Mixture tree model for network inference

JOBIM

Raphaëlle Momal

UMR518 AgroParis Tech/INRA

June 19, 2018

Context

Rising interest in jointly analysed species abundances:

- Metagenomics
- Microbiologie
- Ecology

Context

Rising interest in jointly analysed species abundances:

- Metagenomics
- Microbiologie
- Ecology

Ecological network

Tool to better understand species interactions (direct/indirect, nature), eco-systems organizations (clusters ?)

Context

Rising interest in jointly analysed species abundances:

- Metagenomics
- Microbiologie
- Ecology

Ecological network

Tool to better understand species interactions (direct/indirect, nature), eco-systems organizations (clusters ?)

Allows for resilience analyses, pathogens control, ecosystem comparison, reaction prediction...

Example

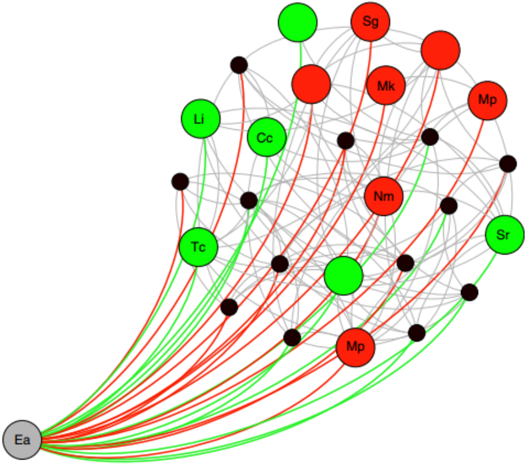


Figure: *Erysiphe alphitoides* pathobiom on oak tree leaves.

Data

- Species : animal, bacteria, gene ...
- Abundances : ecological counts, Next-Generation Sequencing technologies...
- Covariates : coverage, temperature, water depth ...

Repeated signal : n samples of p abundances.

Data

- Species : animal, bacteria, gene ...
- Abundances : ecological counts, Next-Generation Sequencing technologies...
- Covariates : coverage, temperature, water depth ...

Repeated signal : n samples of p abundances.

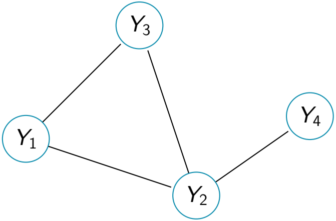
Notations

$Y = [Y_{ij}]_{(i,j)\in\{1,\ldots,n\}\times\{1,\ldots,p\}}$

- Y_{ij} : abundance of the j^{th} species in the i^{th} sample

Infer the species interaction network from Y

Graphical models



- All variables are dependant
- Some are conditionnally independant (i.e. indirectly dependant)

$$Y_4 \perp\!\!\!\perp (Y_1, Y_3) | Y_2$$

Graphical models

Factorization propriety

The joint distribution P is faithful to the graph G iff

$$P(Y_1, \dots, Y_p) \propto \prod_{C \in \mathcal{C}_G} \psi_C(Y_C)$$

where $\mathcal{C}_G =$ set of cliques of G.

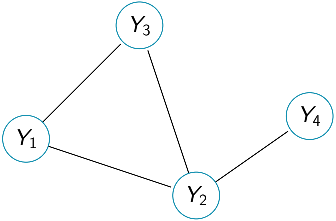
Graphical models

Factorization propriety

The joint distribution P is faithful to the graph G iff

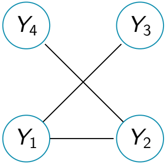
$$P(Y_1, \dots, Y_p) \propto \prod_{C \in \mathcal{C}_G} \psi_C(Y_C)$$

where $\mathcal{C}_G$ = set of cliques of G.



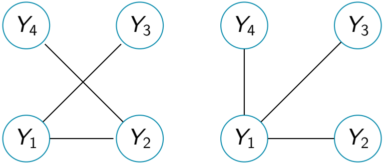
$$P(Y_1, Y_2, Y_3, Y_4) \propto \psi_1(Y_1, Y_2, Y_3) \times \psi_2(Y_3, Y_4)$$

Tree averaging



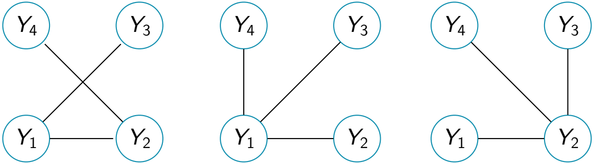
$$P\{T = T_1 | Y\}$$

Tree averaging



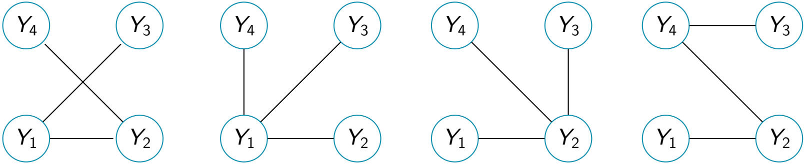
$P\{T = T_1 | Y\}$ $P\{T = T_2 | Y\}$

Tree averaging



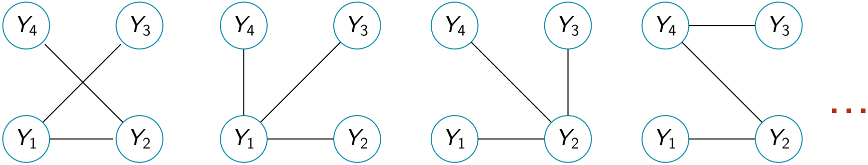
$P\{T = T_1 | Y\}$ $\quad$ $P\{T = T_2 | Y\}$ $\quad$ $P\{T = T_3 | Y\}$

Tree averaging



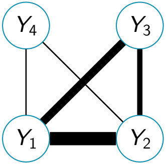
$P\{T = T_1 | Y\}$ $P\{T = T_2 | Y\}$ $P\{T = T_3 | Y\}$ $P\{T = T_4 | Y\}$ ...

Tree averaging



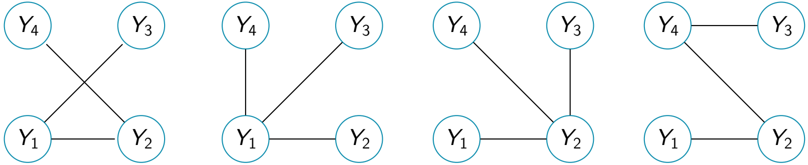
$P\{T = T_1|Y\}$ $P\{T = T_2|Y\}$ $P\{T = T_3|Y\}$ $P\{T = T_4|Y\}$

Compute edge probabilities:



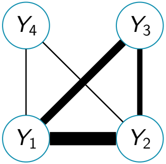
$P\{(j,k) \in T|Y\}$

Tree averaging



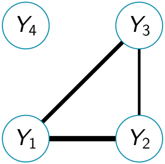
$P\{T = T_1|Y\}$ $P\{T = T_2|Y\}$ $P\{T = T_3|Y\}$ $P\{T = T_4|Y\}$ ...

Compute edge probabilities:



$P\{(j,k) \in T|Y\}$

Thresholding probabilities:



$P\{(j,k) \in T|Y\}$