

WEIGHTED LASSO IN GRAPHICAL GAUSSIAN MODELING FOR LARGE GENE NETWORK ESTIMATION BASED ON MICROARRAY DATA

TEPPEI SHIMAMURA SEIYA IMOTO
shima@ims.u-tokyo.ac.jp imoto@ims.u-tokyo.ac.jp
RUI YAMAGUCHI SATORU MIYANO
ruiy@ims.u-tokyo.ac.jp miyano@ims.u-tokyo.ac.jp

*Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1
Shirokanedai, Minato-ku, Tokyo 108-8639, Japan*

We propose a statistical method based on graphical Gaussian models for estimating large gene networks from DNA microarray data. In estimating large gene networks, the number of genes is larger than the number of samples, we need to consider some restrictions for model building. We propose weighted lasso estimation for the graphical Gaussian models as a model of large gene networks. In the proposed method, the structural learning for gene networks is equivalent to the selection of the regularization parameters included in the weighted lasso estimation. We investigate this problem from a Bayes approach and derive an empirical Bayesian information criterion for choosing them. Unlike Bayesian network approach, our method can find the optimal network structure and does not require to use heuristic structural learning algorithm. We conduct Monte Carlo simulation to show the effectiveness of the proposed method. We also analyze *Arabidopsis thaliana* microarray data and estimate gene networks.

Keywords: Lasso; L_1 -Regularization; Empirical Bayes; Graphical Modeling; Large Gene Network; DNA Microarray Data

1. Introduction

The interest in theoretical structure and constructive methodology for large-scale graphical models has been invigorated by problems of inferring gene networks from microarray gene expression data. Various reverse engineering methods for large gene networks have been proposed in the literature, and graphical Gaussian model has received a lot of attention recently as one of promising approaches for achieving the purpose [4, 14, 16, 19].

Several studies have used regression-based approaches to estimate structures of undirected graphs. The main idea is to apply regression analysis for each variable on the rest of the variables and the regression coefficients with relatively large absolute values correspond to edges in the graph. Meinshausen and Bühlmann [14] considered neighborhood selection as a subproblem of covariance selection and applied least absolute shrinkage and selection operator, called lasso [17], to estimate undirected

2 *T. Shimamura, S. Imoto, R. Yamaguchi & S. Miyano*

graph structures and showed theoretical properties in high-dimensional situations, and Gustafsson *et al.* [6] also used the similar approach to estimate large gene networks from DNA microarray data.

However, recent works suggest that the lasso is not consistent for variable selection [11, 14, 21, 22]. That is, the sets of variables selected are not consistent at finding the true set of important variables. Meinshausen and Bühlmann [14] showed that the optimal regularization parameter for prediction accuracy does not lead to a consistent neighborhood estimate by giving an example in which the number of variables grows to infinity and in which the regularization parameter chosen according to prediction accuracy leads to the wrong model with probability tending to one. As the result, some noisy and irrelevant neighborhoods tend to be selected and the resulting graph is less sparse even if we use the lasso in the neighborhood selection problem.

We propose a new method for improving neighborhood selection in the graphical Gaussian model and define weighted lasso that allows different amount of penalty on each regression coefficient. We show that, by using the ridge estimator, the originally $(p - 1)$ regularization parameter selection problem of each node j reduces to a univariate regularization parameter selection problem for λ_j . For choosing an optimal graph structure, we derived a new information criterion from an empirical Bayes viewpoint.

The rest of this paper is as follows. In Section 2, we describe our method. Especially, we review a class of graphical Gaussian graphical model and neighborhood selection for estimating sparse graph structure, and define a new neighborhood selection approach, in Section 2.1 and 2.2, respectively. In Section 2.3, we also derive a new criterion for choosing an optimal neighborhood for each node. In Section 3, we present the performance of the proposed method through numerical examples. In Section 3.1, We compare the performance of the proposed method with that of conventional graphical modeling approach through Monte Carlo simulation. We also apply our approach to *Arabidopsis thaliana* microarray data and estimate gene networks in Section 3.2. We provide concluding remarks in Section 4.

2. Method for Estimating Large Gene Networks

2.1. Graphical Gaussian Model and Neighborhood Selection

In this section, we present a class of graphical Gaussian model and review the neighborhood selection problem introduced by Meinshausen and Bühlman [14] that is a subproblem of covariance selection. A graphical Gaussian model or covariance selection model for the Gaussian random vector X is represented by an undirected graph $G = (V, E)$ with vertex set $V = \{1, \dots, p\}$ and edge set $E = \{e^{ij}\}$, where $e^{ij} = 1$ or 0 according to whether vertices i and j are adjacent in G or not.

Let X be a p -dimensional random vector which consists of the p elements, $X^{(1)}, \dots, X^{(p)}$. We assume that $X \sim N_p(\mathbf{0}, \Sigma)$ where Σ is a variance-covariance matrix. Our interest is to estimate the underlying graph of the distribution, that is

to identify zero entries of the concentration matrix $\mathbf{C} = \boldsymbol{\Sigma}^{-1} = (c^{ij})$, because \mathbf{C} satisfies the following linear restrictions:

$$e^{ij} = 0 \Rightarrow c^{ij} = 0. \quad (1)$$

The graphical Gaussian model also can be defined in terms of the pairwise conditional independence determined by the Markov properties of G . If $X \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

$$c^{ij} = 0 \Leftrightarrow X^{(i)} \perp\!\!\!\perp X^{(j)} \mid X^{V \setminus \{i,j\}} \Leftrightarrow \rho^{ij} = 0, \quad (2)$$

where

$$\rho^{ij} = \frac{-c^{ij}}{\sqrt{c^{ii}c^{jj}}}, \quad (3)$$

denotes the ij -th partial correlation, that is the correlation between the i -th variable and the j -th variable, say $X^{(i)}$ and $X^{(j)}$, conditional on the rest of the variables $X^{V \setminus \{i,j\}}$.

Given n random samples, $\mathbf{x}_1 \dots \mathbf{x}_n$, of X where $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$, the log-likelihood for $\mathbf{C} = \boldsymbol{\Sigma}^{-1}$ is

$$\frac{n}{2} \log |\mathbf{C}| - \frac{1}{2} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})' \mathbf{C} (\mathbf{x}_i - \boldsymbol{\mu}), \quad (4)$$

up to a constant not depending on $\boldsymbol{\mu}$ and \mathbf{C} . The unbiased estimator for the covariance matrix $\boldsymbol{\Sigma}$ is

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})', \quad (5)$$

where $\bar{\mathbf{x}}$ is the sample mean. Then, the concentration matrix \mathbf{C} can be estimated by $\hat{\boldsymbol{\Sigma}}^{-1}$. In general, the maximum likelihood estimator $\hat{\mathbf{C}} = \hat{\boldsymbol{\Sigma}}^{-1}$ contains no zero entry, and the network corresponding to $\hat{\mathbf{C}}$ is thus the complete graph. However, the structure of the gene networks is known as a 'sparse' structure.

To achieve 'sparse' structure, edge selection is usually carried out by the following steps.

- (1) Compute the sample concentration matrix $\hat{\mathbf{C}}$ via $\hat{\boldsymbol{\Sigma}}^{-1}$.
- (2) Compute the sample estimates of partial correlation ρ^{ij} via equation (3).
- (3) Determine which ρ^{ij} is 0 or not.

However, the sample covariance matrix $\hat{\boldsymbol{\Sigma}}$ is singular in large gene network estimation from microarray data, and we cannot compute the partial correlation coefficients directly. To solve this problem, Meinshausen and Bühlmann [14] proposed the neighborhood selection method.

The neighborhood selection problem is introduced as a subproblem of covariance selection. Let the neighborhood ne_a of a node a be the smallest subset of $V \setminus \{a\}$ so that, given all variables X^{ne_a} in the neighborhood, $X^{(a)}$ is conditionally independent

4 *T. Shimamura, S. Imoto, R. Yamaguchi & S. Miyano*

of all remaining variables. Denote the closure of node $a \in V$ by $\text{cl}_a = \text{ne}_a \cup \{a\}$. Then

$$X^{(a)} \perp\!\!\!\perp \{X^{(k)}; k \in V \setminus \text{cl}_a\} | X^{\text{ne}_a}. \quad (6)$$

Given $n \times p$ data matrix of X , $\mathbf{X} = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(p)})$ where $\mathbf{x}^{(j)} = (x_1^{(j)}, \dots, x_n^{(j)})'$ is the n -dimensional sample vector of node j , the purpose of neighborhood selection is to estimate the neighborhood of any given node individually. We assume that the each column $\mathbf{x}^{(j)}$, $j = 1, \dots, p$ is centered and standardized such that $\sum_i x_i^{(j)}/n = 0$ and $\sum_i \{x_i^{(j)}\}^2/n = 1$. Meinshausen and Bühlmann [14] cast the neighborhood selection problem into a standard linear regression problem. The idea is to apply regression analysis for each variable on the rest of variables and the regression coefficients with nonzero values correspond to edges in the graph. They also show that this problem can be solved efficiently with the lasso [17]. For each node $j = 1, \dots, p$ on the rest of nodes, the lasso estimator $\hat{\boldsymbol{\theta}}^{[-j]} = (\hat{\theta}_j^{(1)}, \dots, \hat{\theta}_j^{(j-1)}, \hat{\theta}_j^{(j+1)}, \dots, \hat{\theta}_j^{(p)})'$ is given by

$$\hat{\boldsymbol{\theta}}^{[-j]} = \arg \min_{\boldsymbol{\theta}^{[-j]}} \left\{ \frac{1}{n} \|\mathbf{x}^{(j)} - \mathbf{X}^{[-j]} \boldsymbol{\theta}^{[-j]}\|^2 + \lambda \sum_{k \neq j} |\theta_j^{(k)}| \right\}, \quad (7)$$

where $\mathbf{X}^{[-j]}$ is the $n \times (p-1)$ matrix resulting from the deletion of the j -th column from the data matrix \mathbf{X} and λ is a non-negative regularization parameter. Then, the neighborhood estimate of node j is defined by

$$\hat{\text{ne}}_j = \{k \in V; \hat{\theta}_j^{(k)} \neq 0\}. \quad (8)$$

Note that it is possible that $\theta_j^{(k)} = 0$ but $\theta_k^{(j)} \neq 0$, or $\theta_k^{(j)} = 0$ but $\theta_j^{(k)} \neq 0$. They suggest that the nodes j and k are connected by the undirected edge if $\hat{\theta}_j^{(k)} \neq 0$ or $\hat{\theta}_k^{(j)} \neq 0$.

2.2. Improving Neighborhood Selection with Weighted Lasso

In this section, we show a solution to overcome the drawback in the neighborhood selection of the graphical Gaussian model with the lasso. The major reason for the inconsistency of variable selection with the lasso comes from the fact that the lasso imposes the same amount of penalty on each regression coefficient, regardless of their relative significance [5, 22]. As a simple approach for allowing different amount of penalty on each regression coefficient, we consider the following problem for node j , $j = 1, \dots, p$:

$$\hat{\boldsymbol{\theta}}^{[-j]} = \arg \min_{\boldsymbol{\theta}^{[-j]}} \left\{ \frac{1}{n} \|\mathbf{x}^{(j)} - \mathbf{X}^{[-j]} \boldsymbol{\theta}^{[-j]}\|^2 + \sum_{j \neq k} \lambda_j^{(k)} |\theta_j^{(k)}| \right\}. \quad (9)$$

Here $\lambda_j^{(k)}$, $k = 1, \dots, (j-1), (j+1), \dots, p$, are $(p-1)$ non-negative regularization parameters. Obviously, this problem depends on $(p-1)$ regularization parameters.

In practice, the choices of these regularization parameters are time-consuming, since we have to choose $(p-1)$ regularization parameters for each node j , which leads to the $p \times (p-1)$ regularization parameter selection problem for all nodes. To solve this problem, we propose to replace $\lambda_j^{(k)}$ by

$$\lambda_j^{(k)} = \frac{\lambda_j}{(\tilde{\theta}_j^{(k)})^\tau}, \quad (10)$$

where τ is a constant and $\tilde{\theta}_j^{(k)}$ is a weight for the significance of the k -th regression coefficient $\theta_j^{(k)}$. Let a $(p-1)$ -dimensional weight vector by $\tilde{\boldsymbol{\theta}}_j = (\tilde{\theta}_j^{(1)}, \dots, \tilde{\theta}_j^{(j-1)}, \tilde{\theta}_j^{(j+1)}, \dots, \tilde{\theta}_j^{(p)})'$. We estimate $\tilde{\boldsymbol{\theta}}_j$ by

$$\tilde{\boldsymbol{\theta}}_j = (\mathbf{X}^{[-j]'} \mathbf{X}^{[-j]} + \gamma \mathbf{I}_{p-1})^{-1} \mathbf{X}^{[-j]} \mathbf{x}^{(j)}, \quad (11)$$

where γ is a pre-specified regularization parameter, which is equivalent to the ridge estimate. Even if the data matrix $\mathbf{X}^{[-j]}$ is rank-deficient, so that $\mathbf{X}^{[-j]'} \mathbf{X}^{[-j]}$ is singular, the regularized matrix $(\mathbf{X}^{[-j]'} \mathbf{X}^{[-j]} + \gamma \mathbf{I}_{p-1})$ is not singular for any nonzero value of γ . The regularization parameter γ also controls the number of neighborhoods. Although γ can be selected by some criteria such as cross-validation or our proposed criterion discussed in next section, we set γ a large value such as 10^{10} . Then, neighborhood selection proceeds by taking a node j to be a neighbor of node k if and only if $\tilde{\theta}_j^{(k)} \neq 0$. In inferring network structure, the nodes j and k are connected by the undirected edge if $\theta_j^{(k)} \neq 0$ or $\theta_k^{(j)} \neq 0$. We call the successive procedure *weighted lasso*.

2.3. A New Criterion for Regularization Parameter Selection

The regularization parameters $\lambda_1, \dots, \lambda_p$ for all nodes in (10) should be chosen in a reasonable manner. We investigate this problem from a statistical model evaluation point of view, and derive a new information criterion for neighborhood selection for each node j . We can choose the optimal neighborhoods for each node by using the derived information criterion.

It is known that the most of regularization approach have Bayesian interpretation, such that the loss function is interpreted as the negative log-likelihood, the penalty term as the negative log-prior density, the regularization parameter as the hyperparameter, and the regularized estimate corresponds to the maximum posteriori estimate. For node j , the loss function can be interpreted as the negative log-likelihood of the following Gaussian distribution

$$f(\mathbf{x}^{(j)} | \mathbf{x}^{[-j]}; \boldsymbol{\theta}^{[-j]}, \sigma_j^2) = \left(\frac{1}{2\pi\sigma_j^2} \right)^{n/2} \exp \left\{ -\frac{1}{2\sigma_j^2} \|\mathbf{x}^{(j)} - \mathbf{X}^{[-j]} \boldsymbol{\theta}^{[-j]}\|^2 \right\}, \quad (12)$$

and the prior density as the independent Laplace prior as

$$\pi(\boldsymbol{\theta}^{[-j]} | \boldsymbol{\lambda}^{[-j]}, \sigma_j^2) = \prod_{k \neq j} \pi(\theta_j^{(k)} | \lambda_j^{(k)}, \sigma_j^2) = \prod_{k \neq j} \left(\frac{\lambda_j^{(k)}}{2\sigma_j^2} \right) \exp \left\{ -\frac{\lambda_j^{(k)} |\theta_j^{(k)}|}{\sigma_j^2} \right\}. \quad (13)$$

6 *T. Shimamura, S. Imoto, R. Yamaguchi & S. Miyano*

From an empirical Bayesian point of view, one can choose $\lambda_j^{(k)}$ by maximizing the marginal likelihood [1, 8, 9]. For node j , the marginal likelihood is computed by integrating over the unknown parameter values $\boldsymbol{\theta}^{[-j]}$ and is defined by

$$\text{ML} = \int f(\mathbf{x}^{(j)} | \mathbf{x}^{[-j]}; \boldsymbol{\theta}^{[-j]}, \sigma_j^2) \pi(\boldsymbol{\theta}^{[-j]} | \boldsymbol{\lambda}^{[-j]}, \sigma_j^2) d\boldsymbol{\theta}^{[-j]}. \quad (14)$$

The marginal likelihood often contains a complicated integral for the parameters, which can be usually approximated by some approximation methods, for example, Laplace's method [18]. However, in situations where the some components of $\hat{\boldsymbol{\theta}}^{[-j]}$ are exactly to zero with L_1 -regularization approaches such as the lasso, the functional in the integral (14) is not differentiable at the origin and thus the above approximation cannot be applied directly.

Let $A_j = \{k; \hat{\theta}_j^{(k)} \neq 0\}$ be *active set* of $\hat{\boldsymbol{\theta}}^{[-j]}$. To overcome such a problem, we define the following partial marginal likelihood given active set A_j by

$$\text{PML} = \int f(\mathbf{x}^{(j)} | \mathbf{x}^{[-j]}; \boldsymbol{\theta}^{[-j]}, \sigma_j^2) \pi(\boldsymbol{\theta}^{[-j]} | \boldsymbol{\lambda}^{[-j]}, \sigma_j^2) d\boldsymbol{\theta}^{A_j}. \quad (15)$$

This quantity is computed by integrating over the unknown parameters $\boldsymbol{\theta}^{A_j}$ included with the set A_j . Suppose that $\text{PML} = O(n)$ and λ depends on the number of sample n . Applying the Laplace method, we have

$$\text{PML} \approx \left(\frac{2\pi}{n}\right)^{|A_j|/2} |\mathbf{H}|^{-1/2} \times f(\mathbf{x}^{(j)} | \mathbf{x}^{[-j]}; \hat{\boldsymbol{\theta}}^{A_j}, \sigma_j^2) \times \pi(\hat{\boldsymbol{\theta}}^{A_j} | \boldsymbol{\lambda}^{[-j]}, \sigma_j^2), \quad (16)$$

with $O(n^{-1})$ where $\hat{\boldsymbol{\theta}}^{A_j} = (\hat{\theta}_j^{(k)})_{k \in A_j}$ and \mathbf{H} is a Hessian matrix given by

$$\mathbf{H} = -\frac{1}{n} \frac{\partial^2}{\partial \boldsymbol{\theta}^{A_j} \partial \boldsymbol{\theta}^{A_j'}} \left[\log \left\{ f(\mathbf{x}^{(j)} | \mathbf{X}^{[-j]}; \boldsymbol{\theta}^{[-j]}, \sigma_j^2) \times \pi(\boldsymbol{\theta}^{[-j]} | \boldsymbol{\lambda}^{[-j]}, \sigma_j^2) \right\} \right]_{\boldsymbol{\theta}^{[-j]} = \hat{\boldsymbol{\theta}}^{[-j]}}.$$

Denote the right part of equation (16) by APML. Then, the estimator of $\hat{\sigma}_j^2$ is given by the equation

$$\frac{\partial}{\partial \sigma_j^2} \log(\text{APML}) = 0. \quad (17)$$

Approximating $-2 \log \text{PML}$ and substituting $\hat{\sigma}_j^2$ for σ_j^2 , we define an empirical Bayes criterion for neighborhood selection, called neighborhood empirical Bayes criterion (NEBC), as

$$\begin{aligned} \text{NEBC} &= (n + |A_j|) \{ \log(2\hat{\sigma}^2) + 1 \} - (n - |A_j|) \log \pi \\ &\quad - 2 \sum_{k \in A_j} \log \lambda_j^{(k)} + \log |\mathbf{X}^{A_j'} \mathbf{X}^{A_j}|, \end{aligned} \quad (18)$$

where

$$\hat{\sigma}_j^2 = \frac{\|\mathbf{x}_j - \mathbf{X}^{[-j]} \hat{\boldsymbol{\theta}}^{[-j]}\|^2 + 2\lambda_j \sum_{k \in A_j} |\hat{\theta}_j^{(k)}|}{n + |A_j|}. \quad (19)$$

For neighborhood selection problem for node j , we can choose λ_j by minimizing NEBC in (18) and estimate neighborhood \hat{n}_j with $\hat{\boldsymbol{\theta}}^{[-j]}$.

3. Numerical Examples

3.1. Monte Carlo Simulation

We perform simulations to evaluate the performance of the proposed method and compare our method with the graphical Gaussian model approach with the lasso by Meinshausen and Bühlmann [14]. Considering finds on the topology of metabolic and protein networks [7, 13], we simulate a scale-free-like graphical model and generate data based on the model by the following steps.

- (1) Generate a set of 20 scale-free networks by Barabási and Albert [3]. For each scale-free network, we start with a single node and no edges in the first time step. In each time step, one node is added and the new vertex initiates some edges to old nodes. The probability of node l , say Prob_l , that an old node is chosen is given by $\text{Prob}_l \propto k_l^\alpha$ where k_l is the in-degree of node l in the current time step, *i.e.*, the number of adjacent edges of node l which were not initiated by l itself. We set $k_l = 1$ and $\alpha = 1$. The node size of each scale-free network is set to 50. Then the total-node size of the simulated network is $20 \times 50 = 1000$.
- (2) Generate a random variable from the resulting scale-free networks. As an initial value of the parent nodes of the 20 scale free-networks, we sample from $N_{20}(\mathbf{0}, \mathbf{\Sigma})$ where $\Sigma_{ii} = 1$ and $\Sigma_{i(i+1)} = \Sigma_{(i+1)i} = 0.5$ if $\{i; i = 1, \dots, 19 \cap i \neq 4, 8, 12, 16\}$. From the j -th scale free network, we generate a random sample x based on

$$x \sim N\left(\sum_l w_l p_l(x), \sigma^2\right), \quad (20)$$

where $p_l(x)$ is the observation of the l -th parent of x and w_l is the coefficient. We sample w_l from the uniform distribution over the interval $[-1, 1]$ and set σ so that the signal-to-noise ratio is 0.1.

We simulate 100 observations with 1000 variables from the above process. The true graph and the estimated graph by the proposed method through the simulation are shown in Figure 1.

The performance of the graphical modeling approaches is evaluated by counting true positives (TP; correctly identified true edges), false positives (FP; spurious edges, that is not recognized as zero-edges), true negatives (TN; correctly identified zero-edges) and false negatives (FN; spurious zero-edges, that is not recognized as true edges). We perform 100 Monte Carlo simulations and calculate the means and standard deviations of TP, FP, TN and FN over these simulations. The result of comparison with the proposed method and the lasso-based neighborhood selection approach by Meinshausen and Bühlmann [14] is shown in Table 1. It can be seen from Table 1 that the neighborhood selection approach with the lasso [14] produces the larger number of false positives and identifies many unrelated edges as important edges. The proposed method reduces the false positives dramatically with a reasonable degree of true positives and succeeds in achieving the high proportion of true-edges within the small number of selected edges.

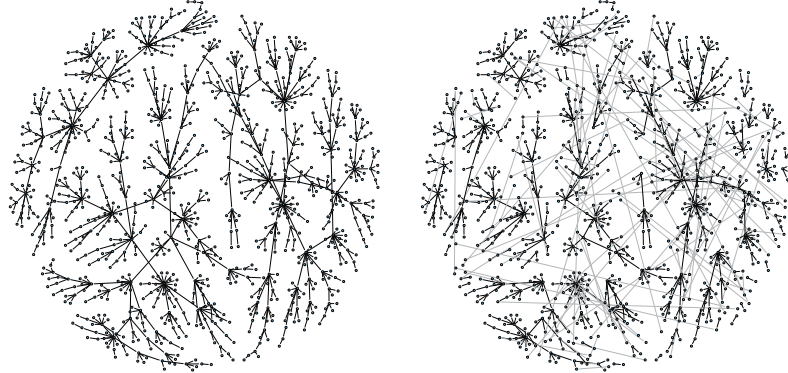


Fig. 1. The left figure is the true graph generated for the Monte Carlo simulations. This network consists of 1000 nodes and 995 edges. The right figure is an example of the estimated network by the proposed method. The black lines describe the true-positive edges and the gray lines the false-positive edges.

Table 1. Comparison of the averages and standard deviations of the true positives (TP), false negatives (FN), true negatives (TN) and false positives (FP) with the lasso-based neighborhood selection [14] and the proposed approach over the 100 simulations. The standard deviations are in parentheses.

Method	TP	FN	TN	FP
Lasso-based Method [14]	909.47 (6.01)	85.53 (6.01)	966630.22 (459.04)	32374.78 (459.04)
Proposed Approach	818.74 (6.29)	176.26 (6.29)	998894.65 (5.87)	110.35 (5.89)

3.2. Example from Experimental Data

We analyze the isoprenoid biosynthetic pathway data in *Arabidopsis thaliana* discussed by Wille *et al.* [20]. Wille *et al.* [20] reported a data set including the gene expression patterns monitored under various experimental conditions using 118 GeneChip microarrays.

It is known that plants contain two pathways, the mevalonate (MVA) pathway and the methylerythritol 4-phosphate (MEP) pathway, for the synthesis of the structural precursors of isoprenoids. To gain insights into the cross-talk between the MVA and MEP pathways at the transcriptional level and construct the gene network, Wille *et al.* [20] focuses on 39 genes where 15 genes were assigned to the cytosolic MVA pathway, 19 to the plastidal MEP pathway, and 5 genes encoding proteins located in the mitochondria.

We first estimate the undirected graph based on the 118 observations of these 39 gene expression profiles by the weighted lasso. Using NEBC in (18) for neighborhood selection, the proposed approach selects 26 gene-gene interactions. These 26 pairs are shown on the true network in Figure 2. We find there is a module with strongly connected genes in each of MVA pathway and MEP pathway. In the MVA pathway,

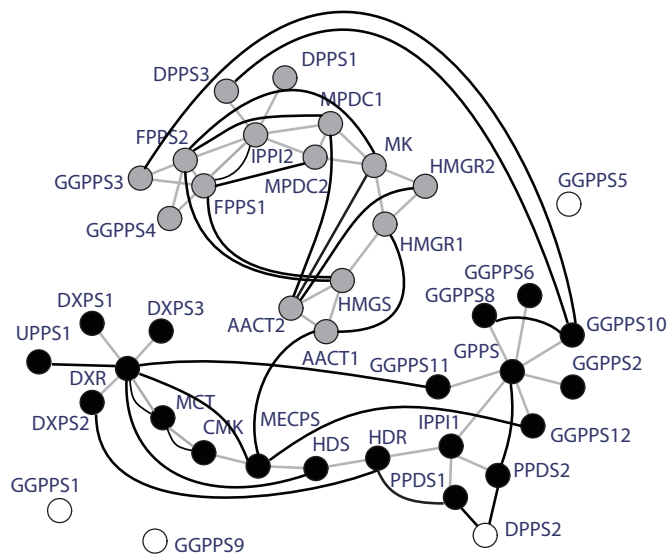


Fig. 2. Pathways identified by the proposed method for the 39 genes in the isoprenoid pathway, where the grey color edges are the true pathways, and black curved edges are the estimated edges by the proposed method. The black nodes also represent a subgraph of the gene module in the MEP pathway, the gray nodes in the MVA pathway, and the white nodes in the Mitochondrion.

DXR, MCT, CMK and MECPS are connected as the known isoprenoid pathway. In the MEP pathway, AACT2, HMGR2, MK, MPDC1, FPPS1 and FPP2 are closely connected. Furthermore, several genes in the MEP pathway are linked to proteins in the mitochondria.

Wille *et al.* [20] also incorporated 795 additional gene expressions from 56 metabolic pathways and investigated which pathways attach significantly well to the MVA and MEP pathways. Among these were genes from pathways downstream of the two isoprenoid biosynthesis pathways, such as phytosterol biosynthesis, mono- and diterpene metabolism, porphyrin/chlorophyll metabolism, carotenoid biosynthesis, plastoquinone biosynthesis, for example. We applied our approach to the 118 observations of these total 835 gene from the 56 metabolic pathways, Mitochondrion, and the MVA and MEP pathways. Then we count the number of pathway-pathway interactions from the estimated graph and generate a “metabolic-pathway relevant network” shown in Figure 4. In the metabolic-pathway relevant network, nodes represent pathways, and two pathways are connected to each other if and only if they share at least 30 interactions.

We find that there are strong connections between the MEP pathway and the 8 other metabolic pathways; calvin cycle, carotenoid, citrate cycle, fatty acid, glycolysis, inositol phosphate, porphyrin chlorophyll, and sucrose pathways. On the other hand, the fatty acid, glycolysis, phytosterol, and sucrose pathways appear to be closely related to the genes of the MVA pathway. By comparing with the results

10 *T. Shimamura, S. Imoto, R. Yamaguchi & S. Miyano*

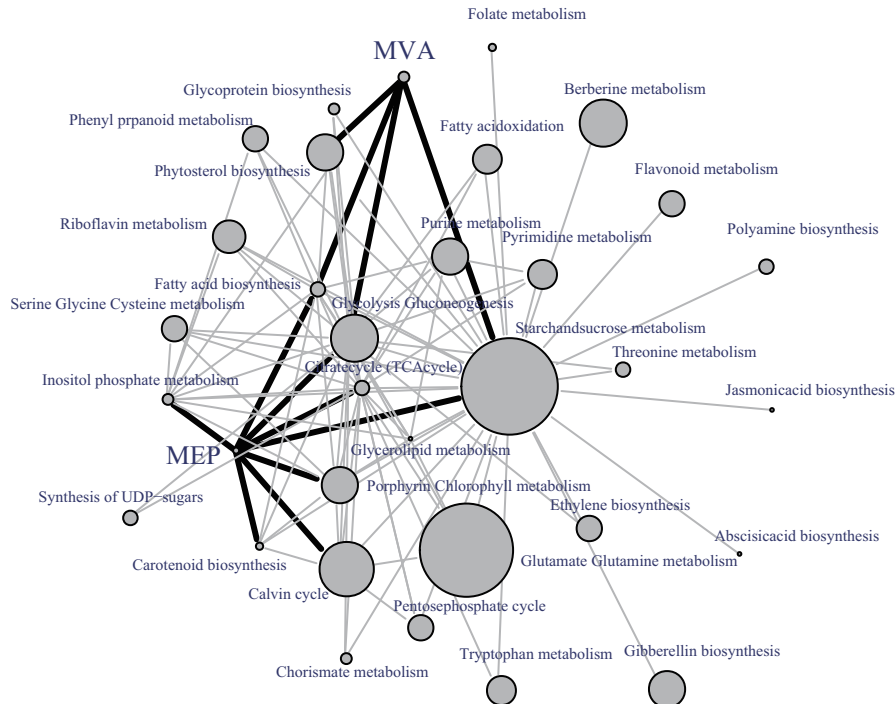


Fig. 3. The metabolic-pathway relevant network identified by the proposed method on 118 microarray data with 835 genes. The network contains the 56 metabolic pathways, mitochondrion, and the MVA and MEP pathways. Nodes represent pathways, and two pathways are connected to each other if and only if they share at least 30 interactions. The diameter of a circle is proportional to the number of genes participating in the corresponding pathways. The black edges represent the edges associated with the MVA and MEP pathways.

in Wille *et al.* [20], some of them are experimentally supported by [2, 10, 12, 15]. It is interesting that our method identifies the fatty acid pathway as a significant one related with both of the MVA and MEP pathways which was not recognized by Wille *et al.* [20].

4. Conclusion

We proposed the new method for improving neighborhood selection in the graphical Gaussian model by the weighted lasso that allows different amount of penalty on each regression coefficient. We showed that, by using the ridge estimator, the originally $(p - 1)$ regularization parameter selection problem of each node j reduces to a univariate regularization parameter selection problem for λ_j . For choosing an optimal graph structure, we derived a new criterion from an empirical Bayes viewpoint.

We conducted Monte Carlo simulation and showed examples from microarray data in *Arabidopsis thaliana* to compare the proposed method with the other graphical Gaussian modeling approach with the lasso [14]. We found that our proposed method was superior to it.

References

- [1] Akaike, H., Likelihood and the Bayes procedure, in J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith (eds.), *Bayesian Statist.*, 143–166, University Press, Valencia, Spain, 1980.
- [2] Arigoni, D., Sagner, S., Latzel, C., Eisenreich, W., Bacher, A. and Zenk, M.H., Terpenoid biosynthesis from I-deoxy-D-xylulose in higher plants by intramolecular skeletal rearrangement, *Proc. Natl. Acad. Sci. USA*, 94:10600–10605, 1997.
- [3] Barabási, A. and Albert, R., Emergence of scaling in random networks, *Science*, 286:509–512, 1999.
- [4] Dobra, A., Jones, B., Hans, C., Nevis, J. and West, M., Sparse graphical models for exploring gene expression data, *J. Multivariate Analysis*, 90:196–212, 2004.
- [5] Fan, J. and Li, R., Variable selection via nonconcave penalized likelihood and its oracle properties, *J. Am. Statist. Assoc.* 96:1348–1360, 2001.
- [6] Gustafsson, M., Hornquist, M. and Lombardi, A., Constructing and analyzing a large-scale gene-to-gene regulatory network-lasso-constrained inference and biological validation, *IEEE Trans. on Computational Biology and Bioinformatics*, 2:254–261, 2005.
- [7] Jeong, H., Tombor, B., Albert, R., Oltvai, Z. and Barabasi, A., The large-scale organization of metabolic networks, *Nature*, 407:651–654, 2000.
- [8] Kass, R. E. and Raftery, A. E., Bayes factor and model uncertainty, *J. Am. Statist. Assoc.*, 90:773–795, 1995.
- [9] Konishi, S., Ando, T. and Imoto, S., Bayesian information criteria and smoothing parameter selection in radial basis function networks, *Biometrika*, 91:27–43, 2004.
- [10] Laule, O., Fürholz, A., Chang, H.S., Zhu, T., Wang, X., Heifetz, P.B., Gruissem W. and Lange, M., Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in *Arabidopsis thaliana*, *Proc. Natl. Acad. Sci. USA*, 100:6866–6871, 2003.
- [11] Leng, C., Lin, Y. and Wahba, G., A note on the lasso and related procedures in model selection, *Statistica Sinica*, 16:1273–1284, 2006.
- [12] Lichtenthaler, H.K., Schwender, J., Disch, A. and Rohmer, M., Biosynthesis of isoprenoids in higher plant chloroplasts proceeds via a mevalonate-independent pathway, *FEBS Lett.*, 400:271–274, 1997.
- [13] Maslov, S. and Sneppen, K., Specificity and stability in topology of protein networks, *Science*, 296:910–913, 2002.
- [14] Meinshausen, N. and Bühlmann, P., High dimensional graphs and variable selection with the lasso, *Ann. Statist.*, 34:1436–1462, 2006.
- [15] Rodriguez-Concepcion, M. and Boronat, A., Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. A metabolic milestone achieved through genomics, *Plant. Physiol.*, 130:1079–1089, 2003.
- [16] Schäfer, J., Strimmer, K., An empirical Bayes approach to inferring large-scale gene association networks, *Bioinformatics*, 21:754–764, 2005.
- [17] Tibshirani, R., Regression shrinkage and selection via the lasso, *J. R. Statist. Soc. B*, 58:267–288, 1996.
- [18] Tierney, L. and Kadane, J. B., Accurate approximations for posterior moments and marginal densities, *J. Am. Statist. Assoc.*, 81:82–86, 1986.
- [19] Toh, H. and Horimoto, K., Inference of a genetic network by a combined approach of

12 *T. Shimamura, S. Imoto, R. Yamaguchi & S. Miyano*

- cluster analysis and graphical Gaussian modeling, *Bioinformatics* 18:287–297, 2002.
- [20] Wille, A., Zimmermann, P., Vranová, E., Fürholz, A., Laule, O., Bleuler, S., Hennig, L., Prelić, A., Rohr, P. V., Thiele, L., Zitzler, E., Gruissem, W. and Bühlmann, Sparse graphical Gaussian modeling of the isoprenoid gene network in *Arabidopsis thaliana*, *Genome Biology*, 5:1-13, 2004.
- [21] Yuan, M. and Lin, Y., On the nonnegative garrote estimator, *J. R. Statist. Soc.*, B, 69, 143–161, 2007.
- [22] Zou, H., Adaptive lasso and its oracle properties, *J. Am. Statist. Assoc.*, 101:1418–1429, 2006.