Inferring a biological network from transcriptomic data ?

J. Chiquet, M-L Martin-Magniette, F. Monéger,
G. Rigaill, L. Schwaller

netbio 2015

Outline

1. Introduction
2. Some network inference approaches
    - Graphical model
    - Tree-structured Graphical Models
3. Some results
4. Conclusion

Network inference and transcriptomic data

Some practical problems

1. Most of the time we measure the transcription of a mixed population of cells (having different phenotypes)
2. The number of replicates is often small
    - We should restrict ourselves to rather small networks.
    - Which genes should we include in the analysis ?

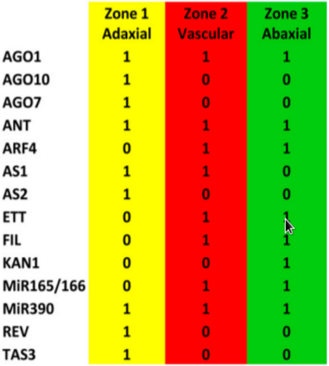
Question

Given a biological network (= given by a biologist), is it possible to recover it from transcriptomic data alone ?

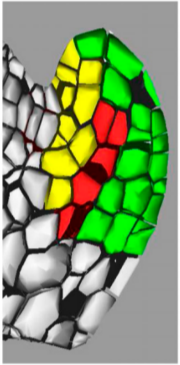
A well characterized network (La Rota et al. 2011)

Construction of a sepal primordium network

- Extensive literature/database search
- Expression pattern of different zones of the sepal primordium

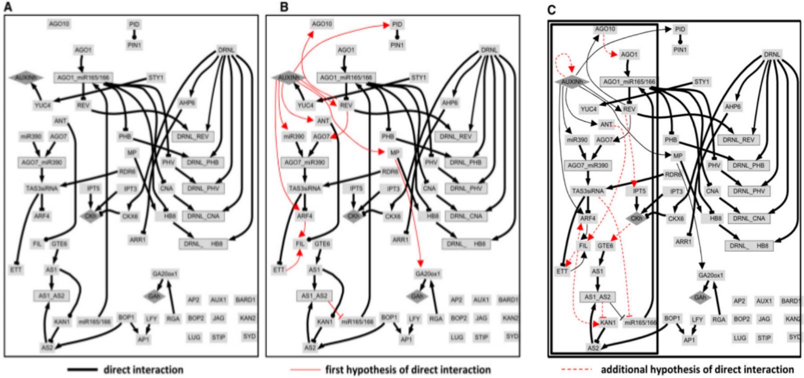


	Zone 1 Adaxial	Zone 2 Vascular	Zone 3 Abaxial
AGO1	1	1	1
AGO10	1	0	0
AGO7	1	0	0
ANT	1	1	1
ARF4	0	1	1
AS1	1	1	0
AS2	1	0	0
ETT	0	1	1
FIL	0	1	1
KAN1	0	0	1
MiR165/166	0	1	1
MiR390	1	1	1
REV	1	0	0
TAS3	1	0	0

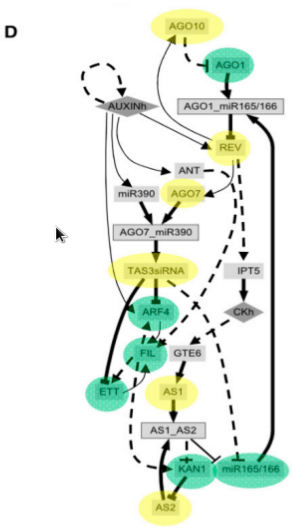


A well characterized network (La Rota et al. 2012)

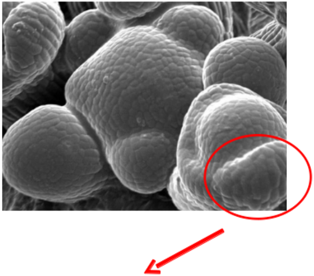
- Successive refinement of the network
- Coherence with observed expression patterns (zone of the sepal primordium)



Final network (La Rota et al. 2012)



Matériel utilisé pour l'analyse du transcriptome



sépale au stade 3

⬇

réseau de La Rota et al, 2011

3 inflorescences = 1 échantillon

⬇

extraction d'ARN

⬇

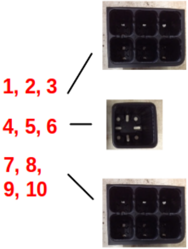
analyse transcriptome

3 chambres de cultures et pots différents

4 plantes/pot

1 plante/pot

Chambre de culture	T°(J/N)	H°(J/N)	Photopériode (J/N)
Madrid	21°/17°	50-65%	16h/8h
Reyjavik	16°	70%	24h
Berlin	21°/17°	50-65%	16h/8h



10 échantillons (répliques biologiques)
Numérotés de 1 à 10

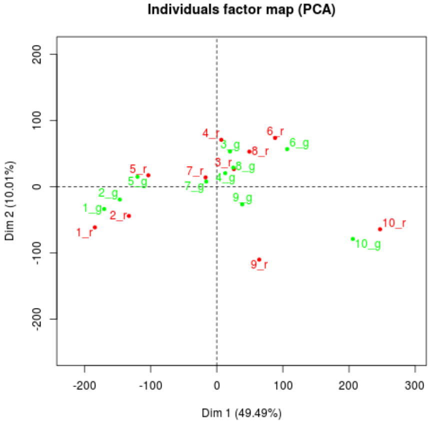
Exploratory Analysis

- 10 samples coming from 3 growing rooms
- 2 measurements per sample (red, green)

- Normalization ?
  - Red-Green or Dye-Swapped
  - Raw or Centered per array/sample
- Further normalization for network inference
  - Pearson or Spearman correlation
  - Copula or non-paranormal skeptic
  - ...

Some preliminary analysis (on all genes)

- Replicates are similar most of the time
- Different groups of samples are visible using PCA or clustering...
  - They do not correspond to growing rooms



Probe Selection

- Map each gene to one probe
- Some genes have more than one probe
  - Highest quality available
  - Highest expression level

Outline

1. Introduction
2. Some network inference approaches
    - Graphical model
    - Tree-structured Graphical Models
3. Some results
4. Conclusion

Correlation

Simple correlation

1. Compute the correlation matrix
2. Predict an edge between two genes if their absolute correlation is above a given threshold

Correlation + hierarchical clustering (hc)

1. Compute the correlation matrix
2. Recover a distance matrix from this correlation matrix
3. Hierarchical clustering
4. For a cut of the tree predict an edge between two genes if they are in the same cluster

...

Outline

1. Introduction
2. Some network inference approaches
    - Graphical model
    - Tree-structured Graphical Models
3. Some results
4. Conclusion

Graphical model

Definition

A graphical model gives a graphical (intuitive) representation of the dependence structure of a probability distribution. It links

1. a random vector $X = \{X_1, \dots, X_p\}$ with distribution $\mathbb{P}$,
2. a graph $\mathcal{G} = (\mathcal{P}, \mathcal{E})$ where
    - $\mathcal{P} = \{1, \dots, p\}$ is the set of nodes associated to each variable,
    - $\mathcal{E}$ is a set of edges describing the dependence relationship of $X \sim \mathbb{P}$.

Conditional independence graph

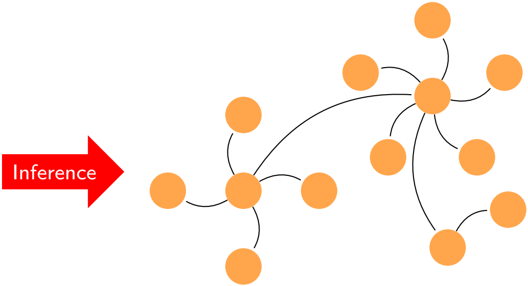
It is the undirected graph $\mathcal{G} = \{\mathcal{P}, \mathcal{E}\}$ where

$$(i,j) \notin \mathcal{E} \Leftrightarrow X_i \perp\!\!\!\perp X_j | \mathcal{P} \backslash \{i,j\}.$$

Gaussian Graphical models

Undirected network with Gaussian distribution

$$\mathbf{X}^{(1)} = (X_1^{(1)}, \dots, X_p^{(1)})$$
$$\mathbf{X}^{(2)} = (X_1^{(2)}, \dots, X_p^{(2)})$$
$$\dots$$
$$\mathbf{X}^{(n)} = (X_1^{(n)}, \dots, X_p^{(n)})$$



i.i.d. sample

$\mathcal{G} = (\mathcal{P}, \mathcal{E})$

Multivariate Gaussian assumption

Let $\mathbf{X} \sim \mathcal{N}(\mathbf{0}_p, \mathbf{\Sigma})$ and $\mathbf{\Theta} = \mathbf{\Sigma}^{-1}$ the precision matrix.

GGM and partial covariance

Gaussian vector

Suppose $X \sim \mathcal{N}(\boldsymbol{\mu}, \begin{pmatrix} \Sigma_{ab} & \Sigma_{ba} \\ \Sigma_{ab} & \Sigma_{bb} \end{pmatrix})$, then

1. X_a is Gaussian with distribution $\mathcal{N}(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_{aa})$
2. $X_a|X_b = x$ is Gaussian with distribution $\mathcal{N}(\boldsymbol{\mu}_{a|b}, \boldsymbol{\Sigma}_{a|b})$.

Partial covariance/correlation and conditional independance

Let X, Y, Z be real random variables.

$$\mathrm{cov}(X, Y|Z) = \mathrm{cov}(X, Y) - \mathrm{cov}(X, Z)\mathrm{cov}(Y, Z)/\mathrm{Var}(Z).$$

$$\rho_{XY|Z} = \frac{\rho_{XY} - \rho_{XZ}\rho_{YZ}}{\sqrt{1-\rho^2_{XZ}}\sqrt{1-\rho^2_{YZ}}}.$$

When X, Y, Z are jointly Gaussian, then

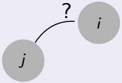
Gaussian Graphical Model and covariance selection

Inverse covariance viewpoint

$$-\frac{\Theta_{ij}}{\sqrt{\Theta_{ii}\Theta_{jj}}} = \operatorname{cor}\left(X_i, X_j | X_{\mathcal{P}\backslash i,j}\right) = \rho_{ij|\mathcal{P}\backslash\{i,j\}},$$

Graphical Interpretation

$\leadsto$ The matrix $\boldsymbol{\Theta} = (\Theta_{ij})_{i,j\in\mathcal{P}}$ encodes the network $\mathcal{G}$ we are looking for.



if and only if

conditional dependency between X_j and X_i
or
non-null partial correlation between X_j and X_i
$\Updownarrow$
$\Theta_{ij} \neq 0$

$\leadsto$ "covariance" selection

Gaussian Graphical Model and Linear Regression

Linear regression viewpoint

Gene expression X_i is linearly explained by the other genes':

$$X_i | X_{\setminus i} = - \sum_{j \neq i} \frac{\Theta_{ij}}{\Theta_{ii}} X_j + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma_i), \quad \varepsilon_i \perp X$$

Conditional on its neighborhood, other profiles do not give additional insights

$$X_i | X_{\setminus i} = \sum_{\color{red}{j \in \text{neighbors}(i)}} \beta_j X_j + \varepsilon_i \quad \text{with } \beta_j = - \frac{\Theta_{ij}}{\Theta_{ii}}.$$

⇝ "Neighborhood" selection

The penalized likelihood approach

Let $\mathbf{\Theta}$ be the model parameter to infer (related to the edges).

Constraint Optimization approach

$$\hat{\mathbf{\Theta}}_\lambda = \arg\max_{\mathbf{\Theta}} \left\{ \underbrace{\log\det\mathbf{\Theta} - \mathrm{trace}\left(\mathbf{S}\mathbf{\Theta}\right)}_{\propto \text{ log-likelihood}} \right\} \quad \text{s.t.} \quad \sum_{i>j} \mathbf{\Theta}_{ij} \neq 0 \leq c,$$

where $\mathbf{S} = n^{-1}\mathbf{X}^\intercal\mathbf{X}$.

Gold standard convexified penalized approaches

Use ℓ_1 as a proxy fot the ℓ_0

Penalized likelihood (Banerjee *et al.*, Yuan and Lin, 2008)

$$\hat{\boldsymbol{\Theta}}_\lambda = \arg\max_{\boldsymbol{\Theta}\in\mathbb{S}_+} \ell(\boldsymbol{\Theta};\mathbf{X}) - \lambda\|\boldsymbol{\Theta}\|_1$$

+ symmetric, positive-definite
− solved by the “Graphical-Lasso” ($\mathcal{O}(p^3)$, *Friedman et al, 2007*).

Neighborhood Selection (Meinshausen & Bülhman, 2006)

$$\widehat{\beta}^{(i)} = \arg\min_{\beta\in\mathbb{R}^{p-1}} \frac{1}{n}\left\|\mathbf{X}_i - \mathbf{X}_{\backslash i}\,\beta\right\|_2^2 + \lambda\left\|\beta\right\|_1$$

CLIME – Pseudo-likelihood (Cai et al., 2011; Yuan, 2010)

$$\widehat{\boldsymbol{\Theta}} = \arg\min_{\boldsymbol{\Theta}} \left\|\boldsymbol{\Theta}\right\|_1 \text{ subjected to } \left\|n^{-1}\mathbf{X}^t\mathbf{X}\boldsymbol{\Theta} - \mathbf{I}\right\|_\infty \leq \lambda$$

Gold standard convexified penalized approaches

Use ℓ_1 as a proxy fot the ℓ_0

Penalized likelihood (Banerjee *et al.*, Yuan and Lin, 2008)

$$\hat{\mathbf{\Theta}}_\lambda = \arg\max_{\mathbf{\Theta}\in\mathbb{S}_+} \ell(\mathbf{\Theta};\mathbf{X}) - \lambda\|\mathbf{\Theta}\|_1$$

Neighborhood Selection (Meinshausen & Bülhman, 2006)

$$\widehat{\boldsymbol{\beta}}^{(i)} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^{p-1}} \frac{1}{n}\left\|\mathbf{X}_i - \mathbf{X}_{\backslash i}\,\boldsymbol{\beta}\right\|_2^2 + \lambda\left\|\boldsymbol{\beta}\right\|_1$$

− not symmetric, not positive-definite

\+ p Lasso solved with Lars-like algorithms ($\mathcal{O}(npd)$ for d neighbors).

CLIME – Pseudo-likelihood (Cai et al., 2011; Yuan, 2010)

$$\widehat{\mathbf{\Theta}} = \arg\min_{\mathbf{\Theta}} \|\mathbf{\Theta}\|_1 \text{ subjected to } \left\|n^{-1}\mathbf{X}^t\mathbf{X}\mathbf{\Theta} - \mathbf{I}\right\|_\infty \leq \lambda$$

Gold standard convexified penalized approaches

Use ℓ_1 as a proxy fot the ℓ_0

Penalized likelihood (Banerjee *et al.*, Yuan and Lin, 2008)

$$\hat{\boldsymbol{\Theta}}_\lambda = \arg\max_{\boldsymbol{\Theta}\in\mathbb{S}_+} \ell(\boldsymbol{\Theta};\mathbf{X}) - \lambda\|\boldsymbol{\Theta}\|_1$$

Neighborhood Selection (Meinshausen & Bülhman, 2006)

$$\widehat{\boldsymbol{\beta}}^{(i)} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^{p-1}} \frac{1}{n}\left\|\mathbf{X}_i - \mathbf{X}_{\backslash i}\boldsymbol{\beta}\right\|_2^2 + \lambda\left\|\boldsymbol{\beta}\right\|_1$$

CLIME – Pseudo-likelihood (Cai et al., 2011; Yuan, 2010)

$$\widehat{\boldsymbol{\Theta}} = \arg\min_{\boldsymbol{\Theta}} \|\boldsymbol{\Theta}\|_1 \text{ subjected to } \left\|n^{-1}\mathbf{X}^t\mathbf{X}\boldsymbol{\Theta} - \mathbf{I}\right\|_\infty \leq \lambda$$

- − not positive-definite
- \+ p linear programs easily distributed ($\mathcal{O}(p^2 d)$ for d neighbors).

Practical implications of theoretical results

Selection consistency (Ravikumar, Wainwright, 2009-2012)

Denote $d = \max_{j \in \mathcal{P}}(\mathrm{degree_j})$. Consistency for an appropriate λ and

- $n \approx \mathcal{O}(d^2 \log(p))$ for the graphical Lasso and Clime.
- $n \approx \mathcal{O}(d \log(p))$ for neighborhood selection (sharp).

Ultra high-dimension phenomenon (Verzelen, 2011)

Minimax risk for sparse regression with d-sparse models: useless when

$$\frac{d \log(p/d)}{n} \geq 1/2, \qquad (\text{e.g.}, n = 50, p = 200, d \geq 8).$$

What about Count data/non Gaussian distribution

Can be handled after data transformation ("skeptic" by Wasserman et al., package `huge`).

Practical implications of theoretical results

Selection consistency (Ravikumar, Wainwright, 2009-2012)

Denote $d = \max_{j\in\mathcal{P}}(\mathrm{degree_j})$. Consistency for an appropriate λ and

- $n \approx \mathcal{O}(d^2 \log(p))$ for the graphical Lasso and Clime.
- $n \approx \mathcal{O}(d \log(p))$ for neighborhood selection (sharp).

Ultra high-dimension phenomenon (Verzelen, 2011)

Minimax risk for sparse regression with d-sparse models: useless when

$$\frac{d\log(p/d)}{n} \geq 1/2, \qquad (\text{e.g.}, n = 50, p = 200, d \geq 8).$$

What about Count data/non Gaussian distribution

Can be handled after data transformation ("skeptic" by Wasserman et al., package huge).

Practical implications of theoretical results

Selection consistency (Ravikumar, Wainwright, 2009-2012)

Denote $d = \max_{j \in \mathcal{P}}(\mathrm{degree_j})$. Consistency for an appropriate λ and

- $n \approx \mathcal{O}(d^2 \log(p))$ for the graphical Lasso and Clime.
- $n \approx \mathcal{O}(d \log(p))$ for neighborhood selection (sharp).

Ultra high-dimension phenomenon (Verzelen, 2011)

Minimax risk for sparse regression with d-sparse models: useless when

$$\frac{d \log(p/d)}{n} \geq 1/2, \qquad (\text{e.g.}, n = 50, p = 200, d \geq 8).$$

What about Count data/non Gaussian distribution

Can be handled after data transformation ("skeptic" by Wasserman et al., package `huge`).

Outline

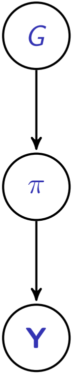
1. Introduction

2. Some network inference approaches
    - Graphical model
    - Tree-structured Graphical Models

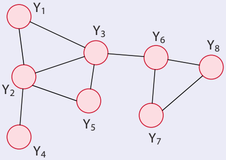
3. Some results

4. Conclusion

Tree-structured Graphical Models



- $G = (V, E_G)$ decomposable graph

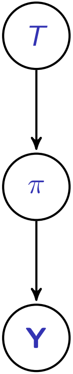


$\mathcal{G}$ = { decomposable graphs }

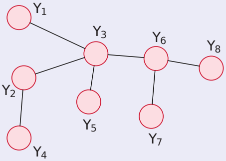
- π distribution for $\mathbf{Y}$
  Markov w.r.t. G

$$\{i,j\} \notin E_G \Leftrightarrow Y_i \perp\!\!\!\perp Y_j | Y_{V\setminus\{i,j\}}$$

Tree-structured Graphical Models



- $T = (V, E_T)$ **spanning tree**

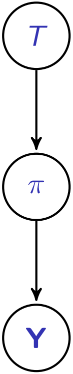


$\mathcal{T} = \{$ spanning trees $\}$

- π distribution for $\mathbf{Y}$
  Markov w.r.t. T

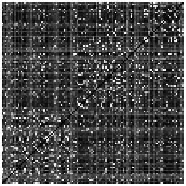
$$\{i, j\} \notin E_T \Leftrightarrow Y_i \perp\!\!\!\perp Y_j | Y_{V \setminus \{i,j\}}$$

Tree-structured Graphical Models



$$T \sim \mathcal{U}(\mathcal{T})$$
$$\pi | T \sim \rho$$
$$\mathbf{Y} | \pi \sim \pi$$

π	ρ
Multinomial	Dirichlet
Gaussian	normal-Wishart for (μ, Λ)
Copula	$\mathcal{U}([-1, 1])$ for the entries of the precision matrix



Posterior Edge Probabilities

$$p(\{i,j\} \in E_T|\mathbf{Y}) = \sum_{\substack{T \in \mathcal{T} \\ T \ni \{i,j\}}} p(T|\mathbf{Y})$$

- Easily computed thanks to an algebra result called the **Matrix-Tree** if ρ has some Markov property

  ▷ Complexity $\mathcal{O}(p^3)$

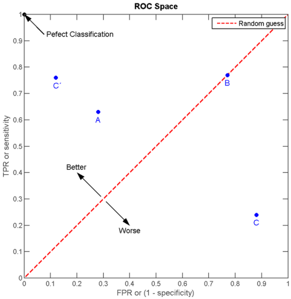
ROC curves

- For a given threshold (λ, correlation, ...) we can compute the table

	Edges	None-Edges
Pred. edges	True Positive	False Positive
Pred. none-edges	.	.

  - TPR : TP / (Number of true edges)
  - FPR : FP / (Number of none-edges)

- Compute the TPR and FPR for various thresholds
- Draw the TPR as a function of FPR



Outline

1. Introduction

2. Some network inference approaches
   - Graphical model
   - Tree-structured Graphical Models

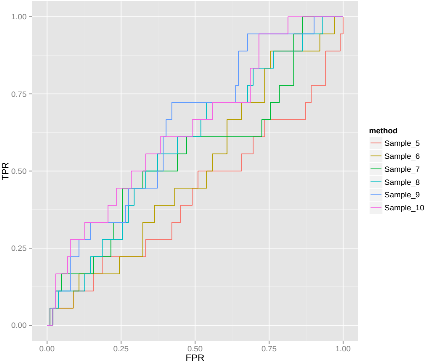
3. Some results

4. Conclusion

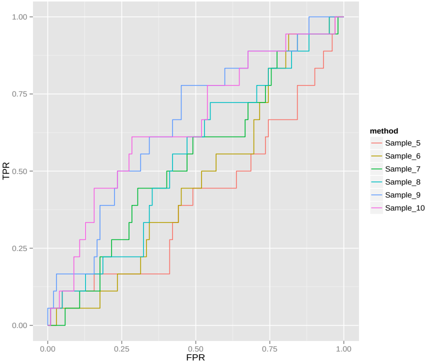
Performances as a function of the number of samples

Dye-swapped and mean centered

- Pearson correlation



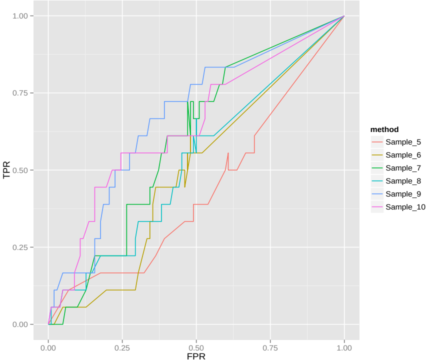
- Spearman correlation



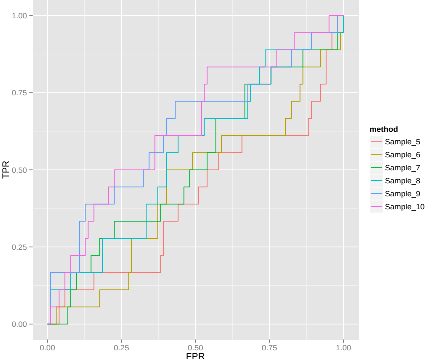
Performances as a function of the number of samples

Dye-swapped and mean centered

- Spearman + Glasso



- Copula + Tree



...

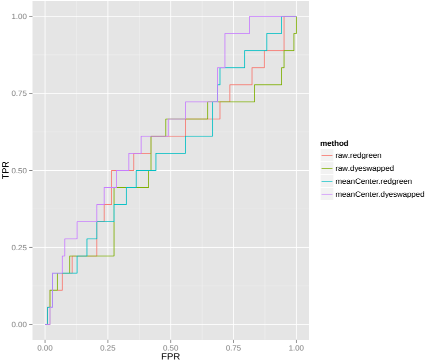
Performances as a function of the number of samples

Conclusion

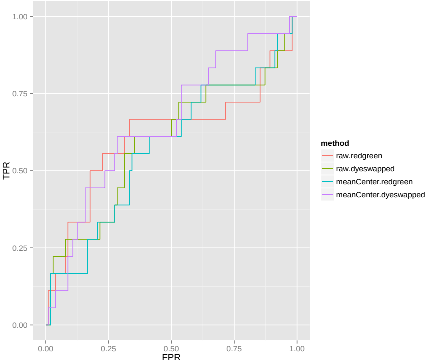
- We recover part of the network !

Performances as a function of the normalization

- Pearson Correlation

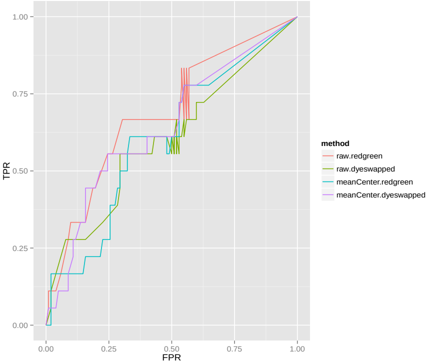


- Spearman Correlation

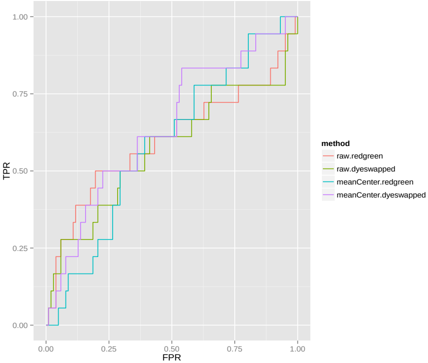


Performances as a function of the normalization

- Spearman + GLasso



- Copula + Tree



...

Performances as a function of the normalization

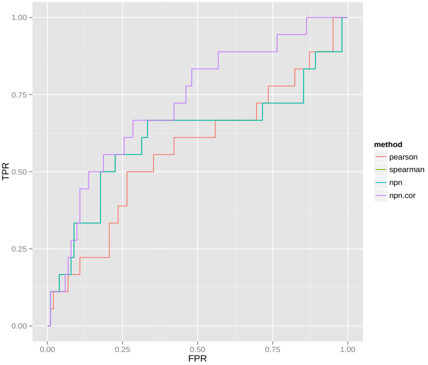
Conclusion

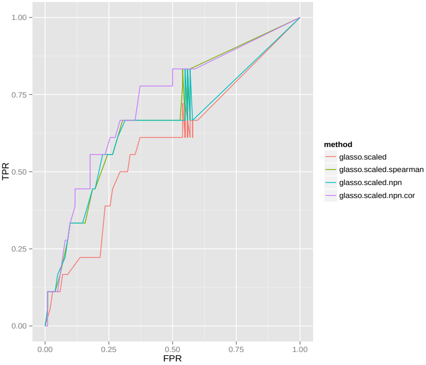
- Normalization counts
- Raw and red-green ?

Performances as a function of the transformation

red-green and raw

- Correlation
- GLasso





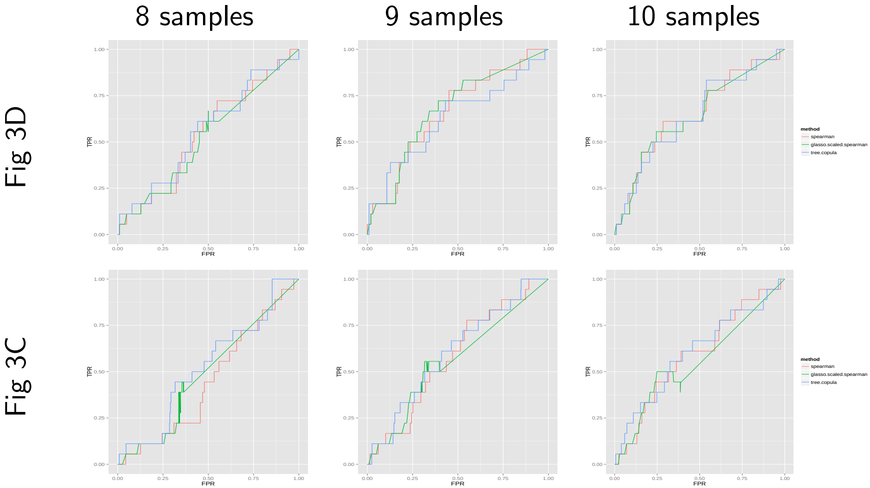
Performances as a function of the transformation

Conclusion

- It seems important to transform the data
- Non-paranormal skeptic + correlation ?

Influence of the difficulty level

dye-swapped mean centered (spearman, glasso+spearman, tree+copula)



8 samples

9 samples

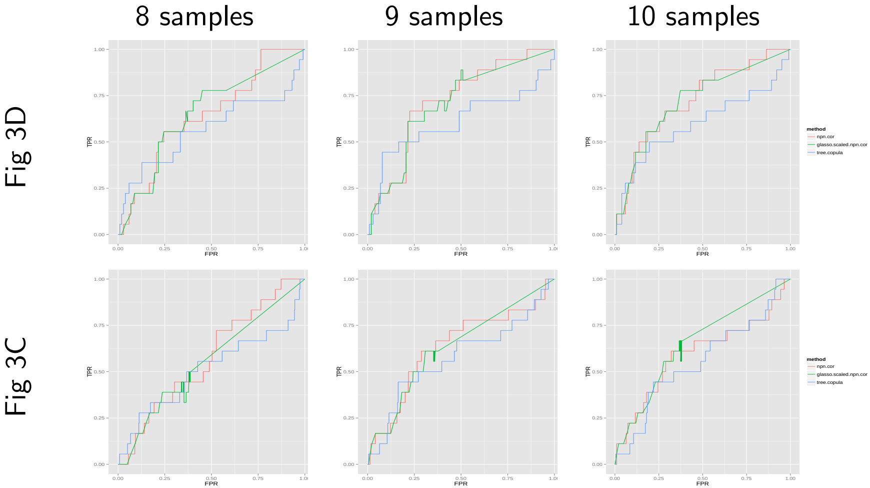
10 samples

Fig 3D

Fig 3C

Influence of the difficulty level

red-green and raw (npn.cor, glasso+npn.cor, tree+copula)



Influence of the difficulty level

Conclusion

- Results are worst when considering more genes

Outline

1. Introduction

2. Some network inference approaches
    - Graphical model
    - Tree-structured Graphical Models

3. Some results

4. Conclusion

Conclusion

Conclusion

- We recover at least part of the network
- More samples leads to better results
- Some normalizations and transformations seem to work better

To do

- Infer the network using none-dedicated data
- More approaches (prior) ?

Questions?

- Raw and red-green
- Non-paranormal skeptic + cor