Penalized Maximum Likelihood Inference for Sparse Gaussian Graphical Models with Latent Structure

Christophe Ambroise, Julien Chiquet and Catherine Matias

Laboratoire Statistique et Génome,
La génopole - Université d’Évry

Journées MAS – 2008
Families of networks

- protein-protein interactions,
- regulation network,
- metabolic pathways.

Regulation example: SOS Network E. Coli

⇒ Let us focus on regulatory networks
Biological networks
Different kinds of biological interactions

Families of networks

- protein-protein interactions,
- regulation network,
- metabolic pathways.

Regulation example: SOS Network E. Coli

Let us focus on regulatory networks
What questions?

- How to find interactions?
- Given a network, what knowledge the structure can provide about the functions?
- Given two nodes, do they interact?
- Given a new node, what are the interaction with the known nodes?
- What are the characteristics of each community?
- Degree distribution
- Spectral clustering
- Stat. model
- Community analysis
- Structure
- Supervised
- Unsupervised
- Inference
Problem
Infer the interactions between genes from microarray data

Microarray gene expression data, \( p \) genes, \( n \) experiments

Which interaction?

Major Issues

▶ combinatorial: \( 2^p \) possible graphs
▶ dimension problem: \( n \ll p \)

Ambroise, Chiquet, Matias
Infer the interactions between genes from microarray data

**Problem**

- **Microarray gene expression data,** $p$ genes, $n$ experiments

**Inference**

- Which interaction?

**Major Issues**

- **combinatory:** $2^p^2$ possible graphs
- **dimension problem:** $n \ll p$
Our ideas to tackle these issues

Introduce prior taking the topology of the network into account for better edge inference

Intuition: use biological constraints

1. few genes effectively interact (sparsity),
2. networks are organized (latent structure).
Our ideas to tackle these issues

Introduce prior taking the topology of the network into account for better edge inference

Intuition: use biological constraints

1. few genes effectively interact (sparsity),
2. networks are organized (latent structure).
Our ideas to tackle these issues

Introduce prior taking the topology of the network into account for better edge inference

Intuition: use biological constraints

1. few genes effectively interact (sparsity),
2. networks are organized (latent structure).
Give the network a model
  Gaussian graphical models
  Providing the network with a latent structure
  The complete likelihood

Inference strategy by alternate optimization
  The E-step: estimation of the latent structure
  The M-step: inferring the connectivity matrix

Numerical Experiments
Outline

Give the network a model
- Gaussian graphical models
- Providing the network with a latent structure
- The complete likelihood

Inference strategy by alternate optimization
- The E-step: estimation of the latent structure
- The M-step: inferring the connectivity matrix

Numerical Experiments
Outline

Give the network a model
  Gaussian graphical models
  Providing the network with a latent structure
  The complete likelihood

Inference strategy by alternate optimization
  The $E$–step: estimation of the latent structure
  The $M$–step: inferring the connectivity matrix

Numerical Experiments
Outline

Give the network a model
  Gaussian graphical models
    Providing the network with a latent structure
    The complete likelihood

Inference strategy by alternate optimization
  The E–step: estimation of the latent structure
  The M–step: inferring the connectivity matrix

Numerical Experiments
The Gaussian model

- Let $X \in \mathbb{R}^p$ be a random vector such as $X \sim \mathcal{N}(0_p, \Sigma)$;
- let $(X^1, \ldots, X^n)$ be an i.i.d. $n$–sample representing the microarray experiments;
- let $X$ be a $n \times p$ matrix such as $(X^k)^\top$ is the $k$th row of $X$;
- let $K = (K_{ij})_{(i,j) \in \mathcal{P}^2} := \Sigma^{-1}$ be the concentration matrix.

The graphical interpretation

\[
X_i \perp \! \! \! \perp X_j | X_{\mathcal{P}\setminus\{i,j\}} \iff \text{edge } (i, j) \notin \text{ network } \iff K_{ij} = 0,
\]

since
\[
r_{ij|\mathcal{P}\setminus\{i,j\}} = -K_{ij}/\sqrt{K_{ii}K_{jj}}.
\]

$K$ describes the graph of conditional dependencies.
The Gaussian model

- Let $X \in \mathbb{R}^p$ be a random vector such as $X \sim \mathcal{N}(0_p, \Sigma)$;
- let $(X^1, \ldots, X^n)$ be an i.i.d. $n$–sample representing the microarray experiments;
- let $X$ be a $n \times p$ matrix such as $(X^k)^\top$ is the $k$th row of $X$;
- let $K = (K_{ij})_{(i,j) \in \mathcal{P}^2} := \Sigma^{-1}$ be the concentration matrix.

The graphical interpretation

$$X_i \perp \perp X_j | X_{\mathcal{P}\backslash \{i,j\}} \Leftrightarrow \text{edge } (i, j) \notin \text{ network } \Leftrightarrow K_{ij} = 0,$$

since $r_{ij}|\mathcal{P}\backslash \{i,j\} = -K_{ij} / \sqrt{K_{ii}K_{jj}}$.

$\leadsto K$ describes the graph of conditional dependencies.
One may use $p$ different linear regressions

$$X_i = (X_{\setminus i})^T \alpha + \varepsilon,$$

where $\alpha_j = -K_{ij}/K_{ii}$,

yet $n \ll p$: the empirical covariance $S$ is singular.

Meinshausen and Bühlman’s approach (06)

Solve $p$ independent Lasso problems:

$$\hat{\alpha} = \arg \min_{\alpha} \frac{1}{n} \| X_i - X_{\setminus i} \alpha \|_2^2 + \rho \| \alpha \|_{\ell_1},$$

where $X_i$ is the $i$th column of $X$, and $X_{\setminus i}$ is the full matrix with $i$th column removed.

Major drawback: need of a symmetrization step to obtain a final estimate of $K$. 
GGMs and regression
Network inference as $p$ independent regression problems

One may use $p$ different linear regressions

$$X_i = (X_{\setminus i})^T \alpha + \varepsilon, \quad \text{where } \alpha_j = -K_{ij}/K_{ii},$$

yet $n \ll p$: the empirical covariance $S$ is singular.

**Meinshausen and Bühlman’s approach (06)**
Solve $p$ independent Lasso problems:

$$\hat{\alpha} = \arg \min_{\alpha} \frac{1}{n} \left\| X_i - X_{\setminus i} \alpha \right\|_2^2 + \rho \left\| \alpha \right\|_{\ell_1},$$

where $X_i$ is the $i$th column of $X$, and $X_{\setminus i}$ is the full matrix with $i$th column removed.

**Major drawback**: need of a symmetrization step to obtain a final estimate of $K$.  

Ambroise, Chiquet, Matias
Consider the approximation $P(X) = \prod_{i=1}^{p} P(X_i|X_{\backslash i})$.

**Proposition**

$$\hat{K} = \arg \max_{K} \log \tilde{L}(X; K) + \rho \|K\|_{\ell_1},$$

(1)

with

$$\tilde{L}(X; K) = \sum_{i=1}^{p} \left( \sum_{k=1}^{n} \log P(X_i^k|X_{\backslash i}^k; K_i) \right),$$

give the same solution as the $p$ independent penalized regressions.

$\implies$ Those $p$ terms are not independent, as $K$ is non diagonal!
Consider the approximation $\mathbb{P}(X) = \prod_{i=1}^{p} \mathbb{P}(X_i | X_{\backslash i})$.

**Proposition**

$$\hat{K} = \arg \max_K \log \tilde{\mathcal{L}}(X; K) + \rho \|K\|_{\ell_1},$$

with

$$\tilde{\mathcal{L}}(X; K) = \sum_{i=1}^{p} \left( \sum_{k=1}^{n} \log \mathbb{P}(X^k_i | X_{\backslash i}^k; K_i) \right),$$

**give the same solution as the $p$ independent penalized regressions.**

\[\Rightarrow \text{Those $p$ terms are not independent, as $K$ is non diagonal!}\]
GGMs and penalized likelihood

The penalized likelihood of the Gaussian observations

Use a penalty term

$$\frac{n}{2} \left( \log \det(K) - \text{Tr}(SK) \right) - \rho \|K\|_1,$$

where $S$ is the empirical covariance matrix.

GGMs and penalized likelihood

The penalized likelihood of the Gaussian observations
Use a penalty term

$$\frac{n}{2} \left( \log \det(K) - \text{Tr}(SK) \right) - \rho \|K\|_{\ell_1},$$

where $S$ is the empirical covariance matrix.

Natural generalization
Use different penalty parameters for different coefficients

$$\frac{n}{2} \left( \log \det(K) - \text{Tr}(SK) \right) - \|\rho_Z(K)\|_{\ell_1},$$

where $\rho_Z(K) = (\rho_{Z_i,Z_j}(K_{i,j}))_{i,j}$ is a penalty function depending on an unknown underlying structure $Z$. 
Outline

Give the network a model
  Gaussian graphical models
  Providing the network with a latent structure
  The complete likelihood

Inference strategy by alternate optimization
  The E-step: estimation of the latent structure
  The M-step: inferring the connectivity matrix

Numerical Experiments
Assumption: there exists a latent structure spreading the vertices into a set \( Q = \{1, \ldots, q, \ldots, Q\} \) of classes of connectivity.

The classes of connectivity
Denote \( Z = \{Z_i = (Z_{i1}, \ldots, Z_{iQ})\}_i \) where \( Z_{iq} = \mathbb{1}_{\{i \in q\}} \) are the latent independent variables, with

- \( \alpha = \{\alpha_q\} \), the prior proportions of groups,
- \( (Z_i) \sim M(1, \alpha) \), a multinomial distribution.

A mixture of Laplace distributions
Assume \( K_{ij} | Z \) independent. Then \( K_{ij} | \{Z_{iq} Z_{j\ell} = 1\} \sim f_{q\ell}(\cdot) \), where

\[
f_{q\ell}(x) = \frac{1}{2\lambda_{q\ell}} \exp \left\{-\frac{|x|}{\lambda_{q\ell}}\right\}, \quad q, \ell \in Q.
\]
Assumption: there exists a latent structure spreading the vertices into a set \( Q = \{1, \ldots, q, \ldots, Q\} \) of classes of connectivity.

The classes of connectivity

Denote \( Z = \{Z_i = (Z_{i1}, \ldots, Z_{iQ})\}_i \) where \( Z_{iq} = 1_{\{i \in q\}} \) are the latent independent variables, with

- \( \alpha = \{\alpha_q\}, \) the prior proportions of groups,
- \( (Z_i) \sim M(1, \alpha), \) a multinomial distribution.

A mixture of Laplace distributions

Assume \( K_{ij} | Z \) independent. Then \( K_{ij} | \{Z_{iq}Z_{j\ell} = 1\} \sim f_{q\ell}(\cdot), \) where

\[
f_{q\ell}(x) = \frac{1}{2\lambda_{q\ell}} \exp \left\{ -\frac{|x|}{\lambda_{q\ell}} \right\}, \quad q, \ell \in Q.
\]
Assumption: there exists a latent structure spreading the vertices into a set $Q = \{1, \ldots, q, \ldots, Q\}$ of classes of connectivity.

The classes of connectivity

Denote $Z = \{Z_i = (Z_{i1}, \ldots, Z_{iQ})\}_i$ where $Z_{iq} = 1_{\{i \in q\}}$ are the latent independent variables, with

- $\alpha = \{\alpha_q\}$, the prior proportions of groups,
- $(Z_i) \sim M(1, \alpha)$, a multinomial distribution.

A mixture of Laplace distributions

Assume $K_{ij} | Z$ independent. Then $K_{ij} | \{Z_{iq} Z_{j\ell} = 1\} \sim f_{q\ell}(\cdot)$, where

$$f_{q\ell}(x) = \frac{1}{2\lambda_{q\ell}} \exp \left\{ -\frac{|x|}{\lambda_{q\ell}} \right\}, \quad q, \ell \in Q.$$
Some possible structures

Figure: From Affiliation to Bipartite

Example
Modular (affiliation) network
Two kinds of Laplace distributions
1. intra-cluster $q = \ell, f_\Lambda (\cdot; \lambda_\Lambda)$;
2. inter-cluster $q \neq \ell, f_\varepsilon (\cdot; \lambda_\varepsilon)$. 

Ambroise, Chiquet, Matias
Some possible structures

Figure: From Affiliation to Bipartite

Example

Modular (affiliation) network
Two kinds of Laplace distributions

1. \( \text{intra-cluster} \ q = \ell, f_\Lambda(\cdot; \lambda_\Lambda) \);
2. \( \text{inter-cluster} \ q \neq \ell, f_\varepsilon(\cdot; \lambda_\varepsilon) \).
Give the network a model
  Gaussian graphical models
  Providing the network with a latent structure
  The complete likelihood

Inference strategy by alternate optimization
  The E-step: estimation of the latent structure
  The M-step: inferring the connectivity matrix

Numerical Experiments
We wish to infer non-null entries of $K$ knowing the data. Then our strategy is

$$\hat{K} = \arg \max_{K \succ 0} \mathbb{P}(K|X) = \arg \max_{K \succ 0} \log \mathbb{P}(X, K).$$

Marginalization over $Z$

Because distribution of $K$ is known conditional on the structure $X$,

$$\hat{K} = \arg \max_{K \succ 0} \log \sum_{Z \in Z} \mathcal{L}_c(X, K, Z),$$

where $\mathcal{L}_c(X, K, Z) = \mathbb{P}(X, K, Z)$ is the so-called complete-data classification likelihood.

$\Rightarrow$ An EM–like strategy is used hereafter to solve this problem.
We wish to infer non-null entries of $K$ knowing the data. Then our strategy is

$$
\hat{K} = \arg \max_{K > 0} \mathbb{P}(K|X) = \arg \max_{K > 0} \log \mathbb{P}(X, K).
$$

### Marginalization over $Z$

Because distribution of $K$ is known conditional on the structure $\hat{K} = \arg \max_{K > 0} \log \sum_{Z \in Z} \mathcal{L}_c(X, K, Z)$,

where $\mathcal{L}_c(X, K, Z) = \mathbb{P}(X, K, Z)$ is the so-called complete-data classification likelihood.

An EM-like strategy is used hereafter to solve this problem.
We wish to infer non-null entries of $K$ knowing the data. Then our strategy is

$$\hat{K} = \arg \max_{K>0} \mathbb{P}(K|X) = \arg \max_{K>0} \log \mathbb{P}(X, K).$$

**Marginalization over $Z$**

Because distribution of $K$ is known conditional on the structure!

$$\hat{K} = \arg \max_{K>0} \log \sum_{Z \in Z} \mathcal{L}_c(X, K, Z),$$

where $\mathcal{L}_c(X, K, Z) = \mathbb{P}(X, K, Z)$ is the so-called complete-data classification likelihood.

$\mapsto$ **An EM–like strategy** is used hereafter to solve this problem.
The complete likelihood

**Proposition**

\[
\log \mathcal{L}_c(X, K, Z) \propto \frac{n}{2} \left( \log \det(K) - \text{Tr}(SK) \right) - \left\| \rho_Z(K) \right\|_{\ell_1}
- \sum_{i,j \in \mathcal{P}, i \neq j} Z_{iq}Z_{j\ell} \log(2\lambda_{q\ell}) + \sum_{i \in \mathcal{P}, q \in \mathcal{Q}} Z_{iq} \log \alpha_q,
\]

(2)

where \( S \) is the empirical covariance matrix and \( \rho_Z(K) = (\rho_{Z_iZ_j}(K_{ij}))_{(i,j) \in \mathcal{P}^2} \) is defined by

\[
\rho_{Z_iZ_j}(K_{ij}) = \sum_{q,\ell \in \mathcal{Q}} Z_{iq} Z_{j\ell} \frac{(K_{ij})}{\lambda_{q\ell}}.
\]

(3)
The complete likelihood

Proposition

\[ \log \mathcal{L}_c(\mathbf{X}, \mathbf{K}, \mathbf{Z}) \propto \frac{n}{2} \left( \log \det(\mathbf{K}) - \text{Tr}(\mathbf{S}\mathbf{K}) \right) - \| \rho_{\mathbf{Z}}(\mathbf{K}) \|_{\ell_1} \]

\[ - \sum_{i,j \in \mathcal{P}, i \neq j} Z_{i,q} Z_{j,\ell} \log(2\lambda_{q,\ell}) + \sum_{i \in \mathcal{P}, q \in \mathcal{Q}} Z_{i,q} \log \alpha_q, \quad (2) \]

where \( S \) is the empirical covariance matrix and \( \rho_{\mathbf{Z}}(\mathbf{K}) = (\rho_{\mathbf{Z}_i \mathbf{Z}_j}(K_{i,j}))_{(i,j) \in \mathcal{P}^2} \) is defined by

\[ \rho_{\mathbf{Z}_i \mathbf{Z}_j}(K_{i,j}) = \sum_{q, \ell \in \mathcal{Q}} Z_{i,q} Z_{j,\ell} \frac{(K_{i,j})}{\lambda_{q,\ell}}. \quad (3) \]

Part concerning \( \mathbf{K} \): PML with a LASSO-type approach.
The complete likelihood

Proposition

\[ \log L_c(X, K, Z) \propto \frac{n}{2} \left( \log \det(K) - \text{Tr}(SK) \right) - \| \rho_Z(K) \|_{l_1} - \sum_{i,j \in \mathcal{P}, i \neq j} Z_{iq} Z_{j\ell} \log(2\lambda_{q\ell}) + \sum_{i \in \mathcal{P}, q \in \mathcal{Q}} Z_{iq} \log \alpha_q, \quad (2) \]

where \( S \) is the empirical covariance matrix and \( \rho_Z(K) = \left( \rho_{Z_i Z_j(K_{ij})} \right)_{(i,j) \in \mathcal{P}^2} \) is defined by

\[ \rho_{Z_i Z_j}(K_{ij}) = \sum_{q, \ell \in \mathcal{Q}} Z_{iq} Z_{j\ell} \frac{K_{ij}}{\lambda_{q\ell}}. \quad (3) \]

Part concerning \( Z \): estimation with a variational \( \text{EM} \) approach.
Plan

Give the network a model
  Gaussian graphical models
  Providing the network with a latent structure
  The complete likelihood

Inference strategy by alternate optimization
  The E-step: estimation of the latent structure
  The M-step: inferring the connectivity matrix

Numerical Experiments
An EM strategy

The conditional expectation to maximize

\[ Q \left( K \mid K^{(m)} \right) = \mathbb{E} \left\{ \log \mathcal{L}_c(X, K, Z) \mid X; K^{(m)} \right\} \]

\[ := \sum_{Z \in \mathcal{Z}} P \left( Z \mid X, K^{(m)} \right) \log \mathcal{L}_c(X, K, Z) \]

\[ = \sum_{Z \in \mathcal{Z}} P \left( Z \mid K^{(m)} \right) \log \mathcal{L}_c(X, K, Z). \]

Problem

- No closed-form of \( Q \left( K \mid K^{(m)} \right) \) because \( P(Z \mid K) \) cannot be factorized.
- We use variational approach to approximate \( P(Z \mid K) \).
An EM strategy

The conditional expectation to maximize

\[
Q\left(K|K^{(m)}\right) = \mathbb{E}\left\{ \log \mathcal{L}_c(X, K, Z)|X; K^{(m)} \right\} \\
:= \sum_{Z \in \mathcal{Z}} \mathbb{P}\left(Z|X, K^{(m)}\right) \log \mathcal{L}_c(X, K, Z) \\
= \sum_{Z \in \mathcal{Z}} \mathbb{P}\left(Z|K^{(m)}\right) \log \mathcal{L}_c(X, K, Z).
\]

Problem

▶ No closed-form of \(Q\left(K|K^{(m)}\right)\) because \(\mathbb{P}(Z|K)\) cannot be factorized.

▶ We use variational approach to approximate \(\mathbb{P}(Z|K)\).
Outline

Give the network a model
  Gaussian graphical models
  Providing the network with a latent structure
  The complete likelihood

Inference strategy by alternate optimization
  The E–step: estimation of the latent structure
  The M–step: inferring the connectivity matrix

Numerical Experiments
Principle

Use an approximation $R(Z)$ of $\mathbb{P}(Z|K)$ in the factorized form, $R_{\tau}(Z) = \prod_i R_{\tau_i}(Z_i)$ where $R_{\tau_i}$ is a multinomial distribution with parameters $\tau_i$.

- Maximize a lower bound of the log-likelihood

\[
\mathcal{J}(R_{\tau}(Z)) = \mathcal{L}(X, K) - D_{KL}(R_{\tau}(Z)\|\mathbb{P}(Z|K)).
\]

- Using its tractable form, we have

\[
\mathcal{J}(R_{\tau}(Z)) = \sum_Z R_{\tau}(Z)\mathcal{L}_c(X, K, Z) + \mathcal{H}(R_{\tau}(Z)).
\]
Principle

Use an approximation \( R(Z) \) of \( P(Z|K) \) in the factorized form, \( R_\tau(Z) = \prod_i R_{\tau_i}(Z_i) \) where \( R_{\tau_i} \) is a multinominal distribution with parameters \( \tau_i \).

- Maximize a lower bound of the log-likelihood

\[
\mathcal{J}(R_\tau(Z)) = \mathcal{L}(X, K) - D_{KL}(R_\tau(Z)\|P(Z|K)).
\]

- Using its tractable form, we have

\[
\mathcal{J}(R_\tau(Z)) = \sum_Z R_{\tau}(Z) \mathcal{L}_c(X, K, Z) + \mathcal{H}(R_\tau(Z)).
\]

This term plays the role of \( \mathbb{E}(\mathcal{L}_c(X, K, Z)|X, K^{(m)}) \)
Outline

- Give the network a model
  - Gaussian graphical models
  - Providing the network with a latent structure
  - The complete likelihood

Inference strategy by alternate optimization
- The E-step: estimation of the latent structure
- The M-step: inferring the connectivity matrix

Numerical Experiments
The M–step
Seen as a penalized likelihood problem

We aim at solving

$$\hat{K} = \arg \max_{K \succ 0} \hat{Q}_{\tau}(K),$$

where

**Penalized likelihood problem**

$$\hat{Q}_{\tau}(K) = \left\{ \frac{n}{2} \left( \log \det(K) - \operatorname{Tr}(SK) \right) - \| \rho_{\tau}(K) \|_{\ell_1} + C \right\},$$

---


⇝ We deal with a more complex penalty term here.
The M–step
Seen as a penalized likelihood problem

We aim at solving

$$\hat{K} = \arg \max_{K \succ 0} \hat{Q}_\tau(K),$$

where

Penalized likelihood problem

$$\hat{Q}_\tau(K) = \left\{ \frac{n}{2} (\log \det(K) - \text{Tr}(SK)) - \| \rho_\tau(K) \|_{\ell_1} + Cst \right\},$$


We deal with a more complex penalty term here.
Let us work on the covariance matrix

**Proposition**

The maximization problem over $\mathbf{K}$ is equivalent to the following, dealing with the covariance matrix $\Sigma$:

$$W := \hat{\Sigma} = \arg \max \log \det(\Sigma), \quad \|\Sigma - S\|_{\infty} / \mathbf{P} \leq 1$$

where $\cdot /$ is the term-by-term division and

$$\mathbf{P} = (p_{ij})_{i,j \in \mathcal{P}} = \frac{2}{n} \sum_{q,\ell} \frac{\tau_{iq} \tau_{j\ell}}{\lambda_{q\ell}}.$$

$\sim\sim$ The proof uses some optimization, primal/dual tricks
Denote

\[ W = \begin{bmatrix} W_{11} & w_{12} \\ w_{12}^\top & w_{22} \end{bmatrix}, \quad S = \begin{bmatrix} S_{11} & s_{12} \\ s_{12}^\top & s_{22} \end{bmatrix}, \quad P = \begin{bmatrix} P_{11} & p_{12} \\ p_{12}^\top & p_{22} \end{bmatrix}, \quad (4) \]

where \( W_{11} \) is a \((p - 1) \times (p - 1)\) matrix, \( w_{12} \) is a \( p - 1 \) length column vector and \( w_{22} \) is a scalar.

Each column of \( W \) satisfies (through Schürr complement)

\[ w_{12} = \arg \min \left\{ y^\top W_{11}^{-1} y \right\}, \quad \left\{ y : \| (y-s_{12})/p_{12} \|_\infty \leq 1 \right\} \]
Proposition

Solving the block-wise problem is equivalent to solve the following dual problem

$$\min_{\beta} \left\| \frac{1}{2} W_{11}^{1/2} \beta - W_{11}^{-1/2} s_{12} \right\|_2^2 + \| p_{12} \ast \beta \|_{\ell_1},$$

where $w_{12}$ and $\beta$ are linked through

$$w_{12} = W_{11} \beta / 2.$$

$\Rightarrow$ At last, this is a LASSO-like formulation!
The full EM algorithm

// Initialization
Compute $W^{(0)}$ by maximizing the penalized likelihood with uniform penalization
Compute $\hat{K}^{(0)}$ by smart inversion of $W^{(0)}$
$m \leftarrow 0$

// The main loop
while $\hat{Q}_\tau(\hat{K}^{(m)})$ is not stabilized or $m = 0$ do
  Compute $\tau$ with fixed-point algorithm, initialized with spectral clustering
  Compute $Q_\tau(\hat{K}^{(m)})$
  while $W^{(m)}$ is not stabilized do
    for each column of $W^{(m)}$ do
      Compute $w_{ij}$ by solving the LASSO-like problem with path-wise coordinate optimization
    end
  end
  Compute $\hat{K}^{(m)}$ by smart inversion of $W^{(m)}$
  $m \leftarrow m + 1$
end
The full EM algorithm

// Initialization
Compute $W^{(0)}$ by maximizing the penalized likelihood with uniform penalization
Compute $\hat{K}^{(0)}$ by smart inversion of $W^{(0)}$
$m \leftarrow 0$

// The main loop
while $\hat{Q}_\tau(\hat{K}^{(m)})$ is not stabilized or $m = 0$ do
    Compute $\tau$ with fixed-point algorithm, initialized with spectral clustering
    Compute $Q_\tau(\hat{K}^{(m)})$
    while $W^{(m)}$ is not stabilized do
        for each column of $W^{(m)}$ do
            Compute $w_{12}$ by solving the LASSO-like problem with path-wise coordinate optimization
        end
    end
    Compute $\hat{K}^{(m+1)}$ by smart inversion of $W^{(m)}$
    $m \leftarrow m + 1$
The full EM algorithm

// Initialization
Compute $W^{(0)}$ by maximizing the penalized likelihood with uniform penalization
Compute $\hat{K}^{(0)}$ by smart inversion of $W^{(0)}$
$m \leftarrow 0$

// The main loop
while $\hat{Q}_\tau (\hat{K}^{(m)})$ is not stabilized or $m = 0$ do

  // The E-Step: latent structure inference
  Compute $\hat{\tau}$ with fixed-point algorithm, initialized with spectral clustering
  Compute $\hat{Q}_\tau (\hat{K}^{(m)})$

  // The M-Step: network inference
  while $W^{(m)}$ is not stabilized do
    for each column of $W^{(m)}$ do
      Compute $w_{12}$ by solving the LASSO-like problem with path-wise coordinate optimization
    end
  end
  Compute $\hat{K}^{(m)}$ by smart inversion of $W^{(m)}$

  $m \leftarrow m + 1$
end
The full EM algorithm

// Initialization
Compute $W^{(0)}$ by maximizing the penalized likelihood with uniform penalization
Compute $\hat{K}^{(0)}$ by smart inversion of $W^{(0)}$
$m \leftarrow 0$

// The main loop
while $\hat{Q}_\tau (\hat{K}^{(m)})$ is not stabilized or $m = 0$ do

    // The E-Step: latent structure inference
    Compute $\hat{\tau}$ with fixed-point algorithm, initialized with spectral clustering
    Compute $\hat{Q}_\tau (\hat{K}^{(m)})$

    // The M-Step: network inference
    while $W^{(m)}$ is not stabilized do
        for each column of $W^{(m)}$ do
            Compute $w_{12}$ by solving the LASSO-like problem with path-wise coordinate optimization
        end
    Compute $\hat{K}^{(m)}$ by smart inversion of $W^{(m)}$

    $m \leftarrow m + 1$
end
The full EM algorithm

// Initialization
Compute $W^{(0)}$ by maximizing the penalized likelihood with uniform penalization
Compute $\hat{K}^{(0)}$ by smart inversion of $W^{(0)}$
$m \leftarrow 0$

// The main loop
while $\hat{Q}_\tau (\hat{K}^{(m)})$ is not stabilized or $m = 0$ do
    // The E-Step: latent structure inference
    Compute $\hat{\Phi}$ with fixed-point algorithm, initialized with spectral clustering
    Compute $\hat{Q}_\tau (\hat{K}^{(m)})$
    // The M-Step: network inference
    while $W^{(m)}$ is not stabilized do
        for each column of $W^{(m)}$ do
            Compute $w_{12}$ by solving the LASSO–like problem with path-wise coordinate optimization
        end
    end
    Compute $\hat{K}^{(m)}$ by smart inversion of $W^{(m)}$
$m \leftarrow m + 1$
end
Plan

Give the network a model
Gaussian graphical models
Providing the network with a latent structure
The complete likelihood

Inference strategy by alternate optimization
The E-step: estimation of the latent structure
The M-step: inferring the connectivity matrix

Numerical Experiments
Simulations settings

Four inference methods

1. **GeneNet** (*Strimmer et al.*)
   Edge estimation based on partial correlation with shrinkage.

2. **GLasso** (*Friedman et al.*)
   Edge estimation uses a uniform penalty matrix.

3. **“perfect” SIMoNe** (*best results our method can aspire to*)
   Edge estimation uses a penalty matrix constructed according to the theoretic node classification.

4. **SIMoNe** (*Statistical Inference for MOdular NEtworks*)
   Edge estimation uses a penalty matrix constructed according to the estimated node classification, iteratively.
Simulated Graphs

- Graphs simulated using an affiliation model (two sets of parameters: intra-groups and inter-groups connections)
- \( p = 300 \) nodes \( \frac{p(p - 1)}{2} = 44850 \) possible interactions.
- 10 graphs (repetitions) were simulated per situation.
- Gene expression data was then simulated using the sampled graph:
  1. Favorable setting \( (n = 4 \times p) \),
  2. Middle case \( (n = p) \)
  3. Unfavorable setting \( (n = p/3) \)
Example of graph recovery
Favorable case

Figure: Theoretical graph and SIvMoNe estimation
Example of graph recovery
Favorable case

simone graph estimation : iter 4

Figure: Theoretical graph and SIMoNe estimation
GeneNet versus GLasso

- GeneNet rather similar to GLasso with a low penalty value.
- Increasing the penalty value flattens the FDR curves from right to left.

**Figure:** GeneNet and GLasso
GeneNet versus GLasso

- GeneNet rather similar to GLasso with a low penalty value.
- Increasing the penalty value flattens the FDR curves from right to left.

**Figure:** GeneNet and GLasso
When no structure SIMoNe is comparable to GeneNet

Figure: Monoclass graph with usual connectivity, low dimensional setting
FDR Curves
Perfect SIMoNe versus GLasso

- low penalty: perfect SIMoNe and GLasso are equivalent
- higher penalty: perfect SIMoNe outperforms normal GLasso
- SIMoNe recovers class information

Figure: GLasso, Perfect SIMoNe
- low penalty: perfect SIMoNe and GLasso are equivalent
- higher penalty: perfect SIMoNe outperforms normal GLasso
- SIMoNe recovers class information

**Figure:** GLasso, Perfect SIMoNe
FDR Curves
Perfect SiMoNe versus GLasso

- Low penalty: perfect SiMoNe and GLasso are equivalent
- Higher penalty: perfect SiMoNe outperforms normal GLasso
- SiMoNe recovers class information

**Figure:** GLasso, Perfect SiMoNe
FDR Curves
Perfect SiMoNe versus GLasso versus SiMoNe

- low penalty: perfect SiMoNe and GLasso are equivalent
- higher penalty: perfect SiMoNe outperforms normal GLasso
- SiMoNe recovers class information

Figure: GLasso, Perfect SiMoNe, SiMoNe
Two types of patients

1. Patient response can be classified as either a pathologic complete response (PCR)
2. or residual disease (Not PCR).

Gene expression data

- 133 patients (99 not PCR, 34 PCR)
- 26 identified genes (differential analysis)
First result on real a dataset
Prediction of the outcome of preoperative chemotherapy

Not PCR
First result on real a dataset
Prediction of the outcome of preoperative chemotherapy
First result on real a dataset
Prediction of the outcome of preoperative chemotherapy
Conclusions

To sum-up

- We proposed an inference strategy based on a penalization scheme given by an underlying unknown structure.
- The estimation strategy is based on a variational EM algorithm, in which a LASSO-like procedure is embedded.
- This is still a work in progress, but preliminary results are very encouraging.
- R package SIMoNe

Perspectives

- Consider alternative prior more biologically relevant: hubs, motifs.
- Time segmentation when dealing with temporal data
Conclusions

To sum-up

▶ We proposed an inference strategy based on a penalization scheme given by an underlying unknown structure.
▶ The estimation strategy is based on a variational EM algorithm, in which a LASSO-like procedure is embedded.
▶ This is still a work in progress, but preliminary results are very encouraging.
▶ R package SIMoNe

Perspectives

▶ Consider alternative prior more biologically relevant: hubs, motifs.
▶ Time segmentation when dealing with temporal data