Undirected Graphical Models

Yanrong Yang

RSFAS/CBE, Australian National University

25th October 2022

Motivation of Undirected Graphical Models

Undirected Graphical Model gives a visual way of understanding the joint distribution of several random variables.

- A graph consists of a set of vertices (nodes), along with a set of edges joining some pairs of the vertices.
- Each vertex represents a random variable.
- The edges are parameterized by values or potentials that encode the strength of the conditional dependence between the random variables at the corresponding vertices.

Let X, Y, Z be random variables. X and Y are conditonally independent given Z (written by $X \perp Y|Z$), if for each value $Z = z$, X and Y are independent in the conditonal distribution given $Z = z$, i.e.

$$f_{X,Y|Z}(x, y|z) = f_{X|Z}(x|z) f_{Y|Z}(y|z). \tag{1}$$

Example of An Undirected Graph

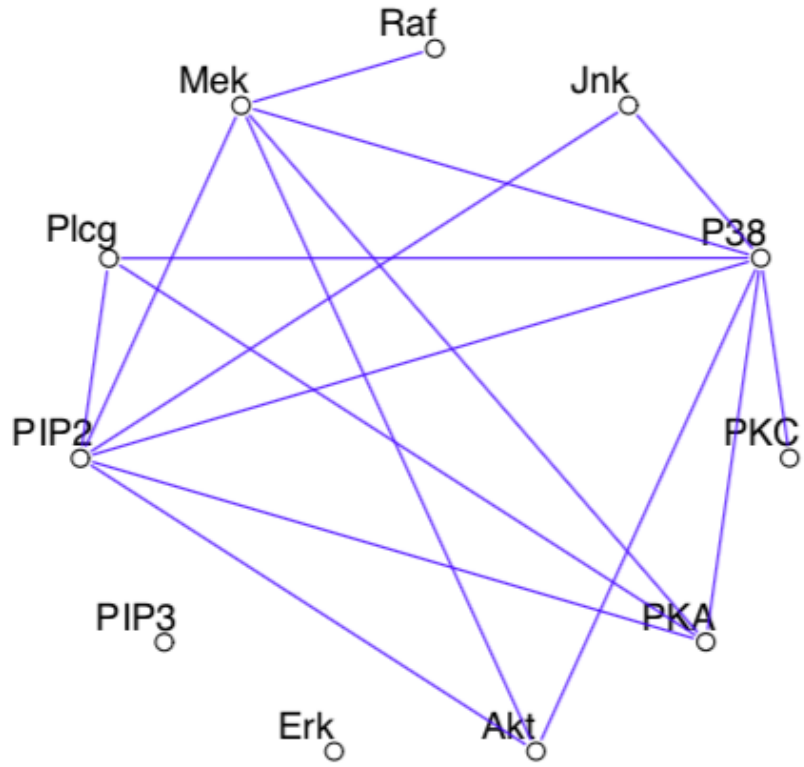


FIGURE 17.1. *Example of a sparse undirected graph, estimated from a flow cytometry dataset, with $p = 11$ proteins measured on $N = 7466$ cells. The network structure was estimated using the graphical lasso procedure discussed in this chapter.*

Basic Elements of Markov Graphs

A markov graph $\mathcal{G}$ consists of a pair (V, E), where V is a set of vertices and E is the set of edges.

- Adjacent: two vertices X and Y are adjacent if there is an edge joining them. This is written as $X \sim Y$.
- Path: a path $X_1, X_2, \ldots, X_n$ is a set of vertices that are joined, that is $X_{i-1} \sim X_i$, $i = 2, \ldots, n$.
- A complete graph is a graph with every pair of vertices joined by an edge.
- A subgraph $U \in V$ is a subset of vertices together with their edges.
- A subgraph C separates subgraphs A and B if every path between A and B intersects a node in C.

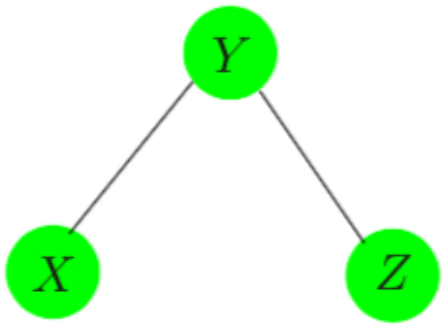
Markov Graph Modelling Conditional Dependence

In a markov graph $\mathcal{G}$, the vertex set V represents a set of random variables having joint distribution P.

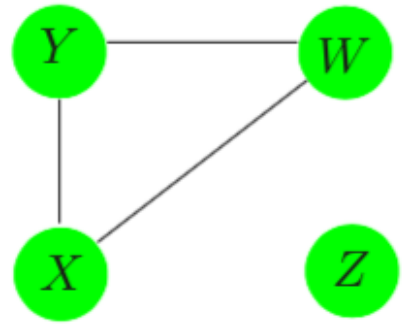
- pairwise Markov independence:
  "No edges joining X and Y" is equivalent to "$X \perp Y | rest$".
- global Markov properties:
  If a subgraph C separates subgraphs A and B, then $A \perp B | C$.

The global Markov property allows us to decompose graphs into smaller more manageable pieces and thus leads to essential simplifications in computation and interpretation.

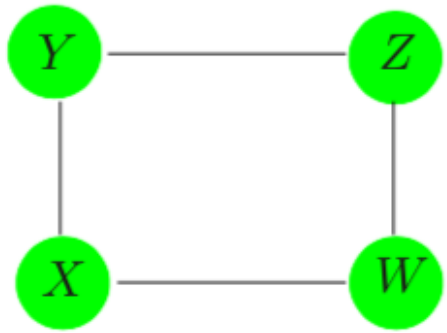
Examples of Markov Graphs



(a)



(b)



(c)

(d)

Gaussian Graphical Models

Precision matrix is informative on conditional dependence among random variables.

- The Gaussian distribution has the property that all conditional distributions are also Gaussian.
- The inverse covariance matrix Σ^{-1} contains information about the partial covariances between the variables, that is, the covariances between pairs i and j, conditonal on all other variables.
- If the (i, j)-th component of $\Theta = \Sigma^{-1}$ is zero, then variables i and j are conditionally independent, given the other variables.

Interpretation based on Quantitative Analysis

Suppose we partition $X = (Z, Y)$, where $Z = (X_1, \ldots, X_{p-1})$ and $Y = X_p$. Then we have the conditional distribution of Y given Z below

$$Y|Z = z \sim \mathcal{N}\left(\mu_Y + (z - \mu_Z)^\top \Sigma_{ZZ}^{-1}\sigma_{ZY}, \sigma_{YY} - \sigma_{ZY}^\top \Sigma_{ZZ}^{-1}\sigma_{ZY}\right). \quad (2)$$

- Zero elements in $\beta := \Sigma_{ZZ}^{-1}\sigma_{ZY}$ means that, the corresponding elements of Z are conditionally independent of Y given the rest.
- Decompose the covariance matrix Σ and the precision matrix Θ as follows.

$$\Sigma = \begin{pmatrix} \Sigma_{ZZ} & \sigma_{ZY} \\ \sigma_{ZY}^\top & \sigma_{YY} \end{pmatrix}, \quad \Theta = \begin{pmatrix} \Theta_{ZZ} & \theta_{ZY} \\ \theta_{ZY} & \theta_{YY} \end{pmatrix}. \quad (3)$$

Interpretation based on Quantitative Analysis

From the relation $\Sigma \cdot \Theta = I$ we have

$$\theta_{ZY} = -\theta_{YY}\Sigma_{ZZ}^{-1}\sigma_{ZY}, \quad \frac{1}{\theta_{YY}} = \sigma_{YY} - \sigma_{ZY}^{\top}\Sigma_{ZZ}^{-1}\sigma_{ZY}. \tag{4}$$

Then

$$\beta = \Sigma_{ZZ}^{-1}\sigma_{ZY} = -\frac{\theta_{ZY}}{\theta_{YY}}. \tag{5}$$

Thus the precision matrix Θ captures all the second-order information needed to describe the condtional distribution of each node given the rest, and is the so-called "natural" parameter for the Gaussian graphical model.

Likelihood Estimation for Precision Matrix

Given some realizations of $X = (X_1, X_2, \ldots, X_p)$, we would like to estimate the parameters Θ of an undirected graph that approximates their joint distribution.

- The sample covariance matrix is $S = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})^\top$.
- The log-likelihood of the Gaussian distributed data is

$$\ell(\Theta) = \log \det(\Theta) - \mathit{trace}(S\Theta). \tag{6}$$

- The quantity $-\ell(\Theta)$ is a convex function of Θ. The MLE of Θ can be derived by maximizing $\ell(\Theta)$.
- It is easy to show that the maximum likelihood estimate of Σ is S.

Penalized Likelihood Estimation for Precision Matrix

A common problem related to Θ is that, we know that there are some elements in Θ are zero.

- We consider the penalized log-likelihood function

$$\ell_C(\Theta) = \log\det(\Theta) - \mathit{trace}(S\Theta) - \sum_{(j,k)\notin E} \gamma_{jk}\theta_{jk}. \qquad (7)$$

- The gradient equation for maximizing (7) can be written as

$$\Theta^{-1} - S - \Gamma = 0. \qquad (8)$$

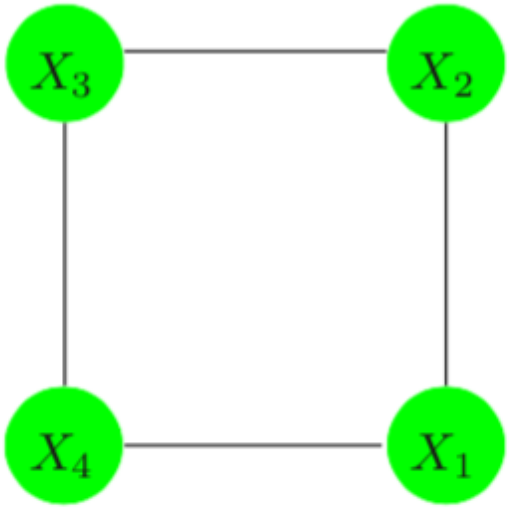
- We can use regression to solve for Θ and its inverse $W = \Theta^{-1}$ one row and column at a time.

Algorithm

Algorithm 17.1 *A Modified Regression Algorithm for Estimation of an Undirected Gaussian Graphical Model with Known Structure.*

1. Initialize $\mathbf{W} = \mathbf{S}$.
2. Repeat for $j = 1, 2, \ldots, p, 1, 2, \ldots, p, \ldots$ until convergence:
   (a) Partition the matrix $\mathbf{W}$ into part 1: all but the jth row and column, and part 2: the jth row and column.
   (b) Solve $\mathbf{W}_{11}^{*}\beta^{*} - s_{12}^{*} = 0$ for the unconstrained edge parameters β^{*}, using the reduced system of equations as in (17.19). Obtain $\hat{\beta}$ by padding $\hat{\beta}^{*}$ with zeros in the appropriate positions.
   (c) Update $w_{12} = \mathbf{W}_{11}\hat{\beta}$
3. In the final cycle (for each j) solve for $\hat{\theta}_{12} = -\hat{\beta} \cdot \hat{\theta}_{22}$, with $1/\hat{\theta}_{22} = s_{22} - w_{12}^{T}\hat{\beta}$.

Toy Example



$$\mathbf{S} = \begin{pmatrix} 10 & 1 & 5 & 4 \\ 1 & 10 & 2 & 6 \\ 5 & 2 & 10 & 3 \\ 4 & 6 & 3 & 10 \end{pmatrix}$$

FIGURE 17.4. *A simple graph for illustration, along with the empirical covariance matrix.*

Toy Example

$$\hat{\mathbf{\Sigma}} = \begin{pmatrix} 10.00 & 1.00 & 1.31 & 4.00 \\ 1.00 & 10.00 & 2.00 & 0.87 \\ 1.31 & 2.00 & 10.00 & 3.00 \\ 4.00 & 0.87 & 3.00 & 10.00 \end{pmatrix}, \quad \hat{\mathbf{\Sigma}}^{-1} = \begin{pmatrix} 0.12 & -0.01 & 0.00 & -0.05 \\ -0.01 & 0.11 & -0.02 & 0.00 \\ 0.00 & -0.02 & 0.11 & -0.03 \\ -0.05 & 0.00 & -0.03 & 0.13 \end{pmatrix}$$

Graphical Lasso

In most cases we do not know which edges to omit from our graph, and so would like to try to discover this from the data itself. A common method is L_1 (lasso) regularization, i.e.

$$\log\det(\Theta) - \mathit{trace}(S\Theta) - \lambda\,||\Theta||_1\,, \tag{9}$$

where $||\Theta||_1$ is the L_1 norm - the sum of absolute values of the elements of Σ^{-1}.

Graphical Lasso Algorithm

Algorithm 17.2 *Graphical Lasso.*

1. Initialize $\mathbf{W} = \mathbf{S} + \lambda\mathbf{I}$. The diagonal of $\mathbf{W}$ remains unchanged in what follows.

2. Repeat for $j = 1, 2, \ldots p, 1, 2, \ldots p, \ldots$ until convergence:

   (a) Partition the matrix $\mathbf{W}$ into part 1: all but the jth row and column, and part 2: the jth row and column.

   (b) Solve the estimating equations $\mathbf{W}_{11}\beta - s_{12} + \lambda \cdot \text{Sign}(\beta) = 0$ using the cyclical coordinate-descent algorithm (17.26) for the modified lasso.

   (c) Update $w_{12} = \mathbf{W}_{11}\hat{\beta}$

3. In the final cycle (for each j) solve for $\hat{\theta}_{12} = -\hat{\beta} \cdot \hat{\theta}_{22}$, with $1/\hat{\theta}_{22} = w_{22} - w_{12}^T\hat{\beta}$.

Illustration of Penalization Parameter