

# Analyzing and learning sparse and scale-free networks using Gaussian graphical models

Melih S. Aslan<sup>1</sup> · Xue-Wen Chen<sup>1</sup> · Hong Cheng<sup>2</sup>

Received: 22 February 2016 / Accepted: 9 April 2016 / Published online: 23 April 2016  
© Springer International Publishing Switzerland 2016

**Abstract** In this paper, we consider the problem of fitting a sparse precision matrix to multivariate Gaussian data. The zero elements in the precision matrix correspond to conditional independencies between variables. We focus on the estimation of a class of sparse precision matrix which represents the scale-free networks. It has been demonstrated that some of the important networks display features similar to scale-free graphs. We propose a new log-likelihood formulation, which promotes the sparseness of the precision matrix as well as the topological structure of scale-free networks. To optimize this new energy formulation, the alternating direction method of multipliers form is used with the general  $L_1$ -regularized loss optimization. We tested our proposed method on various databases. Our proposed method exhibits better estimation performance with various number of samples,  $N$ , and different selection of sparsity parameter,  $\rho$ .

**Keywords** Gaussian networks · Scale-free networks · ADMM · Sparse precision matrix

## 1 Introduction

Undirected graphical models study the relationship between a set of variables. The Gaussian networks are a way of study

to extract undirected relationships among the data. In the Gaussian networks, the variables are assumed to follow the multivariate Gaussian distribution with mean,  $\mu$ , and covariance matrix,  $\Sigma$ .

Gaussian models aim to find out the dependencies between variables. The relationship between variables in these networks can be represented using the precision matrix ( $\Omega$ ) of data sets and a graph in which the nodes and edges represent the variables and their conditional dependencies, respectively. The estimation of precision matrices from data is a problem of many applications and fields such as health, finance and biology. Zero elements reveal the conditional independencies of corresponding variables. If the  $ij$ th component of  $\Omega$  is zero, then variables  $i$  and  $j$  are conditionally independent, given the other variables.

The problem formulation can be summarized as follows: Let  $D$  represent a data set of cases,  $Z$  represent a set of variables,  $B_s = (Z, E)$  be a graph representing interactions between components of  $Z$ .  $B_s$  is any set in all possible networks,  $B$ . If  $E$  consists of an edge between these two variables,  $Z_i$  and  $Z_j$ , then these variables are said to be dependent. The objective is to find the best network which maximizes  $P(B_s|D)$  or  $P(B_s, D)$ . In another definition, the aim is to estimate the best network  $B_s$  which can be described as follows:

$$B_s^* = \arg \min_{B_s} -P(B | D). \quad (1)$$

$P(B|D)$  is modeled using the Gaussian distribution in the Gaussian networks. The inverse of the covariance matrix can be estimated by the log-likelihood estimation with respect to the covariance matrix as follows:

$$\Omega^* = \arg \min_{\Omega^*} (\mathbf{tr}(\Sigma\Omega) - \log |\Omega|), \quad (2)$$

Invited and extended version of our paper [1] presented at the International Conference on Data Science and Advanced Analytics, Shanghai, China, October 30–November 1, 2014.

✉ Melih S. Aslan  
melih.aslan@wayne.edu

<sup>1</sup> Computer Science Department, Wayne State University, Detroit, MI 48202, USA

<sup>2</sup> University of Electronic Science and Technology of China, Chengdu, Sichuan, China

where  $\Omega = \Sigma^{-1}$  is the precision matrix and  $\mathbf{tr}$  denotes the trace.

The number of nonzero elements in the precision matrix and number of edges in the structure are equivalent measures of complexity. Several methods have been proposed to obtain the sparse precision matrix and tackle the existing challenges. For instance, Dempster [2] proposed the idea of the precision matrix estimation. The number of edges in the structure, the number of parameters and the number of nonzero elements in the precision matrix are equivalent measures of complexity in the Gaussian graphical models [3]. Therefore, many studies have been focusing on the sparseness of the precision matrix recently. In the literature, there are various derivations of the following likelihood formulation to estimate the sparse precision matrix:

$$\Omega^* = \arg \min_{\Omega^*} (\mathbf{tr}(\Sigma \Omega) - \log |\Omega| + \rho \|\Omega\|), \quad (3)$$

where  $\Omega = \Sigma^{-1}$  is the precision matrix,  $\mathbf{tr}$  denotes the trace,  $\rho > 0$  is a scalar parameter which controls the size of the penalty, hence the sparsity of the solution,  $\|\Omega\|$  is the  $L_1$  norm which is the sum of absolute values of  $\Omega$ .

Meinshausen and Bühlmann [4] proposed a method to perform edge selection at each node in the graph using the least absolute shrinkage and selection operator (LASSO) which was proposed by Tibshirani [5]. They demonstrated that their estimate is shown to be both more accurate and computationally much more efficient than the traditional forward selection maximum-likelihood estimation (MLE) strategy. The forward MLE estimation has poor accuracy when the number of nodes in the graph is comparable to the number of observations. Furthermore, when the number of observations is remarkably less than the number of nodes, the empirical covariance  $\Sigma$  is singular so that we cannot access the required information about interactions between nodes.

In this context, several algorithms have been proposed such as Glasso (by Friedman et al. [6]), SPICE (by Rothman et al. [7]), SCAD (by Fan et al. [8]), COVSEL (by Banerjee et al. [9]) and CLIME (by Cai et al. [10]). Friedman et al. [6] estimated sparse graphs by a lasso penalty applied to the precision matrix. They proposed a well-known algorithm, named as Glasso. They used the blockwise coordinate descent approach in Banerjee et al. [9] as a launching point and proposed a new algorithm for the rest of the problem solution. They started with a given covariance matrix and regenerated data from the given covariance. The new covariance matrix is estimated from the generated data. Then for each variable they solved the lasso problem ( $\min_{\beta} \{0.5 \|\Sigma_{11}^{1/2} \beta - b\| + \rho \|\beta\|_1\}$  where  $b = \Sigma_{11}^{-1/2} s_{12}$ ,  $\Sigma = \begin{bmatrix} \Sigma_{11} & \sigma_{12} \\ \sigma_{12}^T & \sigma_{22} \end{bmatrix}$ ,  $S = \begin{bmatrix} S_{11} & s_{12} \\ s_{12}^T & s_{22} \end{bmatrix}$ ) using the blockwise

coordinate descent method. This process produces a  $p - 1$  vector solution  $\hat{\beta}$  which is used to calculate  $\sigma_{12} = \sigma_{11} \hat{\beta}$ . This process continues until convergence. They stated that their algorithm is 30 to 4000 times faster than other competing methods.

Later, Dahl et al. [11] solved the likelihood estimation problem using Newton's method and conjugated gradient method. Their proposed method solves problems where a sparse structure of  $\Sigma^{-1}$  is known a priori. Recently, Honorio et al. [3] proposed a method promoting the variable selection in addition to the sparseness of the precision matrix. In this method the sparseness is imposed not only at the edge level but also at the important variable selection level. They applied the block coordinate descent method and compared the results with the graphical lasso and covariance selection methods. Yuan et al. [12] proposed a penalized-likelihood method which has the model selection and parameter estimation simultaneously in the precision matrix estimation. They ensured that the estimator of the precision matrix is positive definite. D'aspremont et al. [13] proposed a likelihood formulation which penalizes the cardinality of the precision matrix. To solve the cardinality penalty term, they derived a convex relaxation solution. Other techniques have been developed in the similar direction and can be found for instance in [14–16] and the references therein.

One of the major limitations in traditional precision matrix estimation methods is that they aim the sparsity uniformly on each variable. In reality, however, most networks display scale-free properties [17]. Hence, the traditional methods would give poor performance to estimate the topology of the specific networks. Hub-dominated networks are usually dominated by a relatively small number of nodes (hubs) which are connected to many other nodes, and these networks are considered resistant to accidental failures but extremely vulnerable to coordinated attacks. Some examples of such data structures can be described as the interactions between research collaborations, accesses to the World Wide Web, role plays in Hollywood, etc. While many studies have been carried out for sparse network learning, little is done on the learning of the sparse Gaussian graphical models aiming to preserve properties of networks which are believed to be scale-free or have dominating hubs. We prefer to name both topologies as 'scale-free' networks for simplicity in this paper.

Liu et al. [18] proposed an  $L_1$  regularization formulations, where coefficients of nodes with higher degree are decreased. They proposed a formulation to promote the features of hub-dominated networks. One of the drawbacks of this paper is that the proposed formulation is nonconvex which does not guarantee the convergency. Also, they tested their method on relatively very small data sets (e.g.,  $p < 120$ ). In this paper our problem can be described as follows: Given a set of data, we solve a convex likeli-

hood formulation to make the resulting graph as sparse as possible, whereas the hub-dominated graphical topology is protected.

Recently, analyzing and using scale-free topology and graphical models improves the computational costs and accuracy. For instance, Liu [19] employed the scale-free network to represent the inter-individual interactions in particle swarm optimization (PSO) algorithm. They reported that in contrast to the traditional PSO with fully connected topology or regular topology, the scale-free topology incorporates the diversity of individuals in searching and information dissemination ability, leading to a quite different optimization process. They obtained results with respect to several standard test functions. The outcome demonstrated that their updated PSO algorithm gives better balance between the convergence speed and the optimum quality than the traditional PSO algorithms. They further explored that the cooperation of hub and regular nodes plays a crucial role in optimizing the convergence process. Zhang and Huang [20] studied the combined impact of reinstalling system and network topology on the spread of computer viruses over the Internet. Their experiments show that the virus-free equilibrium is globally asymptotically stable when its spreading threshold is less than one; nevertheless, it is proved that the viral equilibrium is permanent if the spreading threshold is greater than one. The impacts of different model parameters on spreading threshold are analyzed. Haoran [21] proposed a new scale-free topology model which has both fault tolerance against random faults and intrusion tolerance against selective remove attacks at the same time. The optimal scale-free topology which can keep the fault tolerance and maximize intrusion tolerance is obtained through analyzing the effect of topological degree distribution on these properties of topological fault tolerance and topological intrusion tolerance. The simulation results show that their method also can reduce their fragility for selective remove attacks and further prolong its lifetime.

The major contributions of this paper can be listed as follows: (i) We propose a novel structure learning method which is capable of working on the scale-free networks. (ii) The proposed convex formulation emphasizes the sparsity of the precision matrix and features of the scale-free topology. (iii) The proposed method is able to guard the hubs since hub-dominated networks are considered extremely vulnerable to coordinated attacks. One may diminish the connections on the hubs when the penalty constant is chosen for only the sparsity of the precision matrix.

The rest of the paper is organized as follows. In Sect. 2, the basic formulation of the precision matrix estimation problem and the proposed method is presented. We show the numerical performance of the proposed method and some of alternatives in Sect. 3 using synthetic and real data sets. We conclude our paper in Sect. 4.

## 2 Method

In this paper, we use the alternating direction method of multipliers (ADMM) form [22,23] to solve our proposed log-likelihood formulation [1]. First, we start to explain and derive a simple  $L_1$  problem using ADMM solution. For a generic  $L_1$  problem, the objective is written as follows:

$$\text{minimize } l(x) + \rho \|x\|_1, \tag{4}$$

where  $l$  is any convex loss function. In ADMM form, this problem is written as

$$\begin{aligned} &\text{minimize } l(x) + g(y) \\ &\text{subject to } x - y = 0, \end{aligned} \tag{5}$$

where  $g(y) = \rho \|y\|_1$ . The new updates are estimated as follows:

$$\begin{aligned} x^{k+1} &= \arg \min_x (l(x) + (\kappa/2 \|x - y^k + u^k\|_2^2)), \\ y^{k+1} &= S_{\rho/\kappa}(x^{k+1} + u^k), \\ u^{k+1} &= u^k + x^{k+1} - y^{k+1}, \end{aligned} \tag{6}$$

where the soft thresholding operator,  $S$ , is defined as follows:

$$S_\gamma(a) = \begin{cases} a - \gamma & a > \gamma \\ 0 & |a| \leq \gamma \\ a + \gamma & a < -\gamma. \end{cases} \tag{7}$$

In a scale-free network, some nodes, which are named as hubs in this paper, have a tremendous number of connections to other nodes, whereas most nodes, which are named as normal or regular nodes, have just a few. The hubs can have considerably much larger number of connections with respect to the normal nodes (e.g., hundreds, thousands or even millions of links). Therefore, these networks appear to have no scale [17].

In our problem the variables can be defined as  $Z = \{Z_r \cup Z_h\}$  where  $Z_r$  and  $Z_h$  represent the sets of the normal variables and hubs, respectively. The number of variables in  $Z_r$  and  $Z_h$  classes can be represented as  $p_r$  and  $p_h$ , respectively. We assume that the precision matrix consists of two classes as  $\Omega = [\Omega_r + \Omega_h]$  where  $\Omega \in R^{p \times p}$ ,  $\Omega_r \in R^{p_r \times p_r}$  and  $\Omega_h \in R^{p_h \times p_h}$ .

We propose a new log-likelihood formulation which promotes the sparseness of the precision matrix and features of the scale-free topology. Given a sample covariance matrix  $\Sigma$ , we estimate a precision matrix  $\Omega^*$  for  $p$  variables as follows:

$$\begin{aligned} \Omega^* &= \arg \min_{\Omega^*} (\text{tr}(\Sigma \Omega) - \log |\Omega| + \|G * \Omega\|_1), \\ &\text{subject to } : \Omega_r - Y_r = 0 \quad \text{and} \quad \Omega_h - Y_h = 0, \end{aligned} \tag{8}$$

where  $Y_r \in R^{p \times p}$ ,  $Y_h \in R^{p \times p}$ ,  $G \in R^{p \times p}$ ,  $G = [\rho_{ij}]$ ,  $i = \{1, 2, \dots, p\}$ ,  $j = \{1, 2, \dots, p\}$  and  $\rho_{ij} = \frac{\rho_i + \rho_j}{2}$ . The augmented Lagrangian (using the scaled dual variable) can be written as:

$$L(\Omega, Y, U) = \text{tr}(\Sigma\Omega) - \log |\Omega| + \|G * Y\|_1 + \frac{\kappa}{2} \|\Omega_r - Y_r + U_r\|_F^2 + \frac{\kappa}{2} \|\Omega_h - Y_h + U_h\|_F^2, \tag{9}$$

where  $\kappa = 1$  in our method. The scaled form of ADMM for this problem is:

$$\begin{aligned} \Omega^{k+1} &= \arg \min_{\Omega} L(\Omega, Y^k, U^k), \\ Y^{k+1} &= \arg \min_Y L(\Omega^{k+1}, Y, U^k), \\ \text{and } U^{k+1} &= U^k + \Omega^{k+1} - Y^{k+1}, \end{aligned} \tag{10}$$

where  $k$  is the iteration number. Using the scaled form of ADMM, the optimization steps of  $\Omega$ ,  $Y$  and  $U$  can be described in the following three subsections.

**$\Omega$ -optimization :**

$$\begin{aligned} \Omega^{k+1} &= \arg \min_{\Omega} L(\Omega, Y^k, U^k) \\ &= \arg \min_{\Omega} (\text{tr}(\Sigma\Omega) - \log |\Omega| + \frac{\kappa}{2} \|\Omega_r - Y_r^k + U_r^k\|_F^2 + \frac{\kappa}{2} \|\Omega_h - Y_h^k + U_h^k\|_F^2), \end{aligned} \tag{11}$$

where  $\Omega = [\Omega_r + \Omega_h]$ ,  $Y = [Y_r + Y_h]$  and  $U = [U_r + U_h]$ . After this point,  $\Omega$ -optimization step can be solved using an analytic method. The first-order optimality condition is that the gradient should vanish as follows:

$$\Sigma - \Omega^{-1} + \kappa(\Omega_r - Y_r^k + U_r^k) + \kappa(\Omega_h - Y_h^k + U_h^k) = 0, \tag{12}$$

$$\kappa\Omega - \Omega^{-1} = \kappa(Y^k - U^k) - \Sigma. \tag{13}$$

$\Omega$  is constructed to satisfy this condition and thus minimizes the  $\Omega$ -minimization objective. First, take the orthogonal eigenvalue decomposition of the right-hand side,

$$\kappa(Y^k - U^k) - \Sigma = Q\Lambda Q^T, \tag{14}$$

where  $\Lambda$  is the diagonal matrix which contains the eigenvalues of  $\kappa(Y^k - U^k) - \Sigma$ , and  $Q$  is the matrix which contains eigenvectors of  $\kappa(Y^k - U^k) - \Sigma$ . Note that  $QQ^T = Q^TQ = I$  where  $I$  is the identity matrix. Let us multiply the right side of Eq. 14 with  $Q^T$  on the left and  $Q$  on the right side as follows:

$$Q^T(\kappa(Y^k - U^k) - \Sigma)Q = Q^TQ\Lambda Q^TQ, \tag{15}$$

$$\kappa\tilde{\Omega} - \tilde{\Omega}^{-1} = \Lambda, \tag{16}$$

where  $\tilde{\Omega} = Q^T\Omega Q$ . We can now construct a diagonal solution of this equation, i.e., find positive numbers  $\tilde{\Omega}_i$  that satisfy  $\kappa\tilde{\Omega} - \frac{1}{\tilde{\Omega}} = r_i$ , where  $r_i$  is the corresponding eigenvalue. By the quadratic formula, each elements can be estimated as  $\tilde{\Omega}_{ii} = \frac{r_i + \sqrt{r_i^2 + 4\kappa}}{2\kappa}$ , where the solution is always positive since  $\kappa > 0$ . It follows that  $\Omega = Q\tilde{\Omega}Q^T$  satisfies the optimality condition  $\kappa(Y^k - U^k) - \Sigma = Q\Lambda Q^T$ , so this is the solution to the  $\Omega$ -minimization.

**Y-optimization**

$$\begin{aligned} Y^{k+1} &= \arg \min_Y L(\Omega^{k+1}, Y, U^k) \\ &= \arg \min_Y (\|G * Y\|_1 + \frac{\kappa}{2} \|\Omega_r^{k+1} - Y_r + U_r^k\|_F^2 + \frac{\kappa}{2} \|\Omega_h^{k+1} - Y_h + U_h^k\|_F^2). \end{aligned} \tag{17}$$

One can easily find that

$$\|G * Y\|_1 = \rho_r \|Y_r\|_1 + \rho_h \|Y_h\|_1, \tag{18}$$

where  $\rho_r$  and  $\rho_h$  are the specific constants assigned to  $Z_r$  and  $Z_h$ , respectively. In this problem, we need to ensure that the connections on the hubs will not be lost in the scale-free networks when we seek to have a sparse precision matrix. Therefore, we suggest to use two penalty parameters for the scale-free networks. We offer that constants for the hubs can be accepted as  $\rho_h = -\rho_r/c$  where we choose  $c = N$  in our method. This representation of the proposed formulation aims to minimize  $\|\Omega_r\|_1$ , which promotes the sparseness, and to maximize  $\|\Omega_h\|_1$ , which promotes the presence of hubs. Hence, the scaled form of ADMM for this problem can be written as follows:

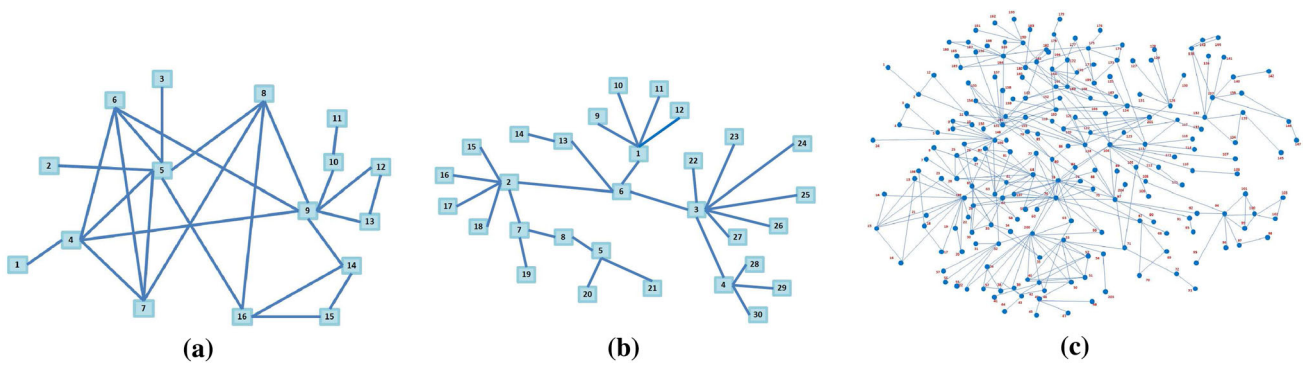
$$\begin{aligned} Y^{k+1} &= \arg \min_Y L(\Omega^{k+1}, Y, U^k) \\ &= \arg \min_Y (\rho_r \|Y_r\|_1 + \rho_h \|Y_h\|_1 + \frac{\kappa}{2} \|\Omega_r^{k+1} - Y_r + U_r^k\|_F^2 + \frac{\kappa}{2} \|\Omega_h^{k+1} - Y_h + U_h^k\|_F^2). \end{aligned} \tag{19}$$

Since

$$\begin{aligned} Y^{k+1} &= \arg \min_Y (\rho_r \|Y_r\|_1 + \frac{\kappa}{2} \|\Omega_r^{k+1} - Y_r + U_r^k\|_F^2) \\ &\quad + \arg \min_Y (\rho_h \|Y_h\|_1 + \frac{\kappa}{2} \|\Omega_h^{k+1} - Y_h + U_h^k\|_F^2), \end{aligned}$$

the final  $Y$ -optimization can be written as follows:

$$Y^{k+1} := [S_{\rho_r/\kappa}(\Omega_r^{k+1} + U_r^k) + S_{\rho_h/\kappa}(\Omega_h^{k+1} + U_h^k)], \tag{20}$$



**Fig. 1** Synthetic scale-free networks with **a**  $p = 16$ , **b**  $p = 30$ . **c** One of the real scale-free network [24] with  $p = 204$

which corresponds to  $Y^{k+1} = [Y_r^{k+1} + Y_h^{k+1}]$  and where  $S$  is the soft thresholding operator.

**U-optimization** The scaled form of ADMM for this problem is:

$$U^{k+1} = U^k + \Omega^{k+1} - Y^{k+1}. \tag{21}$$

We can accept this formulation as follows:

$$U^{k+1} = U_r^k + \Omega_r^{k+1} - Y_r^{k+1} + U_h^k + \Omega_h^{k+1} - Y_h^{k+1}, \tag{22}$$

which corresponds to  $U^{k+1} = [U_r^{k+1} + U_h^{k+1}]$ . Our method is summarized as in Algorithm 1.

**Algorithm 1** Our proposed algorithm

- 
- Input: Data samples,  $D$ , penalty constants,  $\rho_r$  and  $\rho_h$ .  
 Output: Precision matrix,  $\Omega$ .
1. The user defines the number of hubs,  $\tau_{hubs}$ .
  2. Assign the penalty constant  $\rho$  as  $\rho_r = \rho$  and  $\rho_h = -\rho/c$ , where  $c > 0$ .
  3. Estimate the covariance matrix,  $\Sigma$  from the samples.
  4. Do “while  $k < N_{iter}$ ” or “until the optimization is converged”:
    - 4.a)  $\Omega^{k+1}$  estimation begins.
    - 4.b) Find the elements of the hubs,  $\Omega_h$ , and the normal nodes,  $\Omega_r$ . Estimate the nonzero elements for each row/column in the precision matrix as follows:  $N_i = \sum_{j=1}^p (|\Omega_{ij}| > 0)$  for  $i = \{1, \dots, p\}$ . Select  $\tau_{hubs}$  number of maximum values in  $N_i$  and assign the corresponding elements of  $\Omega$  to  $\Omega_h$ . The rest of elements are classified as  $\Omega_r$ . Classify the matrices based on this rule as  $\Omega = [\Omega_r + \Omega_h]$ ,  $Y = [Y_r + Y_h]$  and  $U = [U_r + U_h]$ .
    - 4.c) Assign  $\rho_h$  and  $\rho_r$  for the hubs and normal nodes, respectively.
    - 4.d)  $Y^{k+1}$  estimation begins.
    - 4.e)  $U^{k+1}$  estimation begins.
    - 4.f) Increase “ $k$ ” and return to the step “4.”
- 

**3 Numerical results**

To evaluate the proposed method, we use structures of given scale-free networks that have the ground truth of the precision matrix. The precision matrix is normalized using the

similar approach described in [8]. Nonzero off-diagonal elements in  $\Omega$  are regenerated uniformly from intervals  $[-1 : -0.5] \cup [0.5 : 1]$ . The value of each diagonal element is set as a factor of the sum of the absolute values of its corresponding row elements. A constant value is selected to ensure the positive definiteness of  $\Omega$ . Finally, each row is divided by its corresponding diagonal entry so that the precision matrix has diagonal values of ones. Using this precision matrix, the covariance matrix,  $\Sigma$ , is obtained to generate various number of samples. We simulate data sets each with various number of samples i.i.d. generated from multivariate Gaussian distribution  $N(0, \Sigma^{-1})$ .

The results of the proposed and some of the state-of-the-art methods are evaluated with different formulations such as sensitivity, specificity and Matthew correlation coefficient ( $MCC$ ) defined as follows:

$$\text{Sensitivity} = \frac{TP}{TP + FN}, \tag{23}$$

$$\text{Specificity} = \frac{TN}{TN + FP}, \tag{24}$$

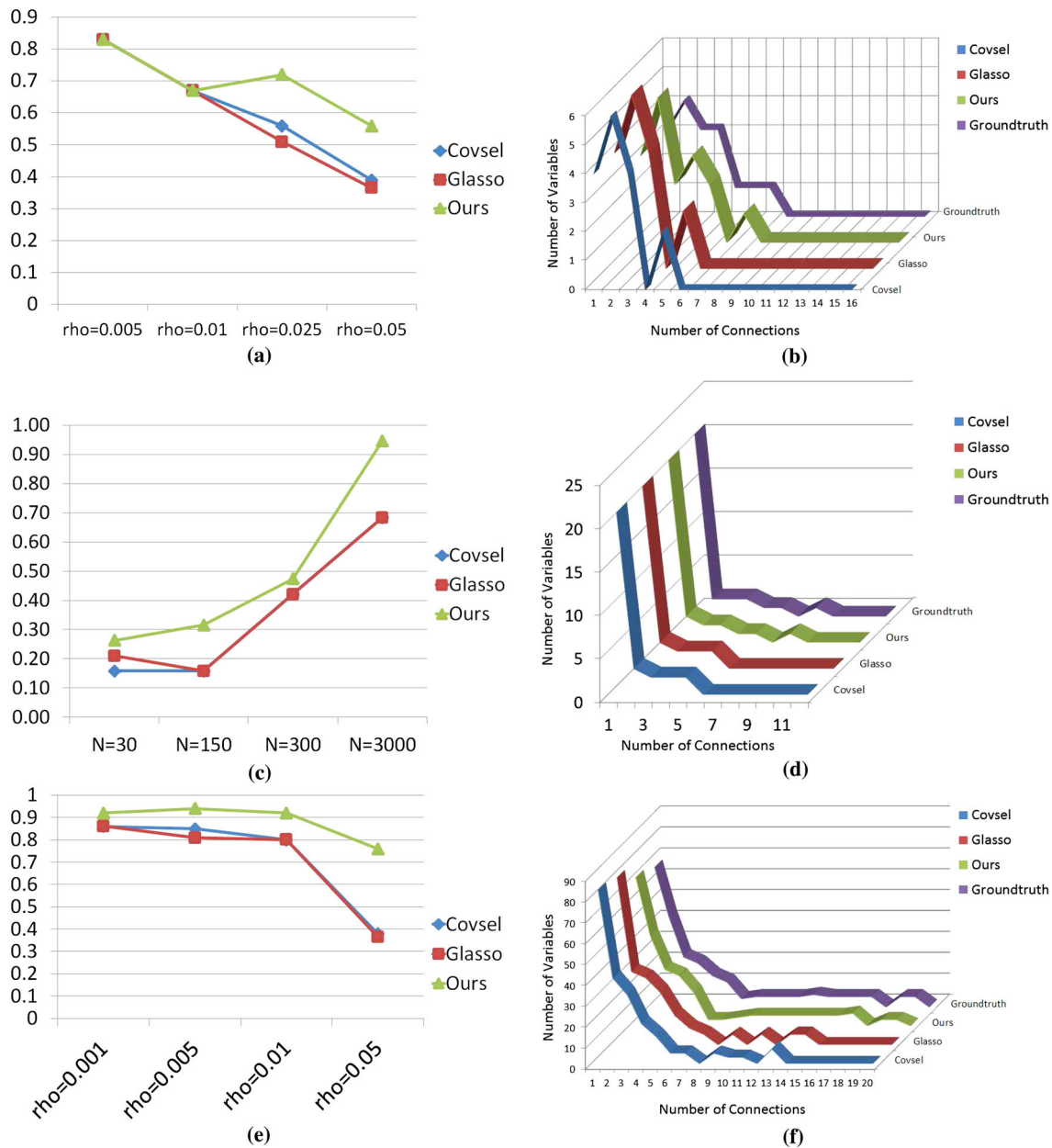
$$\text{Sensitivity}_{HUBS} = \frac{TP_{HUBS}}{TP_{HUBS} + FN_{HUBS}}, \tag{25}$$

$$MCC = \frac{TN \times TP - FN \times FP}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}, \tag{26}$$

where  $TP$ ,  $TN$ ,  $FP$ ,  $FN$  are the number of true positives, true negatives, false positives, false negatives, respectively. Note that  $TP_{HUBS}$  and  $FN_{HUBS}$  represent true positives and false negatives on hubs’ connections, respectively. The bigger sensitivity, specificity and  $MCC$  indicate better estimation.

The proposed method is compared with Covsel [16] and Glasso [6] on both synthetic and real data sets. We test the proposed and alternative methods using various penalty constants and sample size. We choose some of the comparisons in this section. Note that the results are selected when Covsel [16] and Glasso [6] methods give their highest accuracy. The accuracy and robustness of the methods are shown





**Fig. 2** **a** Sensitivity of 30 connections between 6 hubs and other variables in the scale-free network with  $p = 16$  and  $N = 1600$ . **b** Histogram of the variable connections on the data set with  $p = 16$ . **c** Sensitivity of 19 connections between 3 hubs and other variables in the scale-free

network with  $p = 30$  and  $\rho = .045$ . **d** Histogram of the variable connections on the data set with  $p = 30$ . **e** Sensitivity of the hubs in the scale-free network with  $p = 204$  and  $N = 20,400$ . **f** Histogram of the variable connections on the data set with  $p = 204$

using (i) the accuracy table, (ii) histograms analyzing connections between all nodes, (iii) sensitivity measurements on the connections between the hubs and other nodes, (iv) partial regions of the resulting networks of the real data networks.

We tested the methods on two synthetic and three real networks. Two synthetic networks are shown in Fig. 1a, b, whereas one of the real network is shown in Fig. 1c. The number of variables in these networks is  $p = 16$ ,  $p = 30$  and

$p = 204$ , respectively. The second and third real networks have “2361” and “4039” variables, respectively.

### 3.1 Synthetic data

First, we begin with two small synthetic examples to assess the method to recover the sparse and scale-free structures. In this experiment, we measure the robustness of the algorithms

**Table 1** Precision results for the data sets with (i)  $p = 16, N = 1600, \rho = 0.05$ , (ii)  $p = 30, N = 3000, \rho = 0.045$ , (iii)  $p = 204, N = 20,400, \rho = 0.05$ , (iv)  $p = 2361, N = 75,000, \rho = 0.01$

Measurements	(i) Ref. [16]		(ii) Ref. [16]		(iii) Ref. [16]		(iv) Ref. [16]		Ours	
	Ref. [6]	Ours	Ref. [6]	Ours	Ref. [6]	Ours	Ref. [6]	Ours	Ref. [6]	Ours
Sensitivity	0.82	0.91	0.89	0.98	0.87	0.98	0.79	0.94	0.93	0.98
Specificity	0.97	0.97	0.98	0.98	0.98	0.98	0.99	0.99	0.97	0.97
MCC	0.88	0.94	0.94	0.97	0.92	0.97	0.88	0.96	0.96	0.96
Time (s)	0.05	0.06	0.06	0.06	0.19	0.06	1.2	1.3	45.1	61.9
									542.3	

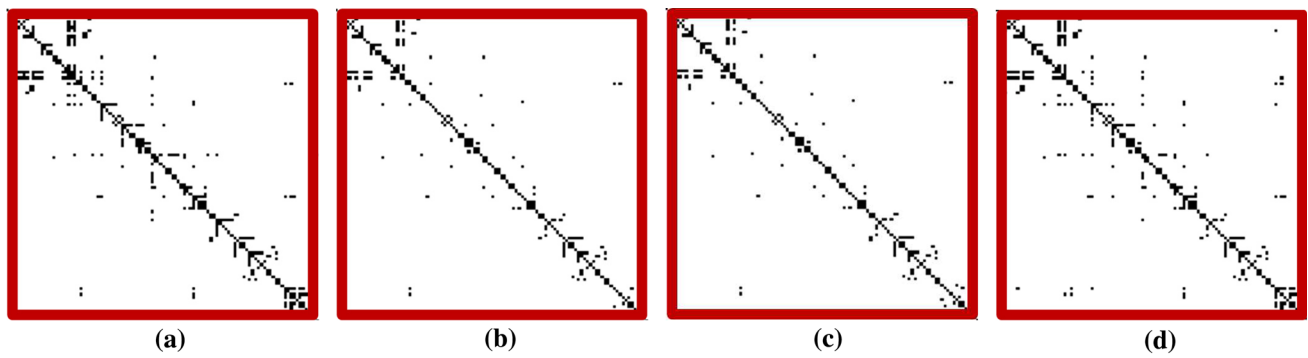
based on different penalty constants and sample size. Note that  $\rho = \rho_R$  in the experiments.

Figure 2a shows the sensitivity only on the connections between the hubs and other nodes under various penalty constants. Here, our method ensures (as the alternatives do not) that the connections on the hubs will not be lost as much as possible in the scale-free networks when we seek to have a sparse precision matrix under various penalty terms. Also, the results show that the proposed method is more robust to the changes. The number of hubs can be selected by users or the power-law distribution. The histogram of “the number of connections” with respect to “the number of variables” is shown in Fig. 2b. Here, we use an illustration to show how the proposed and other two methods estimate the connections between all variables, especially on the hubs (when the number of connections increases). Our proposed method is able to protect the hubs, whereas the sparseness of the precision matrix is warranted. Table 1(i) shows the results on this data set with a specific scenario.

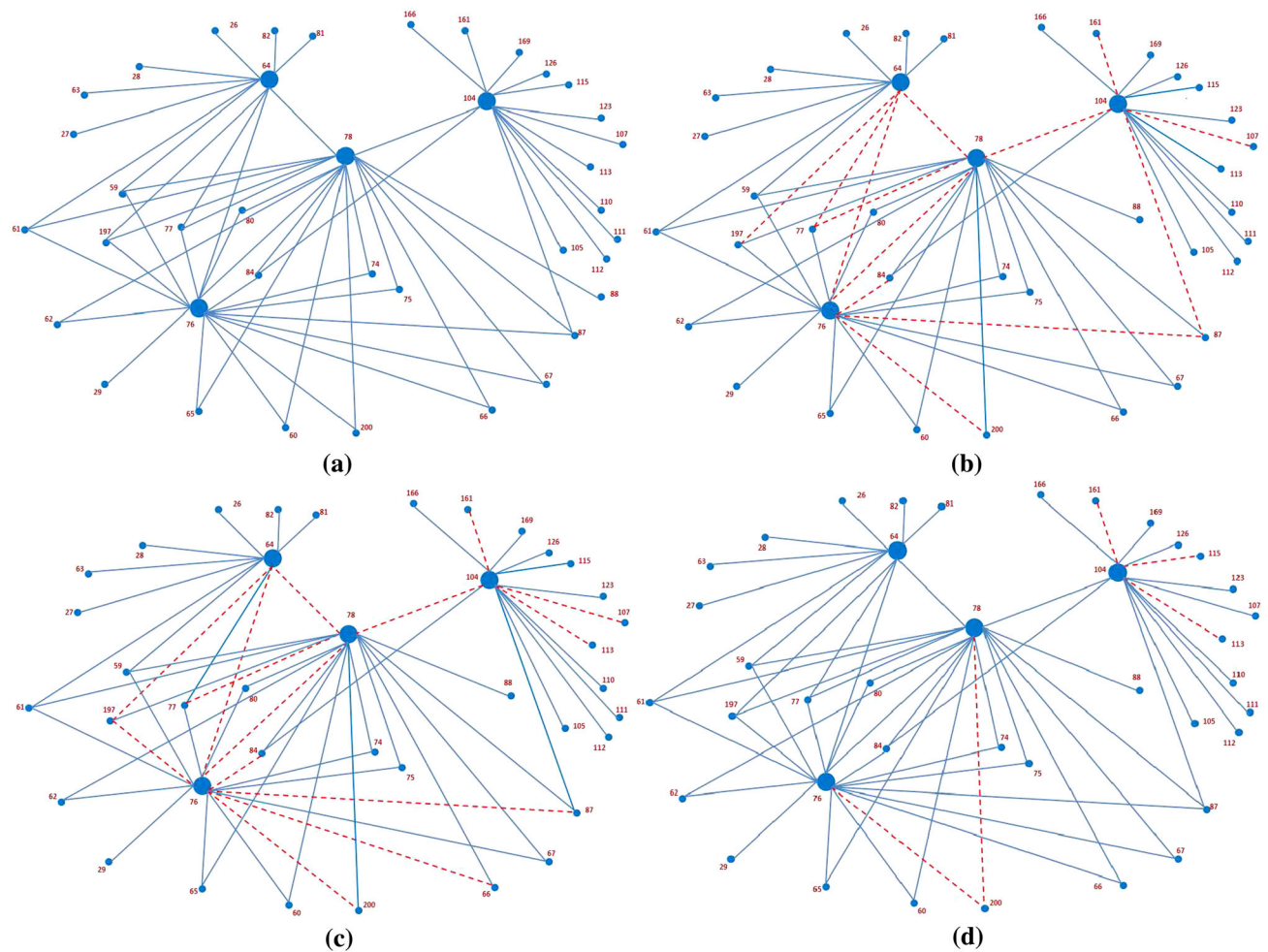
A similar experiment is applied on the network shown in Fig. 1b. Figure 2c shows the sensitivity only on the connections between the hubs and other nodes under various sample sizes. The histogram that shows “the number of connections” with respect to “the number of variables” is shown in Fig. 2d. In this figure, we can see that the numbers of hubs’ connections of our method and the ground truth are almost similar to each other. One of the most important contributions of the proposed method is to guard the hubs since scale-free networks are considered extremely vulnerable to coordinated attacks to hubs. Table 1(ii) shows the results on this data set with a specific scenario.

### 3.2 Real data

A considerably bigger scale-free network [24] with  $p = 204$  (Fig. 1c) is analyzed as one of the real networks. Based on the information given in [24], this graph displays the major component of the network which was generated by proteome data of *Saccharomyces cerevisiae*. Figure 2e shows the sensitivity only on the connections between the hubs and other nodes under various penalty constants. In this figure, the sample size is fixed to  $N = 20,400$  and the constant  $\rho$  changes. Our method is more robust under various constants  $\rho = \{0.001 - 0.05\}$ . The histogram of all connections is shown in Fig. 2f. Table 1(iii) shows the results on this data set with a specific scenario. Figure 3 shows the results of the precision matrix estimation. This figure shows that the proposed algorithm protects the hubs. Figure 4 shows the partial section of this network. In this figure, the quality assessment on 4 hubs and their connections is shown. These results are obtained when the sensitivity measurements of all 204 variables’ connections are 93, 92 and 98% for Covsel [16], Glasso [6] and the proposed method, respectively.



**Fig. 3** Results on the scale-free network with  $p = 204$ . **a** Ground truth of the precision matrix. The results of the **b** Covsel, **c** Glasso and **d** our proposed algorithm. As shown here, the proposed algorithm is capable of to guard the hubs when promoting the sparseness of the precision matrix



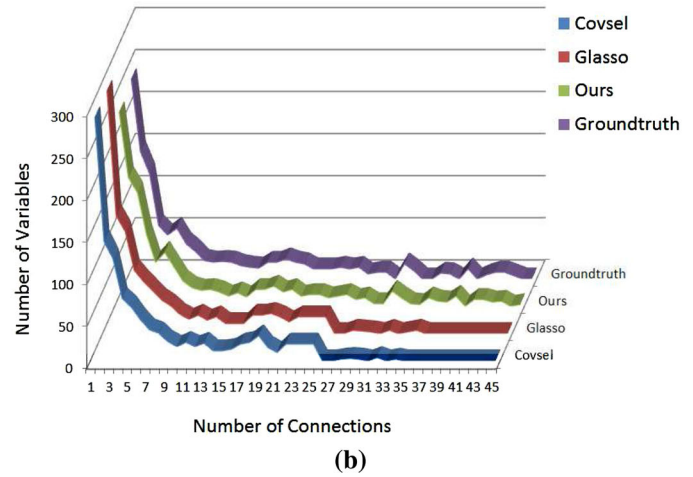
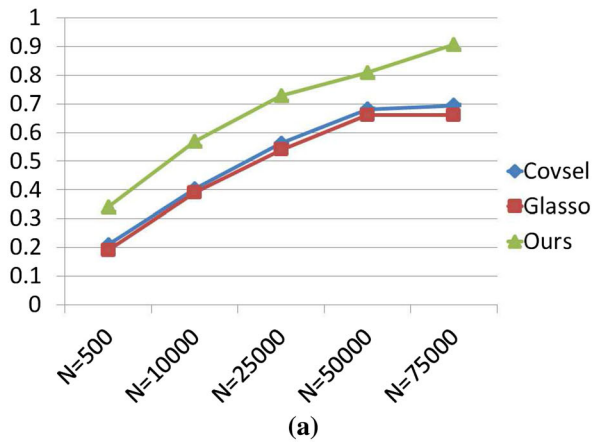
**Fig. 4** A small region of the scale-free network [24] (as shown in Fig. 1c) with  $p = 204$ . This area shows only the 4 hubs and their connections with other nodes. **a** Ground truth of the variable connections. The results of the **b** Covsel, **c** Glasso and **d** our proposed algorithm.

The *red dashed color* represents the *FNs*. These results are obtained when the sensitivity measurements of all 204 variables' connections are 91, 90 and 97% for Covsel [16], Glasso [6] and the proposed method, respectively (color figure online)

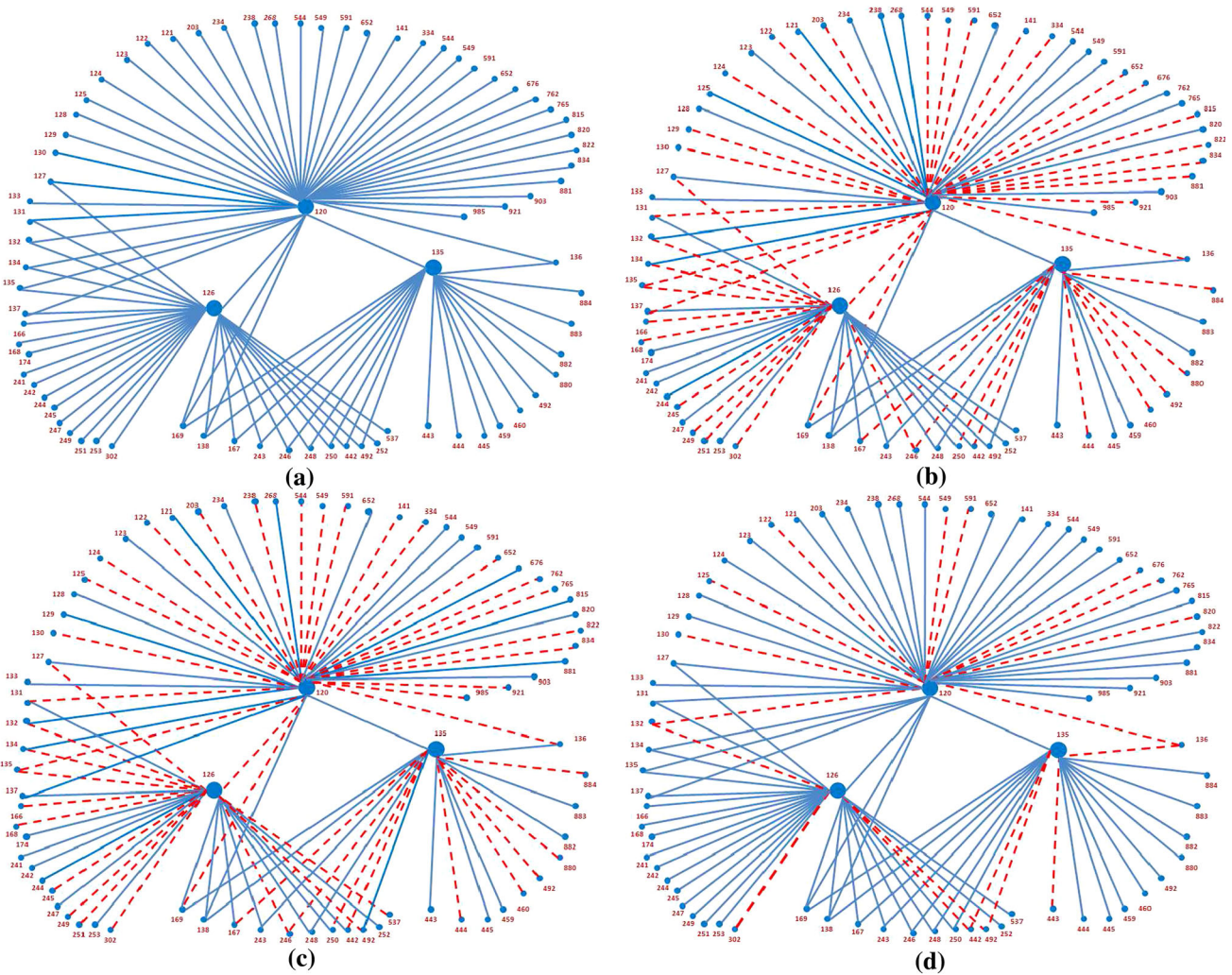
The results reveal that the loss of connections between the hubs and other nodes is decreased by our algorithm when the sparsity is needed. The specificity rates are 98.8, 98.4

and 98.1% for Covsel [16], Glasso [6] and the proposed method, respectively. As a second real data, we test the methods on a yeast data set [27,28] with  $p = 2361$ . Estimation





**Fig. 5** **a** Sensitivity of 841 connections between 54 hubs and other variables in the scale-free network with  $p = 2361$  and  $\rho = .01$ . **b** Histogram of the variable connections on this data set



**Fig. 6** This region shows only the 3 hubs and their connections with other nodes. **a** Ground truth of the variable connections. The results of the **b** Covsel, **c** Glasso and **d** our proposed algorithm. The red color represents the *FNs* (color figure online)

**Table 2** Precision results for the Facebook data sets [25] with  $p = 4039$ ,  $N = 2,019,500$ , and (i)  $\rho = 0.003$ , (ii)  $\rho = 0.008$ , (iii)  $\rho = 0.015$ 

Measurements	(i) Ref. [16]	Ref. [6]	Ours	(ii) Ref. [16]	Ref. [6]	Ours	(iii) Ref. [16]	Ref. [6]	Ours
Sensitivity	0.86	0.85	0.89	0.62	0.61	0.65	0.44	0.42	0.52
Specificity	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	0.99
MCC	0.91	0.92	0.91	0.79	0.78	0.81	0.63	0.63	0.65
Sensitivity (HUBS)	0.87	0.86	0.95	0.84	0.83	0.91	0.74	0.73	0.86
Time (s)	512.3	48.3	53.2	523.2	47.8	52.3	534.2	49.1	54.9

**Table 3** Results for the EVA data set [26] with  $p = 8343$ ,  $N = 2,085,750$ , and (i)  $\rho = 0.0025$ , (ii)  $\rho = 0.0075$ , (iii)  $\rho = 0.015$ 

Measurements	(i) Ref. [16]	Ref. [6]	Ours	(ii) Ref. [16]	Ref. [6]	Ours	(iii) Ref. [16]	Ref. [6]	Ours
Sensitivity	0.81	0.80	0.86	0.56	0.56	0.61	0.40	0.37	0.49
Specificity	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	0.98
MCC	0.87	0.88	0.90	0.76	0.75	0.79	0.58	0.55	0.62
Sensitivity (HUBS)	0.92	0.90	0.98	0.85	0.87	0.96	0.79	0.78	0.92
Time (s)	610.1	76.1	65.3	678.1	49.5	52.3	814.9	56.5	67.3

methods for the interaction detection have led to the discovery of thousands of interactions between proteins. This yeast data set consists of protein–protein interaction network described and analyzed in [27] and available as an example in the software package [28]. Figure 5a shows the sensitivity only on the connections between the hubs and other nodes under various sample sizes. Figure 5b shows the histogram of the number of connections with respect to the number of variables. Table 1(iv) shows the results on this data set with a specific scenario. Also, the execution time of each method is shown in this table. Figure 6 shows the partial section of this network. In this figure, the quality assessment on 3 hubs and their connections is shown. In this real network, we accept that three variables (“120,” “126” and “135”) are the hubs for this partial region. These results are obtained when the sensitivity measurements of all 2361 variables’ connections are 93, 92 and 98 % for Covsel [16], Glasso [6] and the proposed method, respectively. The specificity rates are 98.7, 99.7 and 99.3 % for Covsel [16], Glasso [6] and the proposed method, respectively. The false-positive ( $FP$ ) connections are not shown since there are very few edges showing wrong dependencies for three methods. The results reveal that the loss of connections between the hubs and other nodes is decreased by our algorithm when the sparsity is needed.

A larger real data set [25] is a network that analyze the connections in Facebook. One of the objectives of analyzing the social data is to discover the connections and users’ social circles. The number of subjects or variables is 4039 in these data, whereas the average connection number is 21.88 per variable. The biggest hub in this data set has 1043 connections. We evaluate the performance of the proposed and alternative methods with different scenarios as shown in Table 2. The

results show that hubs are more protected with our proposed method using various parameters, while the other measurements (i.e., sensitivity, specificity and MCC) are comparable with other alternatives.

Finally, we assess our proposed method on the Extraction, Visualization and Analysis (EVA) of corporate inter-relationships (US Corporate Ownership) data [26]. These data are constructed using the telecommunications and media industries with ownership networks with 6726 relationships among 8343 companies. The analysis of these data reveals a highly clustered network, with over 50 % of all companies connected to one another in a single one. In another words, a link  $(X, Y)$  revealing a connection from company  $X$  to company  $Y$  exists in the network if in the company  $X$  has the ownership of the company  $Y$  [29]. Also, the ownership activity is highly unbalanced such that 90 % of companies have no more than one relationship, whereas the top ten companies are parents for over 24 % of all relationships. These data can be a representative population for the current inter-relationships. The results and comparison with the alternatives are shown in Table 3. Again, in this experiment, we prove that our proposed idea is more robust, especially on the hubs when using various parameters.

## 4 Conclusions

This article has presented a new Gaussian graphical model for the scale-free and/or hub-dominated networks. We proposed a new convex log-likelihood formulation, which promotes the sparseness of the precision matrix and scale-free feature of graphical topology together. Our proposed method was assessed on two synthetic and four real networks which have

various variable numbers. As can be seen from the experiments, the adaption of the our method to the various data types such as yeast, social and business interactions is very straightforward. The proposed method is more robust under various effects and is able to guard hubs; hence, scale-free networks are protected to the possible damages to hubs. There are several ways to extend this research. For example, the choice of the penalty constant for the hubs may be studied although our proposed method is not very sensitive to the penalty constant selection based on our experiments. The method appears to be a potentially valuable technique for analyzing various large-scale databases.

**Acknowledgments** This work is supported by the US National Science Foundation award (IIA-1028098).

## References

- Aslan, M.S., Chen, X.W., Cheng, H.: Learning sparse and scale-free networks. In: Data Science and Advanced Analytics (DSAA), 2014 IEEE International Conference on 2014, pp. 326–332 (2014)
- Dempster, A.P.: Covariance selection. *Int. Biom. Sel.* **28**(1), 157–175 (1972)
- Honorio, J., Samaras, D., Rish, I., Cecchi, G.: Variable selection for Gaussian graphical models. In: AISTATS, pp. 538–546 (2012)
- Meinshausen, N., Bühlmann, P.: High dimensional graphs and variable selection with the lasso. *Ann. Stat.* **34**, 1436–1462 (2006)
- Tibshirani, R.: Regression shrinkage and selection via the lasso. *J. R. Stat. Soc.* **58**(1), 267–288 (1996)
- Friedman, J., Hastie, T., Tibshirani, R.: Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9**(3), 432–441 (2007)
- Rothman, A., Levina, P., Zhu, J.: Sparse permutation invariant covariance estimation. *Electron. J. Stat.* **2**, 494–515 (2008)
- Fan, J., Feng, Y., Wu, Y.: Network exploration via the adaptive lasso and scad penalties. *Ann. Appl. Stat.* **3**(2), 521–541 (2009)
- Banerjee, O., Ghaoui, E.L., d’Aspremont, A., Natsoulis, G.: Model selection through sparse maximum likelihood estimation. *J. Mach. Learn. Res.* **9**, 485–516 (2007)
- Guo, J., Wang, S.: A constrained  $l_1$  minimization approach to sparse precision matrix estimation. *Am. Stat. Assoc.* **106**(494), 594–607 (2011)
- Dahl, J., Vandenberghe, L., Roychowdhury, V.: Covariance selection for non-chordal graphs via chordal embedding. *Methods Softw.* **23**(4), 485–516 (2008)
- Yuan, M., Lin, Y.: Model selection and estimation in the gaussian graphical model. *Biometrika* **94**(1), 19–35 (2007)
- d’Aspremont, A., Banerjee, O., El Ghaoui, L.: First-order methods for sparse covariance selection. *SIAM J Matrix Anal. Appl.* **30**(1), 56–66 (2008)
- Lu, Z.: Adaptive first-order methods for general sparse inverse covariance selection. *SIAM J. Matrix Anal. Appl.* **31**(4), 2000–2016 (2010)
- Ravikumar, P., Wainwright, M.J., Raskutti, G., Yu, B.: High-dimensional covariance estimation by minimizing  $l_1$ -penalized log-determinant divergence. *Electron. J. Stat.* **5**, 935–980 (2011)
- 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.* **3**(1), 1–122 (2010)
- Barabási, A.L.: Scale-free networks: a decade and beyond. *Science.* **325**(5939), 412 (2009)
- Liu, Q., Ihler, A.T.: Learning scale free networks by reweighted  $l_1$  regularization. In: International Conference on Artificial Intelligence and Statistics, pp. 40–48 (2011)
- Liu, C., Du, W.B., Wang, W.X.: Particle swarm optimization with scale-free interactions. *PloS one* **9**(5), e97822 (2014)
- Zhang, C., Huang, H.: Optimal control strategy for a novel computer virus propagation model on scale-free networks. *Phys. A Stat. Mech. Appl.* **451**, 251–265 (2016)
- Liu, H., Yin, R., Liu, B., Li, Y.: A scale-free topology model with fault-tolerance and intrusion-tolerance in wireless sensor networks. *Comput. Electr. Eng* (2016)
- Glowinski, R., Marrocco, A.: Sur l’approximation, par éléments finis d’ordre un, et la résolution, par pénalisation-dualité, d’une classe de problèmes de dirichlet non linéaires. *Rev. Franca. Automat. Inf. Rech. Oper.* **9**, 4176 (1975)
- Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximations. *Comput. Math. Appl.* **2**, 1740 (1976)
- Wuchty, S.: Scale-free behavior in protein domain networks. *Mol. Biol. Evol* **18**, 1694–1702 (2001)
- McAuley, J.J., Jure Leskovec.: Learning to discover social circles in ego networks. In: NIPS, vol. 2012, pp. 548–56 (2012)
- Norlen, K., Lucas, G., Gebbie, M., Chuang, J.: EVA: Extraction, visualization and analysis of the telecommunications and media ownership network. In: Proceedings of International Telecommunications Society 14th Biennial Conference (ITS2002), Seoul Korea (2002)
- Sun, S., Ling, L., Zhang, N., Li, G., Chen, R.: Topological structure analysis of the protein–protein interaction network in budding yeast. *Nucleic Acids Res.* **31**(9), 2443–2450 (2003)
- Software package protein interaction network (pin) at <http://www.bioinfo.org.cn/PIN/>
- <http://vlado.fmf.uni-lj.si/>