

Graphical model inference with unobserved variables via latent tree aggregation

Geneviève Robin Christophe Ambroise Stéphane Robin

UMR 518 AgroParisTech/INRA MIA
Équipe Statistique et génome

NETBIO, November 9th 2017



- 1 Introduction
- 2 Gaussian graphical models
- 3 GGM with missing variables
- 4 EM with aggregation of spanning trees
- 5 Experiments - simulations and flow cytometry data

- 1 Introduction
- 2 Gaussian graphical models
- 3 GGM with missing variables
- 4 EM with aggregation of spanning trees
- 5 Experiments - simulations and flow cytometry data

Biological networks

Different kinds of biological interactions

Families of networks

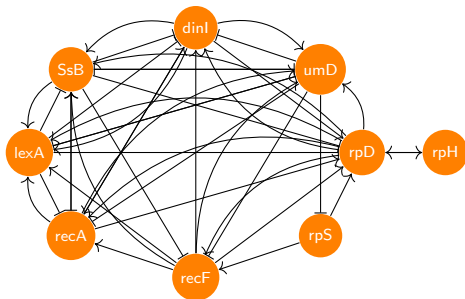
- protein-protein interactions,
- metabolic pathways,
- regulation network,
- ...

Biological networks

Different kinds of biological interactions

Families of networks

- regulation network,
- ...



Regulation example : *SOS Network E. Coli*

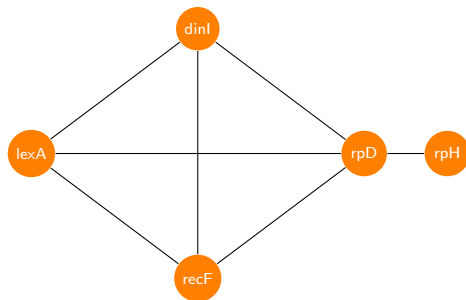
⇒ Let us focus on regulatory networks

Biological networks

Different kinds of biological interactions

Families of networks

- regulation network,
- ...



Regulation example : *SOS Network E. Coli*

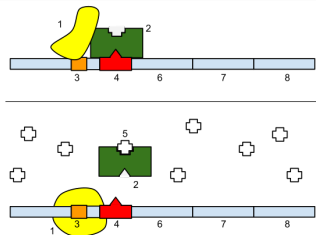
↪ Let us focus on regulatory networks ... and look for *influence* network

Regulation

Gene expression is regulated (inhibited or activated)

- by region (i.e., brain vs liver)
- by development stage (i.e. fetal vs. adult)
- by dynamic response to environment
- by gene status (i.e. mutant vs. wild)

Lactose Operon,
Nobel price, Jacob,
Monod et Lwoff
(1965)

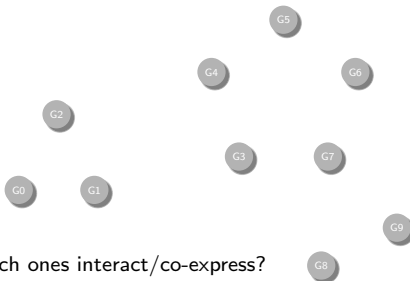


Problem

Infer the interactions between genes from microarray data



Microarray gene expression data,
 p genes, n experiments



Which ones interact/co-express?

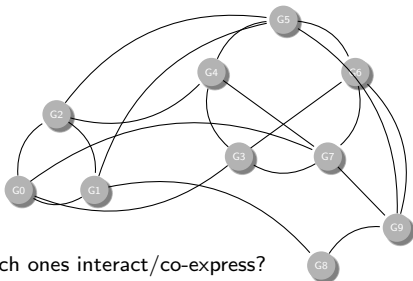
Problem

Infer the interactions between genes from microarray data



Inference

Microarray gene expression data,
 p genes, n experiments



Which ones interact/co-express?

Major Issues

- combinatorial: $2^{\frac{p(p-1)}{2}}$ possible graphs
- dimension problem: $n \ll p$ reduced to $n \approx p$

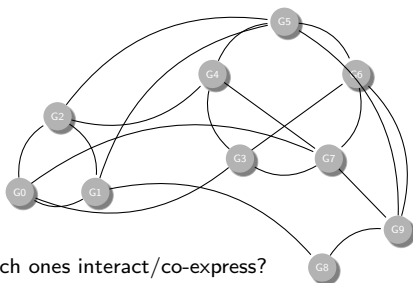
Problem

Infer the interactions between genes from microarray data



Inference

Microarray gene expression data,
 p genes, n experiments



Which ones interact/co-express?

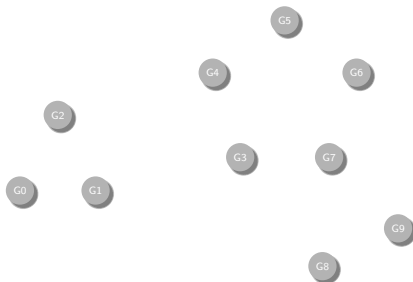
Major Issues

- combinatorial: $2^{\frac{p(p-1)}{2}}$ possible graphs
- dimension problem: $n \ll p$ reduced to $n \approx p$

Here, we **reduce** p to a number of **fixed** genes of interest

Our ideas to tackle these issues

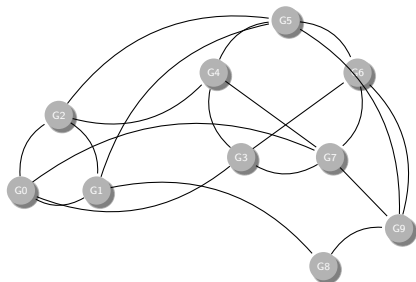
↪ Introduce prior taking the topology of the network into account for better edge inference



Relying on biological constraints

Our ideas to tackle these issues

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

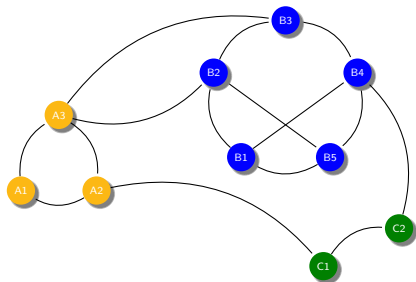


Relying on biological constraints

- 1 few genes effectively interact (**sparsity**),

Our ideas to tackle these issues

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



Relying on biological constraints

- 1 few genes effectively interact (**sparsity**),
- 2 networks are organized (**latent structure** or **Missing variables**).

Outline

Outline

- 1 Introduction
- 2 Gaussian graphical models**
- 3 GGM with missing variables
- 4 EM with aggregation of spanning trees
- 5 Experiments - simulations and flow cytometry data

Gaussian graphical models

General settings

The Gaussian model

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

Gaussian graphical models

General settings

The Gaussian model

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

The graphical interpretation

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

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

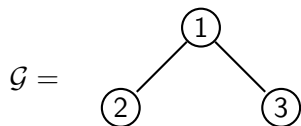
\rightsquigarrow \mathbf{K} describes the graph of **conditional dependencies**.

Gaussian graphical models

Example

$$X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 2 & 1 & -1 \\ 1 & 1.5 & -0.5 \\ -1 & -0.5 & 1.5 \end{pmatrix}$$

$$K = \Sigma^{-1} = \begin{pmatrix} 1 & -0.5 & 0.5 \\ -0.5 & 1 & 0 \\ 0.5 & 0 & 1 \end{pmatrix},$$



- Underlying graph $\mathcal{G} = (V, E)$, $V = \{1, \dots, p\}$
- The edge $\{i, j\}$ is in E if $K_{ij} \neq 0$

Inferring $\mathcal{G} \Leftrightarrow$ inferring the support of K .

Estimate K from data

- Maximum likelihood estimator:

$$\begin{aligned}\hat{K}^{MLE} &= \arg \max_K \log \det(K) - \text{tr}(K\Sigma_n) \\ &= \Sigma_n^{-1}\end{aligned}\tag{1}$$

Hypothesis on the structure of the support of K

- Penalized Log-likelihood
- Tree hypothesis

GGMs and regression

Network inference as p independent regression problems

One may use p different **linear regressions**

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

GGMs and regression

Network inference as p independent regression problems

One may use p different **linear regressions**

$$\mathbf{X}_i = (\mathbf{X}_{\setminus i})^\top \alpha + \varepsilon, \quad \text{where } \alpha_j = -K_{ij}/K_{ii},$$

Meinshausen and Bühlman's approach (06)

Solve p independent Lasso problems (**ℓ_1 -norm** enforces sparsity):

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

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

GGMs and regression

Network inference as p independent regression problems

One may use p different **linear regressions**

$$\mathbf{X}_i = (\mathbf{X}_{\setminus i})^\top \alpha + \varepsilon, \quad \text{where } \alpha_j = -K_{ij}/K_{ii},$$

Meinshausen and Bühlman's approach (06)

Solve p independent Lasso problems (**ℓ_1 -norm** enforces sparsity):

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

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

Major drawback: need of a symmetrization step to obtain a final estimate of \mathbf{K} .

GGMs and Lasso

Solving p penalized regressions \Leftrightarrow maximize the penalized pseudo-likelihood

Consider the approximation $\mathbb{P}(\mathbf{X}) = \prod_{i=1}^p \mathbb{P}(X_i | \mathbf{X}_{\setminus i})$.

Proposition

The solution to

$$\hat{\mathbf{K}} = \arg \max_{\mathbf{K}, K_{ij} \neq K_{ji}} \log \tilde{\mathcal{L}}(\mathbf{X}; \mathbf{K}) + \rho \|\mathbf{K}\|_{\ell_1}, \quad (2)$$

with

$$\tilde{\mathcal{L}}(\mathbf{X}; \mathbf{K}) = \sum_{i=1}^p \left(\sum_{k=1}^n \log \mathbb{P}(X_i^k | \mathbf{X}_{\setminus i}^k; \mathbf{K}_i) \right),$$

shares *the same null-entries* as the solution of the p independent penalized regressions.

GGMs and Lasso

Solving p penalized regressions \Leftrightarrow maximize the penalized pseudo-likelihood

Consider the approximation $\mathbb{P}(\mathbf{X}) = \prod_{i=1}^p \mathbb{P}(X_i | \mathbf{X}_{\setminus i})$.

Proposition

The solution to

$$\hat{\mathbf{K}} = \arg \max_{\mathbf{K}, K_{ij} \neq K_{ji}} \log \tilde{\mathcal{L}}(\mathbf{X}; \mathbf{K}) + \rho \|\mathbf{K}\|_{\ell_1}, \quad (2)$$

with

$$\tilde{\mathcal{L}}(\mathbf{X}; \mathbf{K}) = \sum_{i=1}^p \left(\sum_{k=1}^n \log \mathbb{P}(X_i^k | \mathbf{X}_{\setminus i}^k; \mathbf{K}_i) \right),$$

shares *the same null-entries* as the solution of the p independent penalized regressions.

\rightsquigarrow Those p terms are not independent, as \mathbf{K} is not diagonal !

\rightsquigarrow Still requires the post-symmetrization

GGMs and penalized likelihood

The penalized likelihood of the Gaussian observations [Banerjee et al., 2008]

Use a **penalty term**

$$\frac{n}{2} (\log \det(\mathbf{K}) - \text{Tr}(\mathbf{S}_n \mathbf{K})) - \rho \|\mathbf{K}\|_{\ell_1},$$

where \mathbf{S}_n is the empirical covariance matrix.

Natural generalization

Use different penalty parameters for different coefficients

$$\frac{n}{2} (\log \det(\mathbf{K}) - \text{Tr}(\mathbf{S}_n \mathbf{K})) - \|\rho_{\mathbf{z}}(\mathbf{K})\|_{\ell_1},$$

where $\rho_{\mathbf{z}}(\mathbf{K}) = (\rho_{z_i, z_j}(K_{ij}))_{i,j}$ is a penalty function depending on an unknown underlying structure \mathbf{Z} .

GGMs and Tree structure

The graph is a tree

- Chow-Liu algorithm (1968)

Input Σ_n ,

Output \hat{T}^{CL} , \hat{K}^{CL}

$$\hat{T}^{CL} = \arg \max_{T \text{ arbre}} \underbrace{\log(P(X; T))}_{\sum_{\{i,j\} \in E_T} I(X_i, X_j) + C} \quad (3)$$

- 1 Estimation of mutual information $\hat{I}(X_i, X_j)$
- 2 Maximal Spanning Tree relative to weights $\hat{I}(X_i, X_j)$

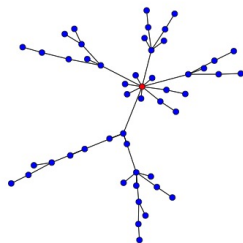


Figure 1: Tree

Outline

- 1 Introduction
- 2 Gaussian graphical models
- 3 GGM with missing variables**
- 4 EM with aggregation of spanning trees
- 5 Experiments - simulations and flow cytometry data

Effect of the missing variables

- Non measured variables
- Experimental conditions



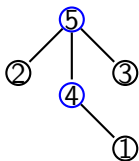
Figure 2: Covariance matrix. WGCNA data - 200 genes

Effect of the missing variables

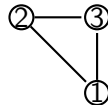
Missing variables

- Missing variables involved in the process of interest but not measured
- $\mathcal{G} = (\{1, \dots, p, p+1, \dots, p+r\}, E)$, $\mathcal{G}_m = (\{1, \dots, p\}, E_m)$
- Problem: inference of \mathcal{G}_m , \mathcal{G}

$$\begin{pmatrix} X_O \\ X_H \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{pmatrix} :$$



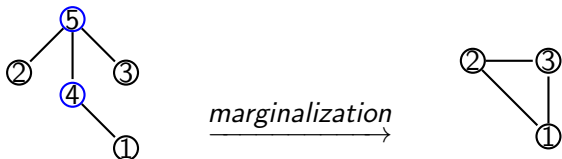
marginalisation \rightarrow



- Apparition of cliques

O = Observed, H = Hidden

Effect of the missing variables

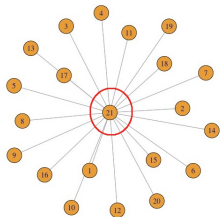


$$\mathcal{G} : K = \underbrace{\begin{pmatrix} K_{OO} & K_{OH} \\ K_{HO} & K_{HH} \end{pmatrix}}_{\text{arêtes de } E} \quad \Sigma = \begin{pmatrix} \Sigma_{OO} & \Sigma_{OH} \\ \Sigma_{HO} & \Sigma_{HH} \end{pmatrix}$$

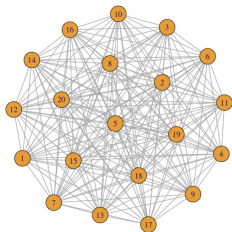
$$\mathcal{G}_m : K_m = \underbrace{K_{OO} - K_{OH}K_{HH}^{-1}K_{HO}}_{\text{arêtes de } E_m} \quad \Sigma_m = \Sigma_{OO}$$

Consequences

- [Chandrasekaran et al., 2012] \mathcal{G}_m is not sparse



(a) Full graph



(b) Marginal graph

Consequences on interpretation + on quality of inference

General conditions in sparse plus low-rank model [Chandrasekaran et al., 2012]

- 1 Support of the low-rank matrix $K_{OH}K_H^{-1}K_{HO}$ not sparse
 - a small number of hidden variables are connected to many observed variables
 - 2 K_O cannot have a low-rank structure
- Typically Graph structures with a small number of central hidden variables (hubs)

Identifiability

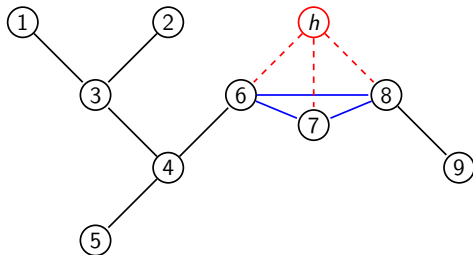
- maximal cliques of a tree are of size two.
- marginalizing a hidden variable produces a clique of size strictly more than two

Tree Structure Identifiability conditions [Choi et al., 2011]

- 1 Every hidden variable has at least 3 children
- 2 No edge between two hidden variables
- 3 No edge has weight 0 or ∞ (connected nodes are neither independent nor completely dependent)

Tree case

\mathcal{G} (red/black) tree



Inference with sparsity penalty

Latent variable selection via convex optimization [Chandrasekaran et al., 2012]

- control the number of latent variables by penalizing the rank of the matrix L

EM algorithm with Glasso [Lauritzen and Meinshausen, 2012]

Parameters: $K = \begin{pmatrix} K_{OO} & K_{OH} \\ K_{HO} & K_{HH} \end{pmatrix}$, $\Sigma = \begin{pmatrix} \Sigma_{OO} & \Sigma_{OH} \\ \Sigma_{HO} & \Sigma_{HH} \end{pmatrix}$

E-step: $\mathbb{E}_{X_H|X_O;K^t}[l_c(X_H, X_O)] = \mathbb{E}_{X_H|X_O;K^t}[\log \det(K^t) - \text{tr}(K^t \Sigma)]$

M-step: $K^{t+1} = \arg \max_K \underbrace{\log \det(K^t) - \text{tr}(K^t \mathbb{E}_{X_H|X_O;K^t}[\Sigma])}_{\text{graphical lasso}} + \lambda \|K_{OO}^t\|$

Inference with tree assumption

Recursive Grouping [Choi et al., 2011]

- building of a latent tree from data in the gaussian case
- heuristic based on the so-called information distances

EM with Chow-Liu M step

- Idea of [Lauritzen and Meinshausen, 2012] replacing Glasso with Chow-liu
- Highly constrained structure

Outline

- 1 Introduction
- 2 Gaussian graphical models
- 3 GGM with missing variables
- 4 EM with aggregation of spanning trees**
- 5 Experiments - simulations and flow cytometry data

Escaping the Tree constraint

Tree as a random variable [Schwaller and Robin, 2015]

$$\pi_{ij} = P(\{i, j\} \in E_T).$$

The edges of T are drawn independently such that

$$P(T) \propto \prod_{\{i, j\} \in E_T} \pi_{ij}. \quad (4)$$

Missing variable structure

We further assume the existence of a full symmetric positive definite matrix

$$K = \begin{pmatrix} K_O & K_{OH} \\ K_{HO} & K_H \end{pmatrix}$$

of which we want to infer the coefficients.

Escaping the Tree constraint

Mixture of trees

(X_O, X_H) is a mixture of centered Gaussian distributions with respective precision matrices K_T :

$$(X_O, X_H) \sim \sum_{T \in \mathcal{T}} p(T) \mathcal{N}(X_O, X_H; 0, K_T^{-1})$$

Conditionally to a Tree

For every $T \in \mathcal{T}$ we define the matrix K_T such that for $(i, j) \in \{1, \dots, p, p+1, \dots, p+r\} \times \{1, \dots, p, p+1, \dots, p+r\}$

$$K_{T,ij} = \begin{cases} K_{ij} & \text{if } \{i, j\} \in E_T \\ 0 & \text{otherwise} \end{cases} .$$

- T and X_H are both latent variables.

Some conditional distributions

Joint conditional distribution of T and X_H given X_O

$$P(T, X_H | X_O) = P(T | X_O) P(X_H | X_O, T).$$

with



$$P(X_H | X_O, T) = \mathcal{N}(\mu_{H|O,T}, K_{H|O,T}) \quad (5)$$



$$\begin{aligned} P(T | X_O) &\propto P(T) P(X_O | T) \\ &\propto \left(\prod_{\{i,j\} \in E_T} \pi_{ij} \right) \underbrace{\frac{\det(K_{T,M})^{\frac{n}{2}}}{(2\pi)^{\frac{np}{2}}}}_{(1)} \underbrace{\exp\left(-\frac{n}{2} \text{tr}(K_{T,M} \Sigma_O)\right)}_{(2)}, \end{aligned} \quad (6)$$

where $K_{T,M} = K_{T,O} - K_{T,OH}(K_{T,H})^{-1}K_{T,HO}$. Terms (1) and (2) can be expressed as products over the edges of T .

EM algorithm

Maximizing the log-likelihood of the observed data $\log p(X_O; K)$ with respect to the parameter K , alternating two steps:

E-step: Evaluation of all the conditional moments involved in the conditional expectation of the so-called complete likelihood with the current value K^h of the parameter, namely:

$$\mathbb{E}_{X_H, T | X_O; K^h} [\log p(X_O, X_H, T; K)]; \quad (7)$$

M-step: Maximization of (7) with respect to K to update K^h into K^{h+1} .

E-step

The conditional expectation of the complete likelihood writes

$$\begin{aligned} & \mathbb{E}_{T|X_O;K^h} \left(\mathbb{E}_{X_H|X_O,T} \log p(X_O, X_H, T; K) \right) \\ &= \mathbb{E}_{T|X_O;K^h} \left(\log p(T) + \mathbb{E}_{X_H|X_O,T;K^h} [\log p(X_O, X_H|T; K)] \right). \end{aligned}$$

Thanks to the tree structure of the graphical model, we have a simple form for the latter term:

$$\mathbb{E}_{X_H|X_O,T;K^h} [\log p(X_O, X_H|T; K)] = \sum_{\{i,j\} \in T} p_{ij}(K)$$

M-step

Combined with $p(T) \propto \prod_{\{i,j\} \in T} \pi_{ij}$ and with the conditional distribution of T , $p(T|X_O; K^h) \propto \prod_{\{i,j\} \in T} \gamma_{ij}$

$$\begin{aligned} & \mathbb{E}_{X_H, T|X_O; K^h} \log p(X_O, X_H, T; K) \\ & \propto \sum_T \left(\prod_{\{k,\ell\} \in T} \gamma_{k\ell}^h \right) \left[\sum_{\{i,j\} \in T} \log \pi_{ij} + p_{ij}(K) \right] \end{aligned}$$

where the normalizing constant does depend on K^h but not on K . Hence, at the M-step we need to maximize wrt K

$$\sum_T \left(\prod_{\{k,\ell\} \in T} \gamma_{k\ell}^h \right) \left[\sum_{\{i,j\} \in T} p_{ij}(K) \right] = \sum_{i < j} A_{ij} p_{ij}(K) \quad (8)$$

where all $A_{ij} = \sum_{T: \{i,j\} \in T} \left(\prod_{\{k,\ell\} \in T} \gamma_{k\ell}^h \right)$ can be computed in $O((p+r)^3)$ using the matrix tree theorem.

Edge probability

Edge probability

We need to compute the probability for an edge to be part of the tree given X_O

$$\alpha_{kl} := P(\{k, l\} \in T | X_O) = 1 - \sum_{T: \{k, l\} \notin E_T} P(T | X_O). \quad (9)$$

This probability can be computed for all edges at a time in $O((p+r)^3)$ thanks to Matrix Tree Theorem

Model Selection

Maximum Log Likelihood

$$\log p(X_O; \hat{K})$$

can be computed as

$$E[\log p(X_O, X_H, T) | X_O; \hat{K}] + H(X_H, T | X_O, \hat{K})$$

BIC

a standard BIC criterion can be defined as

$$BIC(r) = \log p(X_O; \hat{K}) - \text{pen}(r)$$

where

$$\text{pen}(r) = \left(\frac{p(p+1)}{2} + rp + r \right) \frac{\log n}{2}. \quad (10)$$

Model Selection (2)

ICLs

$$ICL_T(r) = \log p(X_O; \hat{K}) - H(T|X_O) - \text{pen}(r)$$

In situations where a reliable prediction of the hidden node X_H is of interest,

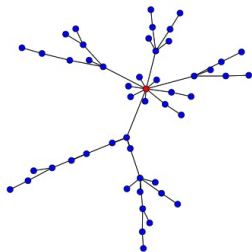
$$ICL_{T, X_H}(r) = \log p(X_O; \hat{K}) - H(T, X_H|X_O) - \text{pen}(r).$$

Outline

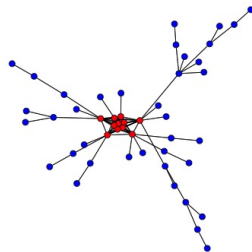
- 1 Introduction
- 2 Gaussian graphical models
- 3 GGM with missing variables
- 4 EM with aggregation of spanning trees
- 5 Experiments - simulations and flow cytometry data**

Simulated data

- Graphs of size $p = 50$: tree, Erdős ($\pi = 0.1$), one Hub with Erdős
- Samples of size $n = 200$

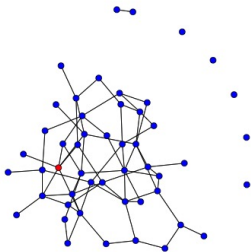


(a) Tree

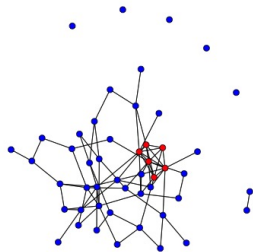


(b) Tree (marg.)

Simulated data

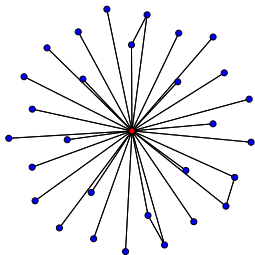


(a) Erdős

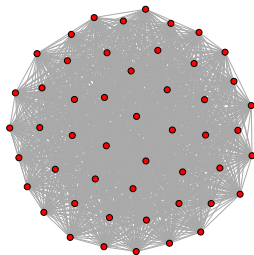


(b) Erdős (marg.)

Simulated data



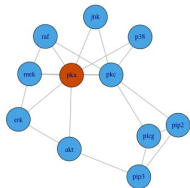
(a) Erdős with Hub



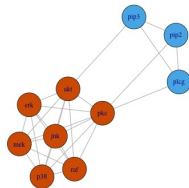
(b) Erdős with Hub (marg.)

Real data

- Raf network (regulation of cellular proliferation)
- Flow cytometry
- $p=11$, $n=100$



(a) Full graph



(b) Marginal graph

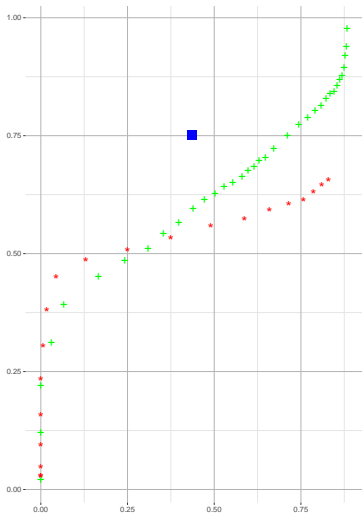
Compared methods

- Chow-Liu
- Recursive Grouping
- Glasso (Meinshausen & Bühlmann approximation)
- EM-Glasso
- EM-Chow-Liu
- EM-aggregation

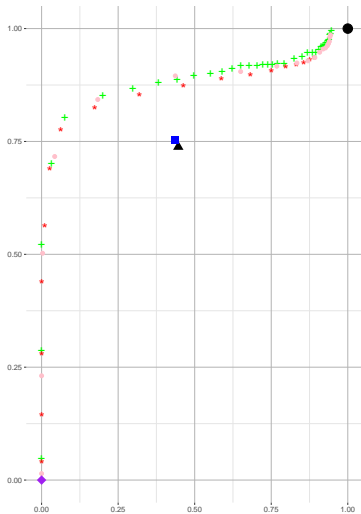
Evaluation criterion

$$\text{power} = \frac{TP}{FN + TP}, \text{FDR} = \frac{FP}{FP + TP}$$

Results

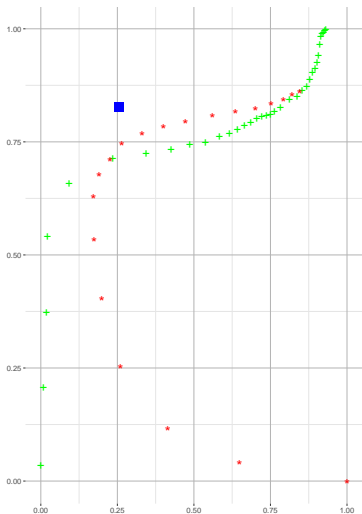


(a) Full graph estimation with the Hub Erdős Data

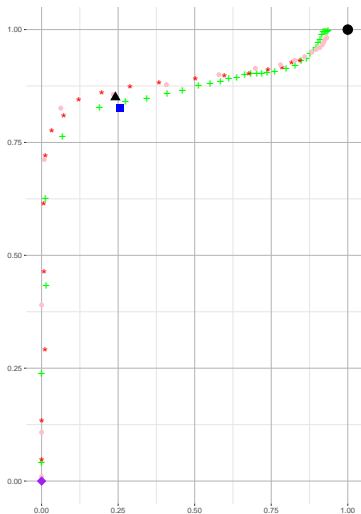


(b) Conditional graph estimation with the Hub Erdős Data

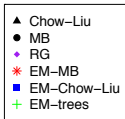
Results



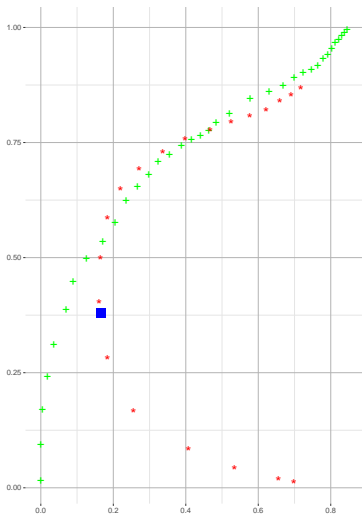
(a) Full graph estimation with the Tree Data



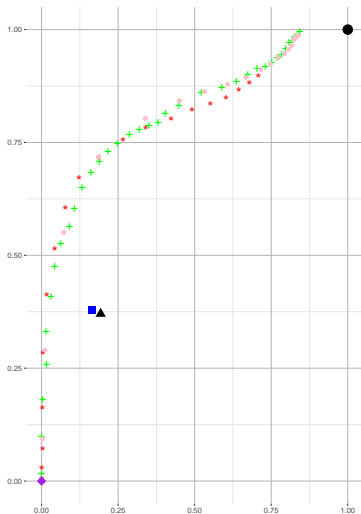
(b) Conditional graph estimation with the Tree Data



Results



(a) Full graph estimation with the Erdős Data



(b) Conditional graph estimation with the Erdős Data

Model Selection for Erdős data

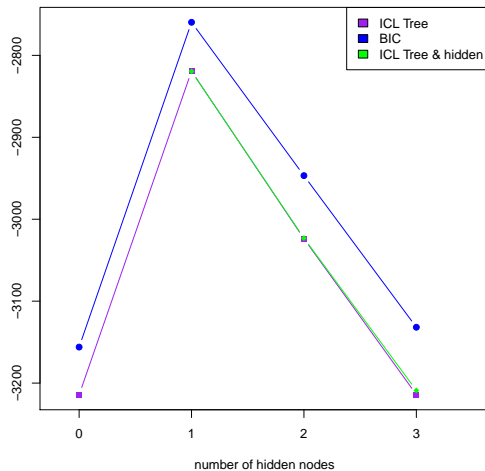


Figure 11: Existence of missing nodes

Flow cytometry

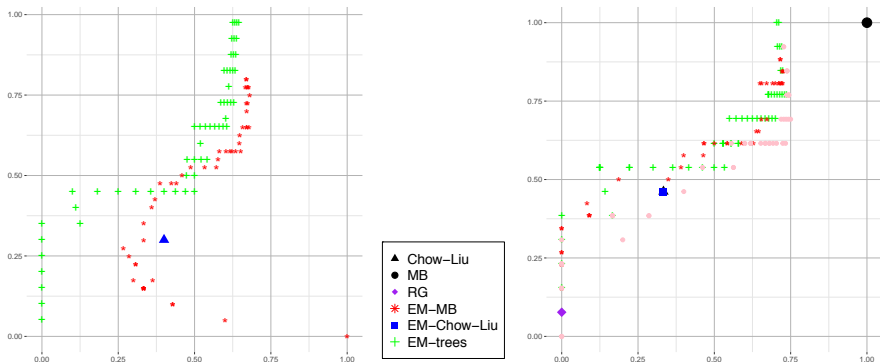


Figure 12: cytometry Precision-Recall curves for graph inference results on flow cytometry data. Full graph (left) and Conditional graph (right)

Model Selection for Flow cytometry

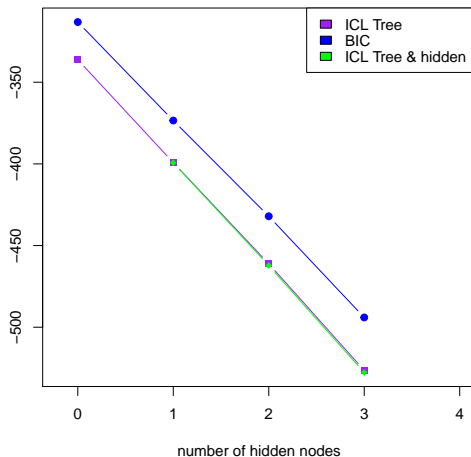


Figure 13: Existence of missing nodes

Improving on the initialization

- Hierarchical classification

Developing R package

Extension to

- Count data (non Gaussian) via Poisson Log-Normal
- Temporal data (Dependence between samples)
- Covariates

References

- Onureena Banerjee, Laurent El Ghaoui, and Alexandre d'Aspremont. Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *Journal of Machine Learning Research*, 9(Mar):485–516, 2008.
- V. Chandrasekaran, P. A. Parrilo, and A. S. Willsky. Latent variable graphical model selection via convex optimization. *The Annals of Statistics*, 40(4):1935–1967, 2012.
- M. J. Choi, V. Tan, A. Anandkumar, and A. S. Willsky. Learning latent tree graphical models. *The Journal of Machine Learning Research*, 12: 1771–1812, 2011.
- S. Lauritzen and N. Meinshausen. Discussion: Latent variable graphical model selection via convex optimization. *The Annals of Statistics*, 2012.
- L. Schwaller and S. Robin. Apprentissage de réseaux par agrégation bayésienne d'arbres couvrants. *Revue d'intelligence artificielle*, 2: 153–172, 2015.

Initialisation

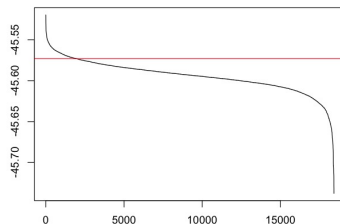


Figure 14: Vraisemblance des observations pour chaque triplet possible

Initialisation

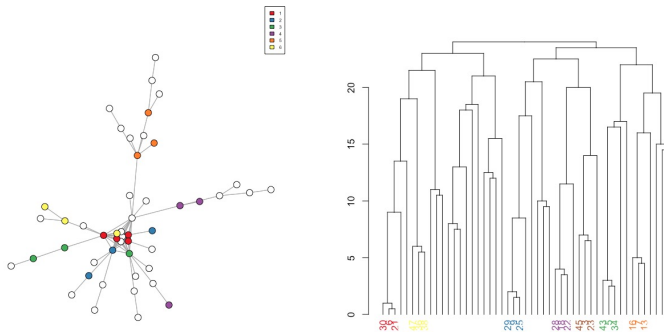


Figure 15: Classification hiérarchique au max du BIC