

Assessment of Covariance Selection Methods in High-Dimensional Gaussian Graphical Models

J. R. MALDONADO^{1*} and S. M. RUIZ²

Received on September 30, 2021 / Accepted on March 24, 2022

ABSTRACT. The covariance selection in Gaussian graphical models consists in selecting, based on a sample of a multivariate normal vector, all those pairs of variables that are conditionally dependent given the remaining variables. This problem is equivalent to estimate the graph identifying the nonzero elements on the off-diagonal entries of the precision matrix. There are different proposals to carry out covariance selection in high-dimensional Gaussian graphical models, such as neighborhood selection and Glasso, among others. In this paper we introduce a methodology for evaluating the performance of graph estimators, defining the notion of non-informative estimator. Through a simulation study, the empirical behavior of Glasso in different structures of the precision matrix is investigated and its performance is analyzed according to different degrees of density of the graph. Our proposal can be used for other covariance selection methods.

Keywords: covariance selection, gaussian graphical model, glasso.

1 INTRODUCTION

In the last two decades, high-dimensional Gaussian graphical models have become in a powerful tool to represent the conditional independence relationships between a large collection of random variables and they have gained importance in the modeling of problems of different disciplines in particular in human genetics (see for instance [18] and [23]).

More formally, if $X = (X_1, \dots, X_p)$ is a p -dimensional multivariate Gaussian vector with covariance matrix Σ , its inverse Ω represents the conditional association structure; i.e. if ω_{ij} denotes the i, j entry of the matrix Ω then $\omega_{ij} = 0$ iff X_i and X_j are conditionally independent given the remaining variables. A Gaussian graphical model (GGM) is the undirected graph $\mathcal{G} = (V, E)$ associated with X , where $V = \{1, \dots, p\}$ and E is the set of edges defined as $(i, j) \in E$ iff $\omega_{ij} \neq 0$. Thus, finding the set of variables that are conditionally dependent is equivalent to determining the set of nonzero elements in the precision matrix.

*Corresponding author: Juliana Maldonado – E-mail: jmaldonado@exa.unrc.edu.ar

¹Departamento de Matemática, FCEFQyN, Universidad Nacional de Río Cuarto, Ruta Nac. 36 - km 601, Río Cuarto, Córdoba, Argentina – E-mail: jmaldonado@exa.unrc.edu.ar <https://orcid.org/0000-0002-7142-6047>

²Departamento de Matemática, FCEFQyN, Universidad Nacional de Río Cuarto, Ruta Nac. 36 - km 601, Río Cuarto, Córdoba, Argentina – E-mail: mruiuz@exa.unrc.edu.ar <https://orcid.org/0000-0002-2336-9316>

Covariance selection (see [7]) is a denomination for the set of procedures that pursue the objective of identifying the conditional dependency in a GGM from a sample.

In high dimension, when the number of variables p is larger than the number n of observations, the sample covariance matrix \mathbf{S} is not invertible and the maximum likelihood estimate (MLE) of $\mathbf{\Omega}$ does not exist. When $p/n \leq 1$, but close to 1, \mathbf{S} is invertible but ill-conditioned, increasing the estimation error [16]. To deal with this problem, several covariance selection procedures have been developed assuming that the precision matrix is sparse, that is, it has few non-zero elements.

Meinshausen and Bühlmann [20] propose to estimate $\mathbf{\Omega}$ via Lasso and lay the foundations of the asymptotic theory when the sample size n and the number of variables p tend to infinity. Friedman et al. [11] propose to estimate $\mathbf{\Omega}$ with a Lasso regularization of the log-likelihood through a coordinate descent algorithm and call their proposal Glasso. Another alternative is a constrained l_1 -minimization for inverse matrix estimation (CLIME) due to Cai et al. [5]. Lafit et al. [14] propose to estimate the graph and the precision matrix with a step-by-step algorithm, called StepGraph, suggesting that both Glasso and CLIME could be highly sensitive to the structure of $\mathbf{\Omega}$ (see also the arxiv version [13]).

In order to evaluate the sensitivity of an estimation method to the structure and density of a GGM, our proposal consists of introducing the notion of “non-informative estimator”, based on the sensitivity and specificity measures. In high-dimensional GGM estimation the computation time is very important. As this time is much higher for StepGraph and CLIME than for Glasso, in our simulation study we will focus only on Glasso. Through this study we will show that the graph recovery performance of Glasso is highly sensitive both to the number of edges of E and to the structure of $\mathbf{\Omega}$.

The rest of the article is organized as follows. Section 2 introduces some general settings and measures to evaluate GGM estimation. Section 3 gives our simulation results. Section 4 submit the analysis of a real data set. Finally, in Section 5 we present general conclusions.

2 EVALUATION OF GGM ESTIMATION

2.1 General settings

In this section we review some necessary definitions and concepts. Let $\mathcal{G} = (V, E)$ be a graph where $V \neq \emptyset$ is the set of nodes and $E \subseteq V \times V = V^2$ is the set of edges. For simplicity we assume that $V = \{1, \dots, p\}$. We assume that the graph \mathcal{G} is undirected, that is, $(i, j) \in E$ if and only if $(j, i) \in E$. Two nodes i and j are called connected, adjacent or neighbors if $(i, j) \in E$.

Let

$$(X_1, \dots, X_p)^\top \sim \mathbf{N}(\mathbf{0}, \mathbf{\Sigma}), \quad (2.1)$$

where $\mathbf{\Sigma} = (\sigma_{ij})_{i,j=1,\dots,p}$ is a positive-definite covariance matrix.

Given (2.1), its *Gaussian graphical model* (GGM) is the graph such that V indexes the set of variables $\{X_1, \dots, X_p\}$ and E is defined by:

$$(i, j) \notin E \text{ if and only if } X_i \perp\!\!\!\perp X_j \mid X_{V \setminus \{i, j\}},$$

where $\perp\!\!\!\perp$ denotes *conditional independence*.

There exists an extensive literature on GGM. For a detailed treatment of the theory see for instance [15], [9], and [4].

In a GGM the set of edges E represents the conditional dependence structure of the vector (X_1, \dots, X_p) . One way to represent this dependence structure as a statistical model is through a parametrization for E . Using well known results of classical multivariate analysis, given the *precision matrix* $\mathbf{\Omega} = (\omega_{ij})_{i, j=1, \dots, p} = \mathbf{\Sigma}^{-1}$, it can be proved that the set of edges is fully characterized by the support of the precision matrix

$$\text{SUPP}(\mathbf{\Omega}) = \{(i, j) \in V^2 : i \neq j \wedge \omega_{ij} \neq 0\}. \tag{2.2}$$

Namely,

$$\forall (i, j) \in V^2, i \neq j : (i, j) \in E \text{ if and only if } \omega_{ij} \neq 0. \tag{2.3}$$

For an exhaustive treatment of these results see, for instance, [2], [6], [15] and [8].

There exist a wide variety of structures for $\mathbf{\Omega}$ models, as for example:

- Nearest neighbors model of order k , denoted $\text{NN}(k)$ and described in [18]. For each node are selected k neighbors at random and then k symmetric entries of $\mathbf{\Omega}$ are chosen.
- Block diagonal matrix model with q blocks of size p/q , denoted $\text{BG}(q)$. Each block has diagonal elements equal to 1 and off-diagonal elements equal to 0.5.
- Random model, denoted $\text{RN}(prob)$. Given two nodes i and j , they are connected with probability “ $prob$ ”, obtaining a graph with approximately $p(p - 1)prob/2$ edges.

Hence, every structure define a family of models $\{\text{NN}(k)\}_k$, $\{\text{BG}(q)\}_q$ and $\{\text{RN}(prob)\}_{prob}$ indexed by the corresponding parameters.

These graph models are widely used in the genetic literature to model gene expression data as detailed in [18] and [17]. Figure 1 displays the graphs with $p = 100$ nodes of the models $\text{NN}(2)$, $\text{BG}(20)$ and $\text{RN}(0.01)$; it can be seen that they represent very distinct structures of conditional association between variables.

There are different options to quantify the sparsity of a graph (see [20], [22]). In this paper, given a GGM, we define a simple sparsity measure of its graph given by

$$d = \frac{N_L}{p(p - 1)/2},$$

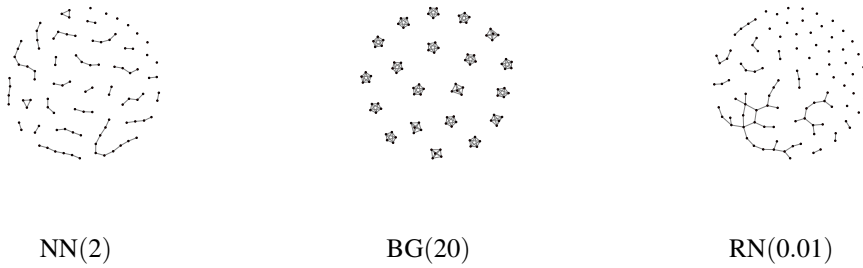


Figure 1: Graphs of NN(2), BG(20) and RN(0.01) graphical models for $p = 100$ nodes.

where N_L denotes the number of edges of the graph and $p(p - 1)/2$ is the total number of edges that a graph with p nodes can have. Thus d measures the density or sparsity of the graph and varies between 0 and 1. When $d = 1$ we call the graph *totally dense* and it has all possible edges and if $d = 0$ we call the graph *totally sparse* and $E = \emptyset$.

For every parameter k, q and $prob$ of NN(k), BG(q) and RN($prob$) respectively, we have an associated density d_k, d_q and d_{prob} . So, every family can be written as $\{NN(k)\}_{d_k}, \{BG(q)\}_{d_q}$ and $\{RN(prob)\}_{d_{prob}}$. In this way, we have three structures, each one being a family of Ω models, indexed by their densities.

Finally, it is important to emphasize that, the covariance selection or precision matrix estimation allows us, based on a data set, to obtain an estimate of the set of edges and therefore an estimate of the GGM.

2.2 Minimum density and not informative estimator

Let \mathbf{X} denote a $n \times p$ data matrix of the normal multivariate distribution (2.1) and let $\hat{\Omega} = (\hat{\omega}_{ij})_{i,j=1,\dots,p}$ be an estimator of Ω based on \mathbf{X} . So, according to (2.2) and (2.3), an estimation of the set of edges can be defined as $\hat{E} = \{(i, j) \in V^2 : i \neq j \wedge \hat{\omega}_{ij} \neq 0\}$.

Assuming that we know the true Ω model, to analyze the performance of the estimator, regarding the identification of the nonzero off-diagonal elements of Ω , we evaluate its ability to recover the graph E based on \hat{E} . For this purpose, we use the Matthews correlation coefficient [19]

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

and the measures

$$\text{specificity} = \frac{TN}{TN + FP} \text{ and } \text{sensitivity} = \frac{TP}{TP + FN},$$

where TP, TN, FP and FN are, in this order, the number of true positives, true negatives, false positives and false negatives. Larger values of MCC, sensitivity and specificity indicate a better performance [3, 10].

Assume that $\mathbb{G} = \{\Omega_d\}_d$ is a family of GGM indexed by the density of its members, like those introduced in Section 2.1 and let $\Omega_d \in \mathbb{G}$. We define

$$\tau = \tau(d, n) =: \min \{ \text{specificity}, \text{sensitivity} \}$$

We adopt as criteria to claim that the estimator is *not informative* if $\tau < 0.7$. The choice of the threshold 0.7 is due to the fact that this value is the maximum attained by τ with estimators such as CLIME and StepGraph that outperform Glasso for different models, sample sizes and values of p as it is shown in [14].

We further define the *minimum density* as

$$d_{\min} = d_{\min}(n) =: \min \{ d : \tau(d) < 0.7 \}.$$

Thus d_{\min} represents the lowest density or the lowest proportion of edges from which the estimator is no longer informative.

3 SIMULATION AND NUMERIC RESULTS

In this section, to estimate the precision matrix we use the algorithm developed by Friedman et al. [11] called Glasso, which is obtained by solving the following ℓ_1 penalized-likelihood problem:

$$\min_{\Omega > 0} \left(-\log \{ \det[\Omega] \} + \text{tr} \{ \Omega \mathbf{X}' \mathbf{X} \} + \lambda \|\Omega\|_1 \right), \tag{3.1}$$

where \mathbf{X} denotes, as before, the data matrix and the minimum in (3.1) is obtained over all non-negative semidefinite matrices. The R-package CVGLASSO implements Glasso by selecting the regularization parameter λ by cross validation with $K = 5$ folds. Thus, setting n and choosing the GGM, we obtain for each replicate an estimate of the precision matrix $\hat{\Omega}$ and therefore an estimate of the set of edges \hat{E} .

In our simulation study we empirically address Glasso’s performance in estimating the graph of a GGM, considering two main objectives: to study how sensitive Glasso is to variations in the structures of the precision matrix and to determine the number of edges that it can estimate before becoming uninformative.

3.1 Simulation scheme

We consider the dimension value $p = 100$ and the Ω structures introduced in Section 2.1:

Model 1. NN(k), with $k = 2, 5, 7, 10, 20, 30, 40, 50, 60$ and 80 . To generate this model we use the “NeighborOmega” function of the R-package TGLASSO.

Model 2. BG(q) with $q = 1, 2, 4, 5, 10, 20, 25, 50$ and 100 blocks. To generate it we use the “Bdiag” function of the R-package MATRIX.

Model 3. RN($prob$) with $prob = 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6$ and 0.7 . To generate it we use the function “huge.generator” of the R package HUGE.

For every Ω belonging to Model 1, 2 and 3 and sample size $n = 30, 50, 100, 200, 500$ we generate $R = 50$ replicates.

For each graph and its respective density d we calculate the means and standard deviations of sensitivity, specificity and MCC for the Glasso estimator in the R replicates of each n . From these measurements we obtained $\hat{\tau}(d, n) = \min \{ \overline{\text{specificity}}, \overline{\text{sensitivity}} \}$ where the bar indicates an average over the R replicates (note that $\hat{\tau}$ is obtained by plug-in the Monte Carlo computation of the sensitivity and specificity of the estimator of the Ω matrix). Then, we compute \hat{d}_{\min} . For simplicity, we will omit the “hat” symbol.

3.2 Results

Figure 2 shows, for Models 1-3, the performance of the MCC for Glasso as a function of the density coefficient of the graph, according to the sample size. Different behavior patterns of the MCC are observed for the three models. Glasso’s performance improves when n grows and it gets worse when the density of the graph increases.

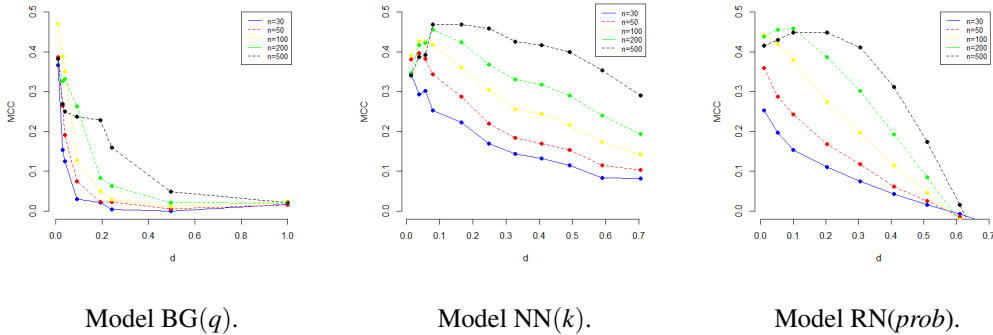
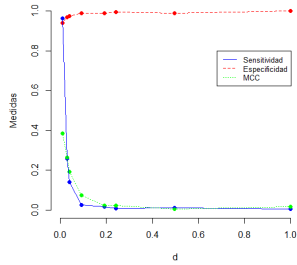


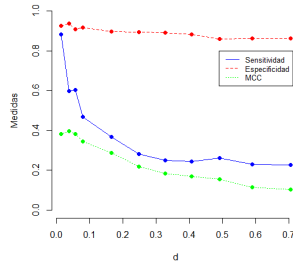
Figure 2: MCC curves for the Glasso estimator, according to sample size n , as functions of the density coefficient of the graph.

Figure 3 display sensitivity, specificity and MCC curves for the Glasso estimator, according to sample size n , as functions of the density coefficient of the graph. Notice that in all three types of models, the Glasso estimator tends to estimate with low sensitivity and high specificity, which implies that there are few false positives but many false negatives. In other words, Glasso is conservative and tends to incorporate few edges but, as shown in [14], it incorporates more false positives than other estimators such as CLIME and StepGraph.

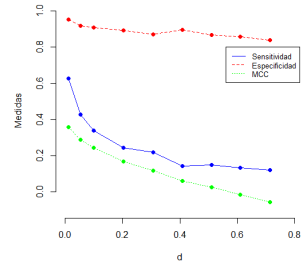
Table 1 shows de minimum (estimated) density d_{\min} for the Models 1-3 and different sample sizes. We observe that for the $BG(q)$, $NN(k)$ and $RN(prob)$ models, Glasso produces poor estimates for graphs that have more than 9%, 25% and 31% of edges, respectively. Furthermore, in the $NN(k)$ model, for example, when $p > n$, good estimates can be obtained for graphs that have only 4% edges, and if $n = 100$ or $n = 200$ this percentage increases. This and what happens for



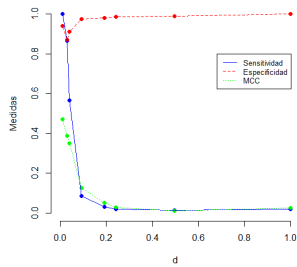
Model BG: $n = 50$



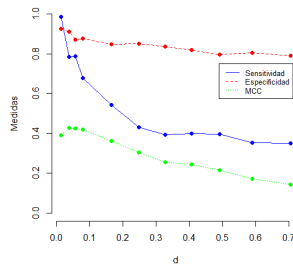
Model NN: $n = 50$



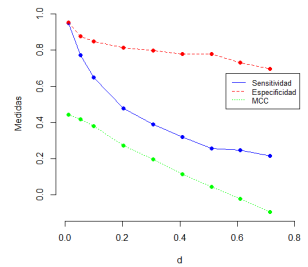
Model RN: $n = 50$



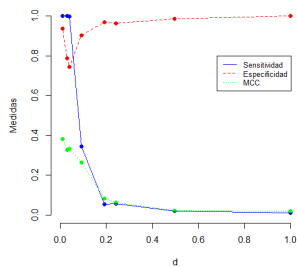
Modelo BG: $n = 100$



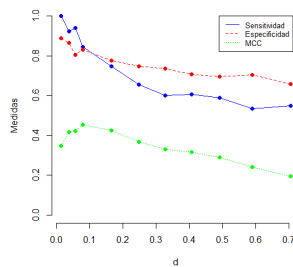
Modelo NN: $n = 100$



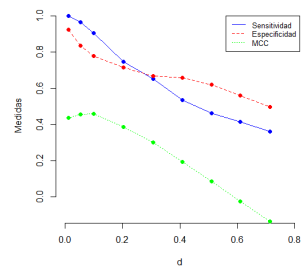
Modelo RN: $n = 100$



Model BG: $n = 200$



Model NN: $n = 200$



Model RN: $n = 200$

Figure 3: Sensitivity, specificity and MCC curves for the Glasso estimator, according to sample size n , as functions of the density coefficient of the graph.

the other two models, indicates that to achieve a good estimate in denser graphs it is necessary to increase the sample size.

Table 1: d_{\min} for the three models and different sample sizes

Model	$n = 30$	$n = 50$	$n = 100$	$n = 200$
BG(q)	0.03	0.03	0.04	0.09
NN(k)	0.04	0.04	0.08	0.25
RN($prob$)	no inf	no inf	0.1	0.31

4 REAL DATA ANALYSIS

Patients with breast cancer treated by neoadjuvant chemotherapy can reach two states, “pathological complete response” (pCR) or “residual disease” (RD). pCR state is associated with the long-term cancer-free survival and RD indicates that disease persists.

Based on measurements of gene expression levels Hess et al. [12] developed a multigene predictor of pCR or RD responses. Their data base has 22283 gene expression levels for 133 patients, with 34 pCR and 99 RD. Based on Natowicz et al. [21], Ambroise et al. [1] study the conditional dependence of 26 key genes estimating the graph, assuming the existence a latent structure of the network consisting of hidden clusters. They impose this structure supposing that a node belongs to only one cluster and they use a multinomial distribution to model this assumption.

In this section using the same dataset than in [1] we compare the estimated networks (graphs) for pCR, RD and for all patients, abbreviated as BOTH. We compare estimated graph using graphical lasso (Glasso), CLIME and StepGraph proposed by [11], [5] and [14], as we mentioned before. For Glasso and CLIME we use the R-packages CVGLASSO and CLIME respectively with the tuning parameter provided by package default. The package STEPGRAPH, written in R, was provided by the authors of [14]. Figures 4, 5 and 6 display the resulting network obtained from each of the estimation methods, and Table 2 exhibits the estimated network density for the 26 genes for each class.

Note that in the three classes, RD, pCR and BOTH, StepGraph estimates a sparse network compared with Glasso and CLIME. Even more, StepGraph does not detect conditional dependence between any of the 26 genes in the pCR class and conversely, Glasso and CLIME estimates a more dense network compared to RD class. StepGraph shows that the behaviour of the structure dependence is very different in pCR state than in RD state.

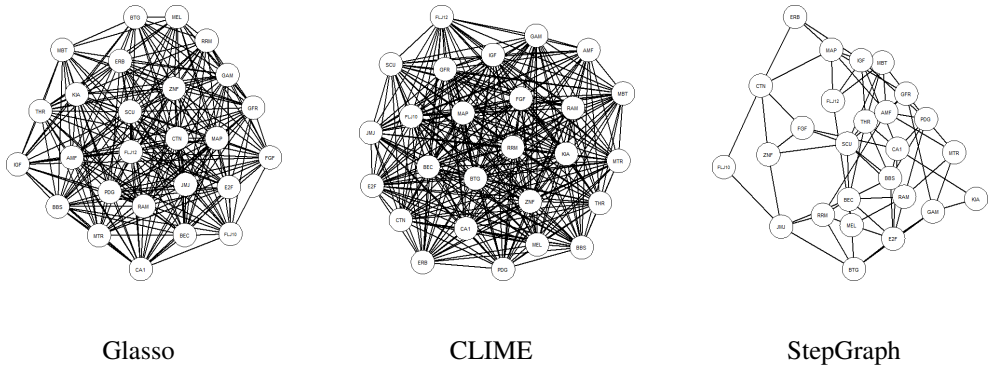


Figure 4: Estimated graph of the GGM for the 26 genes corresponding to RD class.

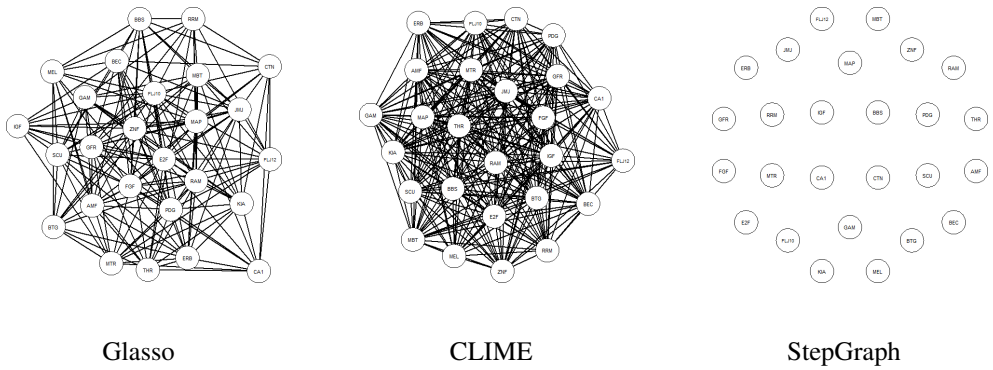


Figure 5: Estimated graph of the GGM for the 26 genes corresponding to pCR class.

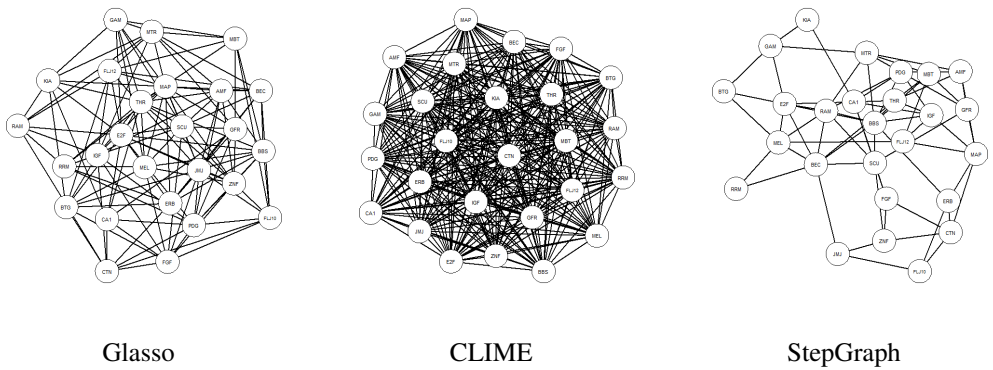


Figure 6: Estimated graph of the GGM for the 26 genes corresponding to BOTH classes.

Table 2: Estimated network density for the 26 genes from breast cancer gene expressions data.

	RD	pCR	BOTH
Glasso	0.745	0.535	0.394
CLIME	0.926	0.895	0.988
StepGraph	0.215	0.000	0.209

5 CONCLUSIONS

This paper introduces a methodology to evaluate the performance of Gaussian graphical model estimators, based on the definition of non-informative estimator notion.

Through a simulation study, the empirical behavior of Glasso in different structures of the precision matrix is investigated and its performance is analyzed according to different degrees of density of the graph. The results obtained show that, in terms of the graph recovery, Glasso is not invariant to the precision matrix structure. Indeed, the graph recovery performance depends significantly on the number of edges or density and this dependence is different according to the structure.

Our proposal can be used for other covariance selection methods.

We compared for a real data set the estimated networks using Glasso, CLIME and StepGraph. Glasso and CLIME estimates more dense graphs than StepGraph. In practice, the researchers' knowledge of the field to which the statistical methodology is applied is decisive.

REFERENCES

- [1] C. Ambroise, J. Chiquet & C. Matias. Inferring sparse Gaussian graphical models with latent structure. *Electronic Journal of Statistics*, **3** (2009), 205–238.
- [2] T. Anderson. “An Introduction to Multivariate Statistical Analysis”. John Wiley (2003).
- [3] P. Baldi, S. Brunak, Y. Chauvin, C. Andersen & H. Nielsen. Assessing the accuracy of prediction algorithms for classification: An overview. *Bioinformatics*, **16**(5) (2000), 412–424.
- [4] P. Bühlmann & S. Van De Geer. “Statistics for high-dimensional data: methods, theory and applications”. Springer Science & Business Media (2011).
- [5] T. Cai, W. Liu & X. Luo. A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, **106**(494) (2011), 594–607.
- [6] H. Cramér. “Mathematical Methods of Statistics”. Princeton University Press (1999).
- [7] A.P. Dempster. Covariance selection. *Biometrics*, (1972), 157–175.
- [8] M.L. Eaton. “Multivariate Statistics : A Vector Space Approach”. Institute of Mathematical Statistics (2007).
- [9] D. Edwards. “Introduction to Graphical Modelling”. Springer Science & Business Media (2000).

- [10] J. Fan, Y. Feng & Y. Wu. Network exploration via the adaptive LASSO and SCAD penalties. *The Annals of Applied Statistics*, **3**(2) (2009), 521–541.
- [11] J. Friedman, T. Hastie & R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, **9**(3) (2008), 432–441.
- [12] K.R. Hess, K. Anderson, W.F. Symmans, V. Valero, N. Ibrahim, J.A. Mejia, D. Booser, R.L. Theriault, A.U. Buzdar, P.J. Dempsey *et al.* Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. *Journal of Clinical Oncology*, **24**(26) (2006), 4236–4244.
- [13] G. Lafit, F. Nogales, M. Ruiz & Z. Ruben. A Stepwise Approach for High-Dimensional Gaussian Graphical Models. *arXiv:1808.06016*, (2018).
- [14] G. Lafit, F. Nogales, M. Ruiz & Z. Ruben. A Stepwise Approach for High-Dimensional Gaussian Graphical Models. *Journal of Data Science, Statistics, and Visualisation*, (2021). In press.
- [15] S.L. Lauritzen. “Graphical Models”. Oxford University Press (1996).
- [16] O. Ledoit & M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, **88**(2) (2004), 365–411.
- [17] W. Lee & Y. Liu. Joint Estimation of Multiple Precision Matrices with Common Structures. *Journal of Machine Learning Research*, **16** (2015), 1035–1062.
- [18] H. Li & J. Gui. Gradient directed regularization for sparse Gaussian concentration graphs, with applications to inference of genetic networks. *Biostatistics*, **7** (2006), 302–317.
- [19] B. Matthews. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta*, **405**(2) (1975), 442–451.
- [20] N. Meinshausen & P. Bühlmann. High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, **34**(3) (2006), 1436–1462.
- [21] R. Natowicz, R. Incitti, E.G. Horta, B. Charles, P. Guinot, K. Yan, C. Coutant, F. Andre, L. Pusztai & R. Rouzier. Prediction of the outcome of preoperative chemotherapy in breast cancer using DNA probes that provide information on both complete and incomplete responses. *BMC bioinformatics*, **9**(1) (2008), 1–17.
- [22] M. Pourahmadi. “High-Dimensional Covariance Estimation”. Wiley (2013).
- [23] J. Yin & H. Li. A sparse conditional Gaussian graphical model for analysis of genetical genomics data. *The Annals of Applied Statistics*, **5**(4) (2011), 2630 – 2650. doi:10.1214/11-AOAS494. URL <https://doi.org/10.1214/11-AOAS494>.

