GRAPHICAL LASSO AND ITS APPLICATIONS

ALEX WASIELEWSKI, HENGRUI LUO, NATE WILAIRAT, XIAOFEI ZHOU

Abstract. In this report we discuss the theoretic motivation for the graphical LASSO methodology and conducted corresponding simulation studies on simulated and real datasets. A thorough discussion of two approaches toward the derivation of graphical LASSO are covered in the first part of the paper. In the second part of the paper a careful data analysis is carried out with appropriate comparison and tuning of parameters. We ended up with the report with a discussion about the graphical LASSO as well as its features.

Contents

1. Introduction to Graphical Models

Graphical models are used everywhere with applications in imaging, network analysis, and Bayesian statistics. It is always of central interest to investigate the dependency relations in the graph model. In this report we briefly introduce the formulation of the problem of estimating the covariance matrix when the underlying probabilistic model of an undirected graph is assumed to be Gaussian. And following [1, 2, 5], we explain how the covariance estimation problem for Gaussian process can be regarded as a regression problem as well as a neighborhood detection problem for the undirected graph. Once we establish these perspectives, we can apply classic regression techniques like LASSO when the dimension of the random variables in the undirected graph is high, even for $p > n$. Traditionally this presents difficulty in estimation as well as consistency of the estimators, however a penalized likelihood estimate like LASSO can accommodate the problem neatly with high efficiency.

The model for continuous i.i.d. p-dimensional data $X = (X_1, \cdots, X_p)$ assumes that the random variables are represented by vertices in the graph and distributed as a multivariate Gaussian distribution $N_p(\mu, \Sigma)$ with mean μ and covariance matrix Σ, or equivalently the precision matrix $\Theta = \Sigma^{-1}$, respectively. In an undirected graph, the random variables are represented by the vertices and the conditional dependency relations are represented by edges connecting two vertices. If there is no edge between two vertices, then it means that the corresponding pair of random variables associated with the vertices are conditionally independent given all the rest of the random variables represented by all other vertices in the graph. Let us consider the following example from [2]. This kind of rule actually provides a way of representing the dependency structure between Gaussian variables.

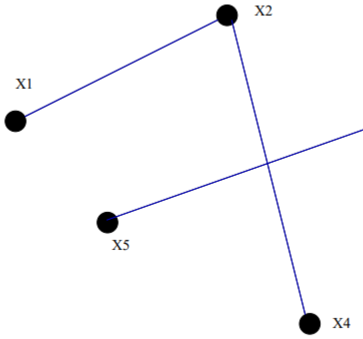


Figure 1.1. Simple example of $\{X_1, \cdots, X_5\}$ represented by an undirected graphs

In the Figure 1.1, we used an example from [2]. We will refer to this example in the following section.

Consider the adjoint matrix (or adjacent matrix or edge matrix) representation of this graph Figure 1.1, we get a 5×5 matrix where 1 means two vertices are joined and 0 means they are not joined. Then the adjoint matrix can be viewed as a forgetful image of the inverse covariance matrix. If we replace the entries in the adjoint matrix with the covariance between two vertices, then we can transform the adjoint matrix into the covariance matrix. Similarly we can also consider their correlation matrix after normalization by their respective variances.

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix} \leftrightarrow \begin{pmatrix} Cov(X_1, X_1) & Cov(X_2, X_1) & 0 & 0 & 0 \\ Cov(X_1, X_2) & Cov(X_2, X_2) & 0 & Cov(X_4, X_2) & 0 \\ 0 & 0 & Cov(X_3, X_3) & 0 & Cov(X_5, X_3) \\ 0 & Cov(X_2, X_4) & 0 & Cov(X_4, X_4) & 0 \\ 0 & 0 & Cov(X_3, X_5) & 0 & Cov(X_5, X_5) \end{pmatrix}$$

For the Gaussian multivariate distribution, the precision matrix captures the conditional distribution of each $X_j, j = 1, \cdots, p$. To see this point we consider the case $p = 2$,

$$N_2(\mu, \Sigma) = N_2\left(\left(\begin{array}{c}\mu_1\\ \mu_2\end{array}\right), \left(\begin{array}{cc}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{array}\right)\right)$$

and $X_1 \mid X_2 = a \sim N\left(\mu_1 + \sigma_{12}\sigma_{22}^{-2}(a - \mu_2), \sigma_{11} - \sigma_{12}\sigma_{22}^{-1}\sigma_{21}\right)$. The precision matrix of Σ is

$$\Theta = \left(\begin{array}{cc}\sigma_{22} & -\sigma_{12}\\ -\sigma_{21} & \sigma_{11}\end{array}\right) \cdot \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}\sigma_{21}}$$

and we can regard the edges in a undirected graph as the nonzero entries in the precision matrix and the zero entries means conditional independence between two variables. The reason why we use the precision matrix instead of covariance matrix directly is that we are capturing the conditional dependency between X's, whose distribution is parameterized by the precision matrix. In another sense, the conditional expectation, as the best linear unbiased predictor (BLUP) and L^2 optimizer, also makes use of the precision matrix in a closed form expression and thus can capture the conditional dependence better. We will also see later that this is a natural choice in terms of algorithmic consideration.

Take the covariance matrix as a functional of the data generating mechanism. The estimation of this functional can be reduced to discrete optimization problem. However, the computational complexity for an exhaustive method of estimation of the covariance matrix is generally exponential [1, 2] and therefore not appropriate. Also, we want to take into consideration the dependence structure exhibited in the undirected graph when making inference, especially when the data is sparse. These two concerns jointly lead to the invention of graphical LASSO, a mixture of two methods which can take care of the sparsity and efficient computing problems.

2. Graphical LASSO and Neighborhood Detection

In the seminal paper [5], the central problem is to estimate the covariance matrix. The author proposed to estimate the pattern zeros in the covariance matrix, which contains information about the conditional independence information in the graph representation of the collection of observed random variables $X = (X_1, \cdots, X_p)$. To estimate those structural zeros in the covariance matrix is equivalent to estimating the neighborhood of each vertex, through this perspective. The covariance estimation problem can be regarded as discovering the conditional independence structure contained in those i.i.d. observations. A stronger dependency between two random variable vertices means a partial correlation in the correlation matrix with absolute value closer to one; a weaker dependency between two random variable vertices means a partial correlation in the correlation matrix with absolute value closer to zero. Now we examine the conditional distribution of $X_j \mid X_{-j}$ which means we conditioned on all other vertices in the graph except for X_j. Since we assume the underlying model is Gaussian, we can use the closed form conditional distribution. In the following notations, we assume $\Sigma = [\![\sigma_{ij}]\!]_{i,j}^n, \Sigma^{-1} = \Theta = [\![\theta_{ij}]\!]_{i,j}^n$ and more precisely $\theta_{ij} = -\sigma_{ij|-ij} \det \Sigma_{-ij} \det \Sigma^{-1}$ where $\sigma_{ij|-ij}$ is the marginal covariance conditioning on all other random variables except for the i, j-th random variables; Σ_{-ij} is the covariance matrix with the i-th row and j-th column removed. We then have,

$$X_j \mid X_{-j} \sim N_1\left(\sum_{i \neq j} X_i\beta_{ij}, \sigma_{jj}\right)$$
$$\beta_{ij} = -\frac{\theta_{ij}}{\theta_{jj}}$$
$$\sigma_{jj} = \frac{1}{\theta_{jj}}$$

If we regarded the partial correlation between two random variables as geometric "angles" between two random variables, the closed form of the conditional distribution above shows that the mean of $X_j \mid X_{-j}$ is the sum of projected images of X_j onto other random variables $X_{-j} = \{X_1, \cdots, X_{j-1}, X_{j+1}, \cdots, X_n\}$ and the variance is the norm of this mean. Moreover,

$$\beta_{ij} \propto \frac{\theta_{ij}}{\sqrt{\theta_{ii} \cdot \theta_{jj}}}$$

and $\beta_{ij} = 0 \Leftrightarrow \theta_{ij} = 0 \Leftrightarrow X_i \perp X_j \mid X_{-ij}$ (i.e. X_i, X_j are independent conditioning on all other random variables except for i, j-th random variable). For our example in Figure 1.1, $X_2 \perp X_3 \mid X_1, X_4, X_5$. Therefore there is an one-to-one correspondence between neighborhoods and nonzero entries of the covariance matrix. Observe that the conditional distribution with mean $\sum_{i \neq j} X_i \beta_{ij}$ of $X_j \mid X_{-j}$ can be regarded as a regression model. Therefore, it is tempting to estimate the covariance matrix using the classic estimate from the regression setup. That is to say we are trying to estimate the response X_j by regressing on the rest of the random variables, X_{-j}, with coefficients $\beta_{ij}, i \neq j$. Currently based on the Gaussian graphical model, the author [1] proposed to derive the joint likelihood based (i.e. the joint likelihood of $\{X_1, \cdots, X_n\}$) estimate

$$\ell_\lambda(\Theta) = \log\det\Theta - Tr(S\Theta) \tag{2.1}$$

such that S is the empirical covariance matrix of the data. We can differentiate with respect to the Θ and solve following estimate equation

$$\frac{\partial \ell(\Theta)}{\partial \Theta} = \Theta^{-1} - S = 0$$

However, such an estimate is not very useful since it is simply the moment estimate without considering the underlying dependence structure provided by the undirected graph.

There are two ways of deriving the graphical LASSO method from this point. One way is to follow [5], that is to say we can regress each vertex X_j as a regression function of all the other random variables that is a neighbor to X_j (i.e. there is at least an edge connecting X_j and the vertex representing the random variable). Again in our simplest example above, X_2 is regressed on X_1, X_4 only. Then $\widehat{\beta_{ij}}$ can be used for estimating the covariance/correlation between X_2 and other random variables respecting the structure of the undirected graph. After doing the same regression procedure for each $X_1, \cdots, X_n$ we can obtain an estimate for Σ^{-1} and hence an estimate to the covariance matrix with respect to the underlying graphical dependency structure. If we take this approach, then this becomes an obvious problem of iterative updating of each vertex in the undirected graphs until the final solution to the estimate actually converges. One common method that can be applied to fit this conditional model is the cyclical coordinate descent algorithm [8] that optimizes the objective function (e.g. L^2 loss function) with respect to only one regressor coordinate at one time and cyclically updates coordinates until the algorithm converges.

The other way [1] is to modify the likelihood equation directly. Instead of doing a maximal likelihood equation, we consider its L^1 penalized version of the likelihood above with a tuning parameter $\lambda > 0$.

$$\ell_\lambda(\Theta) = \log\det\Theta - Tr(S\Theta) - \lambda\|\Theta\|_1 \tag{2.2}$$

$$\frac{\partial \ell(\Theta)}{\partial \Theta} = \Theta^{-1} - S - \lambda \cdot sgn\,(\Theta) = 0$$

where the function $sgn\,(\Theta) \coloneqq [\![sgn(\theta_{ij})]\!]_{ij}^{n}$ is simply applying the sign function to each entry of the precision matrix. This reproduces our adjoint matrix representation of the undirected graph. One step further we can realize that this is exactly the penalized estimate equation of LASSO on regression coefficients. One comment is that if we return to our geometric interpretation of the regression coefficients β_{ij} as “angles” between two random variables, then the penalized estimation in this scenario is no more than scaling the angles to achieve a dimension reduction for the space consisting of the regression basis. Consider the following equation in W: $W = \Theta^{-1}$.

$$\begin{pmatrix} W_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix}_{p \times p} \begin{pmatrix} \Theta_{11} & \theta_{12} \\ \theta_{12}^T & \theta_{22} \end{pmatrix}_{p \times p} = I_p$$

by the partitioned matrix multiplication formula we can solve $w_{12} = -W_{11}\frac{\theta_{12}}{\theta_{22}} = W_{11}\beta_{12}$. Therefore the upper right block of the equation

$$\begin{aligned}\Theta^{-1} - S - \lambda \cdot sgn\left(\Theta\right) &= \begin{pmatrix} W_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix}_{p\times p} - \begin{pmatrix} S_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix} - \lambda \begin{pmatrix} sgn\Theta_{11} & sgn\theta_{12} \\ sgn\theta_{12}^{T} & sgn\theta_{22} \end{pmatrix} \\ &= \begin{pmatrix} W_{11} & W_{11}\beta_{12} \\ w_{21} & w_{22} \end{pmatrix}_{p\times p} - \begin{pmatrix} S_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix} - \lambda \begin{pmatrix} sgn\Theta_{11} & sgn\theta_{12} \\ sgn\theta_{12}^{T} & sgn\theta_{22} \end{pmatrix}\end{aligned}$$

can be recognized as the LASSO penalized equation $W_{11}\beta_{12} - s_{12} + \lambda \cdot sgn(\beta_{12}) = 0$ for the regression coefficient β_{12}. These two methods are shown to coincide by the consideration above. The following graphical LASSO method is mentioned in [2]:

- Initialize $W = S + \lambda I$ where S is the empirical covariance matrix of the data. The diagonal of W remains unchanged in what follows because the w_{22} is never involved in the derivation above.
- Repeat for $j = 1, 2, ...p$ until certain convergence criterion is met.
  - Partition the matrix W into
    * W_{11}=all but the jth row and column are zeros.
    * W_{-11}=Only the jth row and column are nonzero.
  - Solve the estimating equations $W_{11}\beta_{12} - s_{12} + \lambda \cdot sgn(\beta_{12}) = 0$ above using the *cyclical coordinate-descent algorithm* [8] for β_{12} and obtain $\hat{\beta}$. Update $w_{12} = W_{11}\hat{\beta}$ and then obtain $\hat{\theta_{12}}$ from the regression function.
- In the final cycle (for each j) solve for $\hat{\theta_{12}} = -\hat{\beta} \cdot \hat{\theta_{22}}$ with $\frac{1}{\hat{\theta_{22}}} = w_{22} - w_{12}^{T}\hat{\beta}$.

Two major challenges exist when high dimensionality exists. First the computation complexity grows very fast; secondly classic estimates like MLE and method of moments may fail to exist. In a high dimensional sample space $\mathcal{X}$ where the observed dependence relations are sparse, the estimation is hindered by the lack of sample size n of the data.

From the theoretic consideration above, we can see that penalized regressions are powerful tools for handling sparsity both in computation and in estimation. In our setting of estimating the precision matrices, they provide faster convergence rates as well as adaptation to more sparse data as shown in our data analysis part. Therefore, graphical LASSO shall be considered as a central tool to the construction and estimation of Gaussian graphical models.

3. Tuning of Penalty Parameters

As other penalization methods, the graphical LASSO also needs a way of choosing a value of penalty parameter λ in (2.2). In this section we conducted a couple simulation studies in order to investigate the effect of different parameters.

Selection of the penalty parameter λ can be conducted via cross-validation or using an information criterion, most commonly BIC. A common rule of thumb to select a range of feasible λ is $(0, \max(|S^{-1}|_{ij}))$.
Our k-fold cross validation procedure would proceed as follows.

- Partition the dataset into k subsets. Let A_s represent the s-th subset and A_s^c be the complement of the s-th subset.
- For $s = 1, ..., k$, train the Graphical LASSO model on A_s^c .
- Evaluate the log-likelihood (2.1) using $\hat{\Theta}_{-s}$ (i.e. the estimated precision matrix without s-th subset) and the empirical covariance matrix S_{A_s} from the test subset.
- Set the estimated cross-validated likelihood $\hat{\alpha}_k(\lambda) = \frac{1}{k}\sum_{s=1}^{k} \ell(\hat{\Sigma}_{-s}, S_{A_s})$.
- Select the value of λ that maximizes the penalized log-likelihood using some criterion.

An alternative approach is to use an information criterion, such as Bayes Information Criterion (BIC), to select an optimal λ. [4] have shown that using BIC can lead to "consistent graphical model selection" when p is fixed. The effective degrees of freedom parameter (denoted by p in standard notation) is the sum of non-zero off-diagonal entries in the upper triangular part of the estimated precision matrix $\hat{\Theta}$ given by graphical LASSO. The value of λ that minimizes the BIC is selected as the penalty parameter. Here we define BIC where $p = dim(S)$ but p is not "model size" parameter. Denote that $\Theta = [\![\theta_{ij}]\!]_{ij}^{n}$ and $\hat{\Theta} = \left[\!\left[\hat{\theta}_{ij}\right]\!\right]_{ij}^{n}$.

$$BIC_{\lambda} = -n \log |\hat{\Theta}| + nTr(S\hat{\Theta}) + \log(n) \sum_{1 \le i < j \le p} (\hat{\theta}_{ij} \neq 0) \tag{3.1}$$

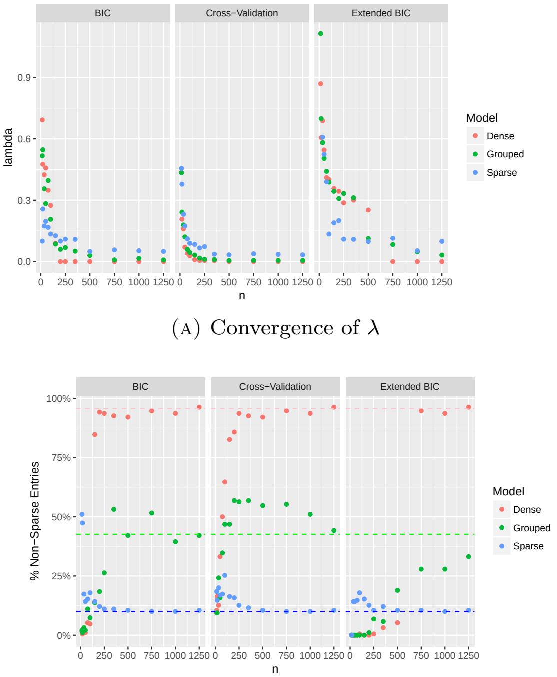
The extended BIC adds an additional term and is argued to be consistent when $p \to \infty$ and $n \to \infty$ [3]. Note that $\gamma \in [0, 1]$ where $\gamma = 1$ represents maximum penalization of large graphs. Previous simulation studies suggest extended BIC performs well when p is close to n.

$$BIC_{\gamma,\lambda} = BIC_{\lambda} + 4\gamma \log(p) \sum_{1 \le i < j \le p} I(\hat{\theta}_{ij} \neq 0) \tag{3.2}$$

To assess the performance of penalty parameter selection for the graphical LASSO, we created three simulations: a sparse precision matrix, a grouped precision matrix, and a dense precision matrix. The sparse precision matrix is an $AR(1)$ model, with $\theta_{ii} = 1$, $\theta_{i,i-1} = \theta_{i-1,i} = \frac{1}{2}$, and 0 otherwise. The grouped precision matrix defines two conditionally independent groups (block diagonal with two square matrices of dimension $\frac{p}{2}$), and the dense precision matrix has no zeroes. We used the gLASSO package in R to conduct the simulation study and selected a fixed dimension $p = 20$ with a range of sample sizes from 15 ($p < n$) to 1250 ($n \gg p$). The gLASSO package has no build-in penalty parameter selection function. For $n = 15$ and $n = 20$, we use 5-fold cross-validation and for $n > 20$ we used 10-fold cross-validation. For extended BIC, we used $\gamma = 1$.

In our simulation study we observed that for fixed p, as n increases, the log-likelihood evaluated at the average of estimated precision matrices from different cross-validation sets $\hat{\Theta}_{CV}(\lambda)$ converges to the true log-likelihood evaluated with the known Θ as expected [4]. We also observed that $||\hat{\Theta}(\lambda)||_1$ converges to the true $||\Theta||_1$ more quickly under cross-validation than BIC penalty parameter selection (see Appendix).

We also have $\lambda \to 0$ as $n \to \infty$ for the dense and grouped models. When $\lambda = 0$, then $\hat{\Theta} = S^{-1}$ since S^{-1} is the MLE, and we must round $\hat{\Theta}$ to induce sparsity. This is also the case for values of λ close to zero. We rounded elements of $\hat{\Theta}$ less than 0.05 to 0 in these cases. After rounding, we observe the sparsity of $\hat{\Theta}$ appears to converge towards the true sparsity (denoted by horizontal dashed lines below) under cross-validation, BIC, and extended BIC, but the rates of convergence differ.



(A) Convergence of λ

(B) Convergence of the estimated graph to the true graph

The cross-validation and BIC penalty parameter selection methods performed similarly in the limit $n \to \infty$. The BIC criteria perform better for the sparse model but poorly on the dense model (especially extended BIC) for small n. The BIC penalty appears to generally induce greater sparsity in the resulting estimated precision matrix than the CV approach at the same n.

In practice, this suggests that the BIC approach may be advantageous when the researcher expects a highly sparse Θ, whereas the CV approach may be preferred when Θ is dense or not known to be sparse. In the appendix we show the true sparse, grouped, and dense graphs and the estimated graphs for $n = 15$, $n = 150$, and $n = 1250$ under λ_{CV}, λ_{BIC}, and $\lambda_{Ext.BIC}$.

4. Data Analysis Application: Malaria Gene Pathway

Due to the high cost of the technologies, the sample size is always much smaller than the number of genes, rendering the gene covariance matrix noninvertible, and hence makes it impossible to obtain the precision matrix that indicates the conditional dependence among the genes. On the other hand, the pathways (i.e. dependence between genes) are sparse among genes, therefore graphical LASSO provides an effective way of estimation such a dependency relation between genes represented by the vertexes and the dependency represented by the edges.

Gene Expression Omnibus (GEO) is a public database that provides gene expression data with easy user access. Among the 4,349 datasets available in this database, we found gene expression data in malaria infection (GSE 5418) that provides pairwise transcription readings (experimental malaria-infected versus baseline uninfected) for 22 human subjects. The expression data for the 22 individuals were generated from Affymetrix's GPL96 microarray platform. Note that the cost of microarray technology has decreased significantly over the years (and is much lower than that of the next sequencing technologies), but the total expense of the biological analysis can still be high when the number of test subjects is large. GSE 5418 is the largest dataset we can find on GEO that has pairwise design on human subjects, and it provides readings on 22,283 different genes and genomic segments.

To identify the differentially expressed genes, we utilize the DIME software [7] (Differential Identification using Mixture Ensemble) that considers gene log ratios following some finite mixture model. The models under evaluation are gamma-normal-gamma (gng), normal-uniform (nudge), and multiple normal-uniform (inudge). Since the log intensities also play a role in indicating genetic signalÂŹ significance, the DIME software applies Huber weights that upweight genes with larger log Gaussian intensities. Here, we have tens of thousands of genes, so the false discovery rate (FDR) is used for multiple comparison correction.

The interaction among these differentially expressed genes are of particular interest. However, the number of differentially expressed genes may still be larger than the sample size and the covariance matrix is still noninvertible. Hence, graphical LASSO can be applied so that precision matrix can be achieved and weak conditional dependence among genes will be penalized to 0. Of course, we can use cross validation or extended BIC to identify the ideal tuning parameter for the restriction. Nevertheless, the graph resulting from the precision matrix should be sparse enough to only show the important information, so the number of connections can be prefixed as well.

We provide the results of applying graphical LASSO to our data under two scenarios. In the first, we focus on the genes that belong to the interferon signaling pathways, while, in the second we expand our focus to the top 100 differentially expressed genes (that have FDR smaller than 10^{-7}). In both cases the penalty parameter was chosen under the extended BIC criterion. Our estimated graphs are both shown on the following page.

The genes that extend multiple connections are called the hub genes and they can be considered as the control center for the infection process, as they share conditional dependence with various other genes. The hub genes are essential for diagnosing and treating diseases, and they are the keys for the genetic medicine development. In the plots that follow, CCL2 is a hub gene with several connections, and it is an immunoregulatory gene from the Chemokine family. The hub gene, CCL2, is apparent in both of the estimated graphs below Figure 4.1.

5. Conclusion

As we have seen, graphical LASSO [1] offers an efficient, quick solution to the problem of estimating the precision matrix of an undirected graph model. It applies very well to the data set where there exists some kind of sparsity. Compared to the naive approach of inverting the empirical covariance matrix, graphical LASSO provides a more consistent estimate to the precision matrix Θ with sparsity constraints such as high dimensionality. Moreover, numerically graphical LASSO is more stable compared to inversion of covariance. In the second perspective, we can equivalently treat the graphical LASSO as conditional iterative least

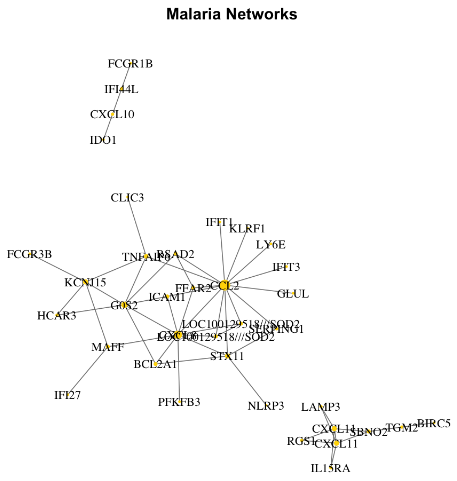


Figure 4.1. No Pathway (Top 100) Estimate

square estimation, which provides more insights into the dependency relations between the random variables represented by the vertexes. This makes the result of the method very intepretable especially attractive in fields such as high dimensional network, spatial analysis and genetics.

In addition, we discussed three ways proposed in the literature for selection of an optimal tuning parameter λ. Methods of cross-validation, BIC and extended BIC each have their strengths and weaknesses which we investigated deeply in our simulation study. Although the problem of choosing a penalty parameters for a regularization method like graphical LASSO is generally difficult [3, 4], we still implement and evaluate three different ways of choosing penalty parameters therein. To sum up our result, BIC or information criteria approach may be advantageous when the researcher expects a highly sparse precision matrix Θ, whereas the cross-validation approach may be preferred when Θ is dense or not known to be sparse.

Graphical LASSO is a valuable tool in network construction when sparsity is the goal or when the number of variables p is larger than the sample size n as we discussed above [7]. We illustrated this point further in our gene analysis example, where the pathway dependency between genes are sparse due to the sequencing technologies. In the real data application, we utilize the malaria data from 22 subjects on more than 20,000 genes. Even though the number of genes is significantly reduced from DE gene selection, it is still much higher than the number of subjects. By applying graphical LASSO, we are able to penalize the weak connections and construct a precision matrix to indicate conditional dependence.

In all, we explained the methodology of the graphical LASSO in two equivalent ways of precision matrix based undirected graph [1] and the iterative least square estimation [5]; and then discussed the effect of penalty parameters by simulation study comparing existing criteria [3, 4] and eventually we apply the graphical LASSO methodology onto the genetic analysis and showed its effectiveness when the dependency between variables are sparse [7].

References

[1] Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. "Sparse inverse covariance estimation with the graphical LASSO." Biostatistics 9.3 (2008): 432-441. (document)

[2] Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. The elements of statistical learning. Vol. 1. New York: Springer series in statistics, 2001. (document)

[3] Foygel, R., & Drton, M. "Extended Bayesian information criteria for Gaussian graphical models." In Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010, NIPS 2010. (document)

[4] Gao, Xin, Daniel Pu, Yuehua Wu, and Hong Xu. "Tuning parameter selection for penalized likelihood estimation of Gaussian graphical model." Statistica Sinica 22 (2012): 1123-1146. (document)

[5] Meinshausen, Nicolai, and Peter . "High-dimensional graphs and variable selection with the LASSO." The annals of statistics (2006): 1436-1462. (document)

[6] Mazumder, Rahul, and Trevor Hastie. "The graphical LASSO: New insights and alternatives." Electronic journal of statistics 6 (2012): 2125.

[7] Taslim, Cenny, Tim Huang, and Shili Lin. "DIME: R-package for identifying differential ChIP-seq based on an ensemble of mixture models." Bioinformatics 27.11 (2011): 1569-1570. (document)

[8] Wright, Stephen J. "Coordinate descent algorithms." Mathematical Programming 151.1 (2015): 3-34.

(document)