Inference of the interactions within the pathobiome of *Erysiphe alphitoides*

Sarah Ouadah, Stéphane Robin, Loïc Schwaller
& Corinne Vacher

AgroParisTech

INRA
SCIENCE & IMPACT

5th *Workshop on*
***A**lgorithmic issues for **I**nference in **G**raphical **M**odels*

September 28, 2015

Microbial Interaction Networks

Plan-inhabiting microorganisms can interact with each other forming **complex interaction networks**.

- Direct interactions
  - Predation
  - Parasitism
  - Mutualism, etc
- Indirect interactions
  - Shared resources
  - Shared environments, etc

Photo by Zelimir Borzan, University of Zagreb, Bugwood.org

- Pedunculate oak (*Quercus robur L.*)
- Infected by *Erysiphe alphitoides*
  - Species of fungus
  - Responsible for powdery mildew

- Pathogenicity resulting from **specific interactions with the microbial environment**
  ▷ Pathobiome

Microbial Environment

- Operational taxonomic units (OTUs)
  - **Fungal** (among which *E. alphitoides*) .... 48 OTUs
  - **Bacterial** .... 66 OTUs

114 OTUs

- Abundances measured on **120 leaves** taken from **3 trees**

  highly susceptible / intermediately resistant / strongly resistant

Covariates

- Distance to
  - the base of the branch
  - the tree trunk
  - the ground
- The orientation of the branch .... SW/NE

Problem

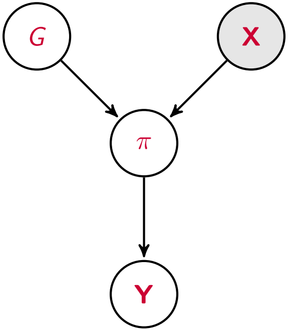
- Interactions can be concealed by **indirect relationships** resulting from **the environment**.

The inference of the interactions must take covariates into account.

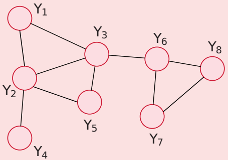
Notations

p	OTUs	$V = \{1, \dots, p\}$
n	leaves	
Y_i^k	abundance of OTU i on leaf k	$1 \leqslant i \leqslant p,\ 1 \leqslant j \leqslant n$
$\mathbf{Y}^k$	abundances of all OTUs on leaf k	$\mathbf{Y} = (\mathbf{Y}^1, \dots, \mathbf{Y}^n)$
α^k	$\in \{1, 2, 3\}$ tree of leaf k	
D^k	orientation of the branch of leaf k	
$\{X_d^k\}_{d=1}^3$	distance of leaf k to the {base of the branch, trunk, ground}	
$\mathbf{X}^k$	$= (\alpha^k, D^k, X_1^k, X_2^k, X_3^k)$	$\mathbf{X} = (\mathbf{X}^1, \dots, \mathbf{X}^n)$

Model



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

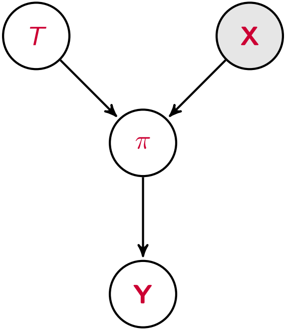


$\mathcal{G} = \{$ decomposable graphs $\}$

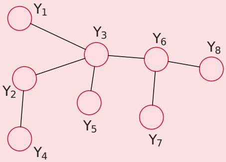
- π distribution for $\mathbf{Y}$
  Markov w.r.t. G
  depending on $\mathbf{X}$

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

Model



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

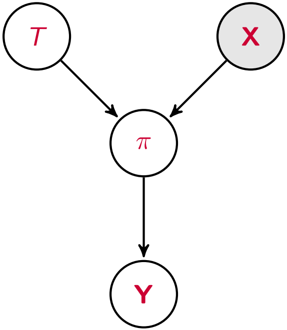


$$\mathcal{T} = \{ \text{ spanning trees } \}$$

- π distribution for **Y**
  Markov w.r.t. T
  depending on **X**

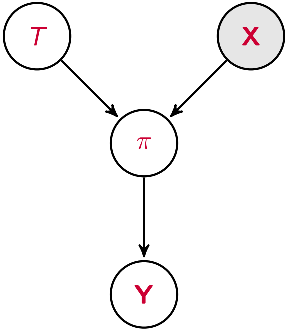
$$\{i,j\} \notin E_T \Leftrightarrow Y_i \perp\!\!\!\perp Y_j | Y_{V \setminus \{i,j\}}$$

Model



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

Model

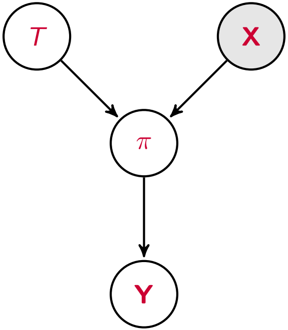


$$T \sim \mathcal{U}(\mathcal{T})$$
$$\pi | T, \mathbf{X} \sim \rho(T, \mathbf{X})$$
$$\mathbf{Y} | \pi \sim \pi$$

Posterior Edge Probabilities

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

Model



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

Posterior Edge Probabilities

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

- Summation over $\mathcal{T}$
- Choice of $\rho(T, \mathbf{X})$

Suppose that

$$p(T|\mathbf{Y},\mathbf{X}) \propto \prod_{\{k,\ell\}\in E_T} \omega_{k\ell}(Y_k, Y_\ell, \mathbf{X})$$

Suppose that

$$p(T|\mathbf{Y},\mathbf{X}) \propto \prod_{\{k,\ell\}\in E_T} \omega_{k\ell}(Y_k, Y_\ell, \mathbf{X})$$

then

$$p(\{i,j\} \in E_T|\mathbf{Y},\mathbf{X}) = \frac{Z^{(ij)}}{Z}$$

with

$$Z = \sum_{T\in\mathcal{T}} \prod_{\{k,\ell\}\in E_T} \omega_{k\ell} \qquad Z^{(ij)} = \sum_{\substack{T\in\mathcal{T}\\ T\ni\{i,j\}}} \prod_{\{k,\ell\}\in E_T} \omega_{k\ell}$$

Matrix-Tree Theorem

- ω weight matrix
- Δ Laplacian matrix

$$\Delta_{k\ell} = \begin{cases} -\omega_{k\ell} & \text{if } k \neq \ell \\ \sum_j \omega_{k\ell} & \text{if } k = \ell \end{cases}$$

Let $\overline{\Delta}_{uv}$ denote the $(u, v)^{th}$ minor of Δ.

Theorem (Matrix-Tree Theorem, Chaiken, 1982)

All $\overline{\Delta}_{uv}$ are equal and the following identity holds

$$\overline{\Delta}_{uv} = \sum_{T \in \mathcal{T}} \prod_{\{k,\ell\} \in E_T} \omega_{k\ell} = Z$$

Matrix-Tree Theorem

$$p(\{i,j\} \in E_T | \mathbf{Y}, \mathbf{X}) = \frac{Z^{(ij)}}{Z}$$

$$\underbrace{\begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}}_{\omega} \xrightarrow{MT} Z \qquad \underbrace{\begin{matrix} & i & & j & \\ i & \cdot & \cdot & 0 & \cdot \\ & \cdot & \cdot & \cdot & \cdot \\ j & 0 & \cdot & \cdot & \cdot \\ & \cdot & \cdot & \cdot & \cdot \end{matrix}}_{\omega^{(ij)}} \xrightarrow{MT} Z - Z^{(ij)}$$

- **Posterior probabilities for the edges**

  Naive implementation
  Computing $Z^{(ij)}$ .... $O(p^3)$
  Total complexity .... $\mathbf{O(p^5)}$

  Generalized MT theorem and some algebra
  Total complexity .... $\mathbf{O(p^3)}$

Hyperdistributions

$\{\rho(T, \mathbf{X})\}_{T \in \mathcal{T}}$ defined on **the set of distributions over $\mathbf{N}^p$**

$\triangleright$ $\rho(T, \mathbf{X})$ defined on $\{\pi |$ Markov w.r.t. $T\}$

Hyperdistributions

$\{\rho(T, \mathbf{X})\}_{T \in \mathcal{T}}$ defined on **the set of distributions over $\mathbf{N}^p$**

▷ $\rho(T, \mathbf{X})$ defined on $\{\pi |$ Markov w.r.t. $T\}$

- required to satisfy some **Markov property** w.r.t. T

▷ Factorisation property for $p(T|\mathbf{Y}, \mathbf{X})$

$$p(T|\mathbf{Y}, \mathbf{X}) \propto \prod_{\{k,\ell\} \in E_T} \omega_{k\ell}(Y_k, Y_\ell, \mathbf{X}, T)$$

Hyperdistributions

$\{\rho(T, \mathbf{X})\}_{T \in \mathcal{T}}$ defined on **the set of distributions over $\mathbf{N}^p$**

▷ $\rho(T, \mathbf{X})$ defined on $\{\pi |$ Markov w.r.t. $T\}$

- required to satisfy some **Markov property** w.r.t. T

▷ Factorisation property for $p(T|\mathbf{Y}, \mathbf{X})$

$$p(T|\mathbf{Y}, \mathbf{X}) \propto \prod_{\{k,\ell\} \in E_T} \omega_{k\ell}(Y_k, Y_\ell, \mathbf{X}, T)$$

- Compatible family over $\mathcal{T}$

▷ Local terms do not depend on T

$$\omega_{k\ell}(Y_k, Y_\ell, \mathbf{X}, T) \longrightarrow \omega_{k\ell}(Y_k, Y_\ell, \mathbf{X})$$

Univariate GLMs & Copula

- $Y_i|\mathbf{X}$ given by **Poisson GLMs**

- Dependence structure of $\mathbf{Y}|\mathbf{X}$ given by a **Gaussian copula**

Univariate GLMs & Copula

- $Y_i|\mathbf{X}$ given by **Poisson GLMs**

$$\begin{cases} Y_i^k & \sim \mathcal{P}(\Lambda_i^k) \\ \log(\Lambda_i^k) & = \mathbf{X}^k \beta_i \end{cases}$$

- Dependence structure of $\mathbf{Y}|\mathbf{X}$ given by a **Gaussian copula**

Univariate GLMs & Copula

- $Y_i|\mathbf{X}$ given by **Poisson GLMs**

$$\begin{cases} Y_i^k & \sim \mathcal{P}(\Lambda_i^k) \\ \log(\Lambda_i^k) & = \mathbf{X}^k \beta_i \end{cases}$$

- Dependence structure of $\mathbf{Y}|\mathbf{X}$ given by a **Gaussian copula**

$\Phi_i^k(\cdot|\beta_i)$ cumulative distribution function of $Y_i^k|\mathbf{X}^k$

$C(\cdot|\Gamma)$ Gaussian copula with precision matrix Γ

$$\Phi(\cdot|\Gamma, \beta) = C(\Phi_1^k(\cdot|\beta_1), \dots, \Phi_p^k(\cdot|\beta_p)|\Gamma)$$

cumulative distribution function of $\mathbf{Y}^k|\mathbf{X}^k$

Univariate GLMs & Copula

- $Y_i|\mathbf{X}$ given by **Poisson GLMs**

$$\begin{cases} Y_i^k & \sim \mathcal{P}(\Lambda_i^k) \\ \log(\Lambda_i^k) & = \mathbf{X}^k \beta_i \end{cases}$$

- Dependence structure of $\mathbf{Y}|\mathbf{X}$ given by a **Gaussian copula**

$\Phi_i^k(\cdot|\beta_i)$ cumulative distribution function of $Y_i^k|\mathbf{X}^k$

$C(\cdot|\Gamma)$ Gaussian copula with precision matrix Γ

$\Phi(\cdot|\Gamma,\beta) = C(\Phi_1^k(\cdot|\beta_1), \dots, \Phi_p^k(\cdot|\beta_p)|\Gamma)$

cumulative distribution function of $\mathbf{Y}^k|\mathbf{X}^k$

- $\rho(T,\mathbf{X})$ induced by a distribution on $\{\Gamma|$ compatible with $T\}$
  a distribution on β

Heuristic

- $Y_i|\mathbf{X}$ given by **Poisson GLMs**

$$\left\{\begin{array}{lcl} Y_i^k & \sim & \mathcal{P}(\Lambda_i^k) \\ \log(\Lambda_i^k) & = & \mathbf{X}^k\beta_i \end{array}\right. \longrightarrow \widehat{\beta}$$

- Dependence structure of $\mathbf{Y}|\mathbf{X}$ given by a **Gaussian copula**

$\Phi_i^k(\cdot|\widehat{\beta}_i)$ cumulative distribution function of $Y_i^k|\mathbf{X}^k$

$C(\cdot|\Gamma)$ Gaussian copula with precision matrix Γ

$$\Phi(\cdot|\Gamma,\widehat{\beta}) = C(\Phi_1^k(\cdot|\widehat{\beta}_1), \ldots, \Phi_p^k(\cdot|\widehat{\beta}_p)|\Gamma)$$

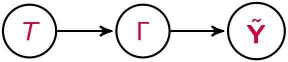
cumulative distribution function of $\mathbf{Y}^k|\mathbf{X}^k$

- $\rho(T)$ induced by a distribution on $\{\Gamma|$ compatible with $T\}$

In Practice

- R_i^k Pearson residual of Y_i^k
  $\tilde{F}_i$ Empirical cdf of $\{R_i^k\}_{k=1}^n$

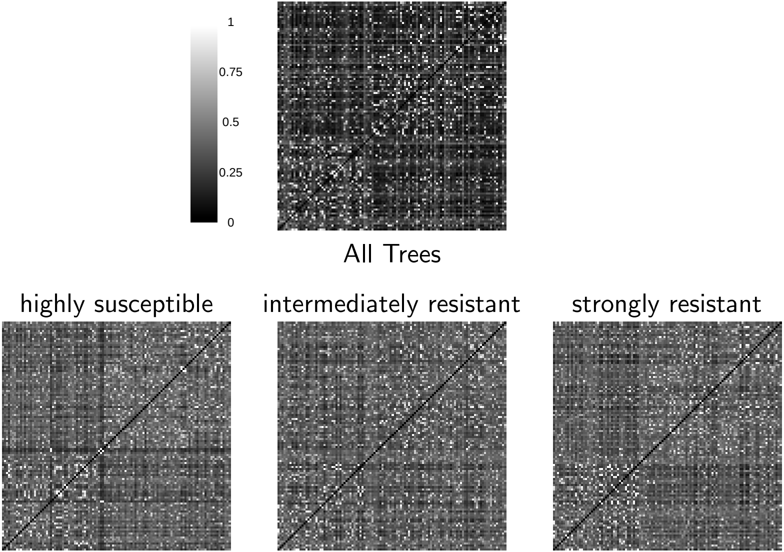
$$\tilde{Y}_i^k = \tilde{F}_i(R_i^k)$$



$$T \sim \mathcal{U}(\mathcal{T})$$

$$\Gamma_{ij}|T \sim \begin{cases} \mathcal{U}([-1;1]) & \text{if } \{i,j\} \in E_T \\ \delta_0 & \text{otherwise} \end{cases}$$

$$\tilde{\mathbf{Y}}|\Gamma \sim C(\cdot|\Gamma)$$


1
0.75
0.5
0.25
0
All Trees
highly susceptible
intermediately resistant
strongly resistant

Interactions of *Erysiphe alphitoides*

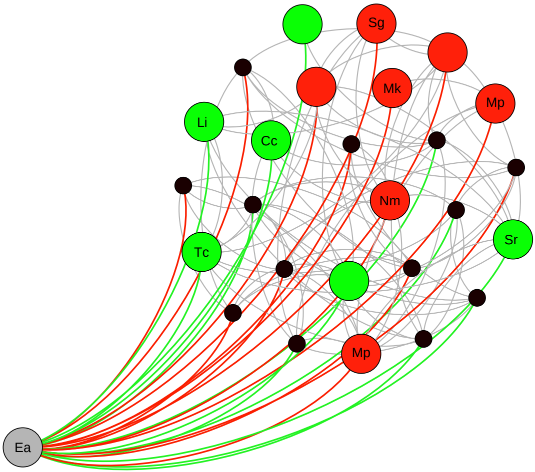
Susceptible tree

Positive edges

Negative edges

Fungal species

Bacterial species



- Exploration of graphs restricted to **tree structures**
  - Posterior edge appearance probabilities
  - Exact & efficient computation
    thanks to the Matrix-Tree theorem

- Covariates
  - No full model
  - Inference performed after fitting univariate GLMs

References

- Seth Chaiken. “A Combinatorial Proof of the All Minors Matrix Tree Theorem”. In: *SIAM Journal on Algebraic Discrete Methods* 3.3 (1982), pp. 319–329.
- A Philip Dawid and Steffen L. Lauritzen. “Hyper Markov Laws in the Statistical Analysis of Decomposable Graphical Models”. In: *The Annals of Statistics* 21.3 (1993), pp. 1272–1317.
- Marina Meilă and Tommi Jaakkola. “Tractable Bayesian learning of tree belief networks”. In: *Statistics and Computing* 16.1 (2006), pp. 77–92.
- L. Schwaller, S. Robin, and M. Stumpf. “Bayesian Inference of Graphical Model Structures Using Trees”. In: *ArXiv e-prints* (Apr. 2015). arXiv: 1504.02723 [stat.ML]. URL: http://arxiv.org/abs/1504.02723.