. 40-48.

Tan, K.M.*, et al.* Learning graphical models with hubs. *Journal of Machine Learning Research* 2014;15(1):3297-3331.

van den Heuvel, M.P. and Sporns, O. Network hubs in the human brain. *Trends in cognitive sciences* 2013;17(12):683-696.

Witten, D.M. and Tibshirani, R. Covariance‐regularized regression and classification for high dimensional problems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2009;71(3):615-636.

Zou, H. and Hastie, T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2005;67(2):301-320.