Network inference from incomplete abundance data

Raphaëlle Momal

Supervision: S. Robin[1] and C. Ambroise[2]
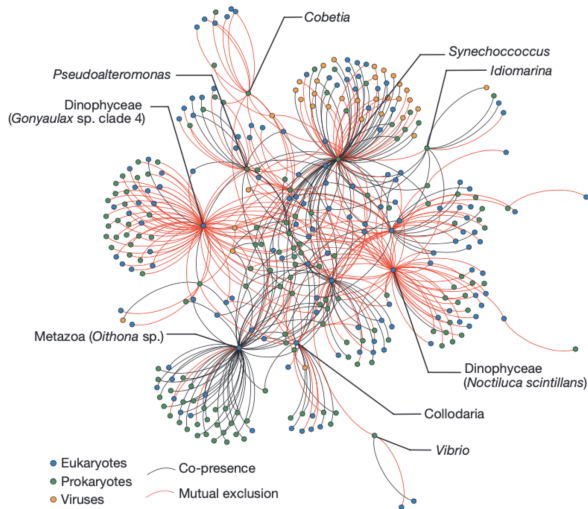
[1]UMR AgroParisTech / INRA MIA-Paris
[2]LaMME, Evry

December 8th, 2020

Species co-occurrence network

Integrated plankton community network related to carbon export at 150m (Guidi et. al, 2016)

# Reasons for species co-occurrence

Two species can co-occur due to:

1. a similar response to the same environmental variable,
2. their response to a third species prensence/abundance (mediator species), even if they do not directly depend on one another,
3. their direct association.

Taking environmental effects into account is paramount, yet not enough to separate (2) from (3).

Simple dependencies

After adjusting for environmental covariates, we obtain (residual) correlations between species.

$$\text{correlation} \neq 0 \iff \text{dependence}$$
$$\text{(Gaussian framework)}$$
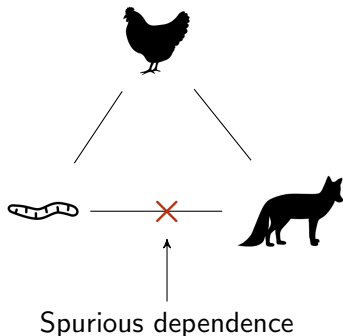
Dependencies can be direct, or indirect/spurious and due to a mediator species (or unaccounted environmental factor).

$\Rightarrow$ Conditional dependencies are always direct links.

Spurious dependence

Interpretation of conditional dependencies

Measure of the dependence link between two species after having controlled for the effect of all others.
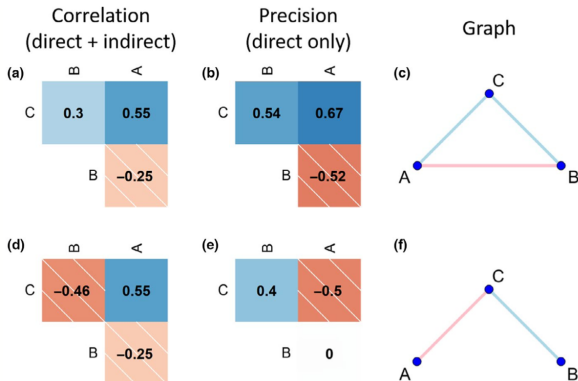
Regression: $Y = \beta_X X + \beta_Z Z + \varepsilon$.

- Y and X are dependent conditionnally on Z $\iff \beta_X \neq 0$.
- Partial correlations quantify this dependence: correlation between the residuals of the regressions of X with Z and of Y with Z ($cos(\varphi)$).

Graphically: are the projections of X and Y on the hyperplan of Z orthogonal?

Two scenarios

Correlation (direct + indirect) / Precision (direct only) / Graph
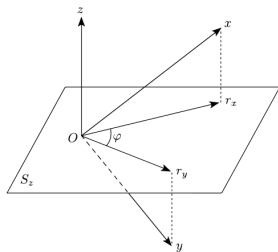
- 1^{rst} line: $A \sim B$, 2nd line: $A \nsim B$.

- Same $Cor(A, B)$ in both scenarios.

- Only conditional dependences can separate scenarios.

Toy-example with Gaussian data (Popovic et al., 2019)

Aim of network inference from abundance data

EFI	ELA	GDE	GME	date	site
71	1	5	6	apr93	km03
118	2	3	0	apr93	km03
69	0	6	2	apr93	km03
56	0	0	0	apr93	km03
0	1	1	0	apr93	km17
0	0	2	0	apr93	km17
⋮	⋮	⋮	⋮	⋮	⋮
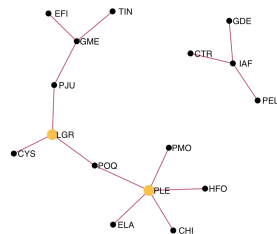
(a) species abundances **Y**   (b) covariates **X**   (c) **G**

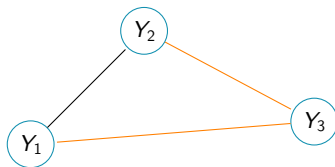Data sample from the Fatala river dataset (Baran 1995).

Incomplete data: a missing species/covariate

Marginalization of graphs:

Complete graph:

$\Longrightarrow$
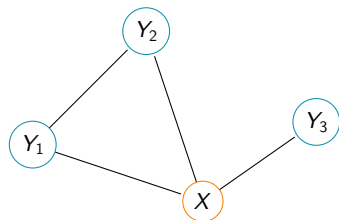
Marginal graph:

Spurious edges leading to
wrong interpretation

X is a missing actor.

Incomplete abundance data

EFI	ELA	GDE	GME
71	1	5	6
118	2	3	0
69	0	6	2
56	0	0	0
0	1	1	0
0	0	2	0
⋮	⋮	⋮	⋮

date	site
apr93	km03
apr93	km03
apr93	km03
apr93	km03
apr93	km17
apr93	km17
⋮	⋮

$\Longrightarrow$?

(a) incomplete abundances $\mathbf{Y}$ (b) incomplete $\mathbf{X}$ (c) complete $\mathbf{G}$
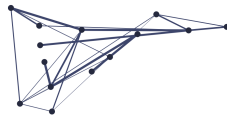
Example with $x \sim \mathcal{N}(1, 1)$

True graph
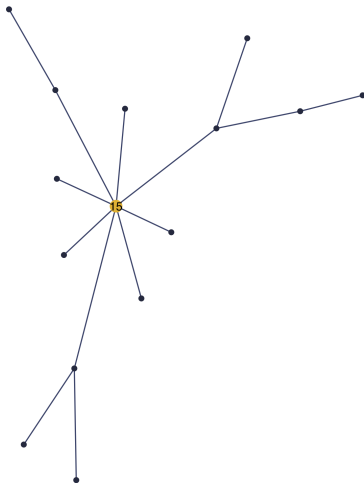
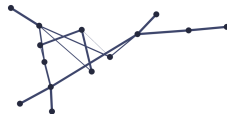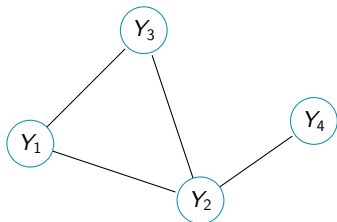**Null fit**

Fit with covariate

Covariate and one missing actor

Mathematical framework

  i Graphical Models

 ii Graph exploration with trees

iii Poisson log-Normal model

Graphical Models

Global Markov:
Y_2 separates Y_3 from $Y_4 \Rightarrow Y_3 \perp\!\!\!\perp Y_4 \mid Y_2$.

Hammersley-Clifford:
Strictly positive and continuous density f:
f global Markov $\iff f(\boldsymbol{Y}) = \prod_{c \in \mathcal{C}} \psi(Y_c)$.
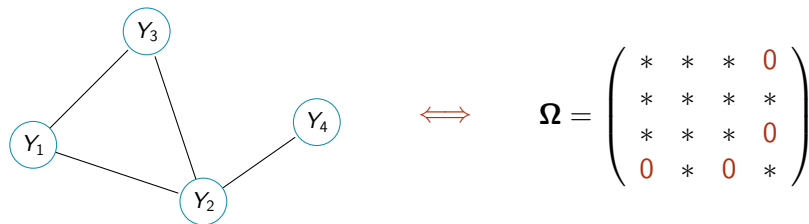
Here $\mathcal{C} = \big\{\{1, 2, 3\}, \{2, 4\}\big\}$:

$$f(\boldsymbol{Y}) = \psi(Y_1, Y_2, Y_3) \times \psi(Y_2, Y_4)$$

Gaussian Graphical Models (GGM)

Let $\boldsymbol{Y} \sim \mathcal{N}(\mu, \boldsymbol{\Sigma})$ with precision matrix $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1} = (\omega_{jk})_{jk}$:

$$f(\boldsymbol{Y}) \propto \prod_{j,k,,\omega_{jk} \neq 0} \exp(-Y_k \omega_{jk} Y_j / 2).$$

Faithful Markov property:

$$\Longleftrightarrow \qquad \boldsymbol{\Omega} = \begin{pmatrix} * & * & * & 0 \\ * & * & * & * \\ * & * & * & 0 \\ 0 & * & 0 & * \end{pmatrix}$$

Gaussian precision terms and conditional dependence

Regression : $X \sim \mathcal{N}(\mu, \boldsymbol{\Omega}^{-1})$. In the regression $X_j = \sum_{k \neq j} \theta_{jk} X_k + \varepsilon_j$, it holds that $\varepsilon_j \sim \mathcal{N}(0, \omega_{jj}^{-1})$ and $\theta_{jk} = -\omega_{jk}/\omega_{jj}$. Thus $\omega_{jk} \propto \theta_{jk}$

Covariance/Correlation matrix

$\downarrow$ Inverse

Precision matrix $(\omega_{jk})_{jk}$

$\downarrow$ -Normalized

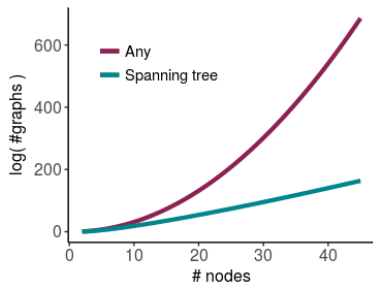Partial correlations $(\rho_{jk} = -\omega_{jk}/\sqrt{\omega_{jj}\omega_{kk}})$

partial correlation/precision $\neq 0 \iff$ conditional dependence
(Gaussian framework)

Exploring the graph space

Aim: infer **G**.

Very large space to explore: $\#\mathcal{G}_p = 2^{\frac{p(p-1)}{2}}$

Spanning trees are sparse and simple structures:

- no loops
- $(p-1)$ edges

Much smaller space to explore:

$$\#\mathcal{T}_p = p^{(p-2)}$$

Summing over spanning trees

Let $\mathbf{W} = (w_{jk})_{jk}$ be a matrix with null diagonal and positive entries, and $\mathbf{Q}$ its Laplacian:
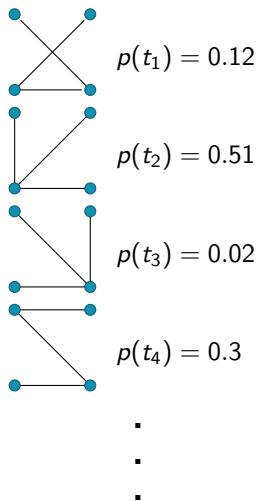
$$[\mathbf{Q}]_{jk} = \begin{cases} \sum_k w_{jk} & \text{if } j = k \\ -w_{jk} & \text{otherwise} \end{cases}$$

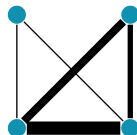### Matrix-tree Theorem (Chaiken and Kleitman, 1978)

All minors of $\mathbf{Q}$ are equal, and for any $1 \leq u, v, \leq p$:

$$|\mathbf{Q}^{uv}| = \sum_{T \in \mathcal{T}} \prod_{jk \in T} w_{jk}$$

Allows to sum over $p^{(p-2)}$ trees in $\mathcal{O}(p^3)$ operations.

Exploring $\mathcal{T}$ with tree averaging

$p(t_1) = 0.12$

$p(t_2) = 0.51$
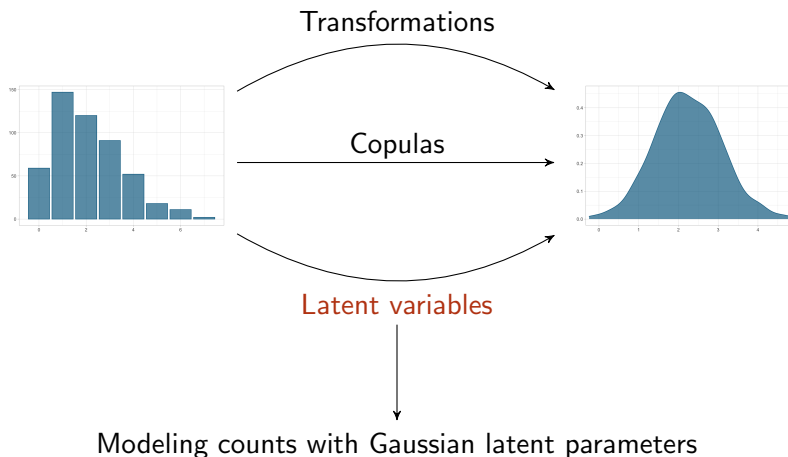
$p(t_3) = 0.02$

$p(t_4) = 0.3$

Network inference
$=$ edge probabilities:

$$\mathbb{P}\{k\ell \in T\} = \sum_{\substack{T \in \mathcal{T} \\ k\ell \in T}} p(T)$$

$$p(T) \propto \prod_{kl \in T} w_{kl}$$

Getting back to Gaussian data

Transformations

Copulas

Latent variables

Modeling counts with Gaussian latent parameters

Poisson log-normal model

PℓN model (Aitchison and Ho, 1989) for sample i and species j:

$$\boldsymbol{Z}_i \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

$$Y_{ij} \mid \boldsymbol{Z}_i \sim \mathcal{P}(\exp(\underbrace{o_{ij} + \boldsymbol{x}_i^\mathsf{T}\boldsymbol{\theta}_j}_{\text{fixed}} + Z_{ij})).$$

- Latent variables are iid, observed data are independent conditionally on the $\boldsymbol{Z}_i$.
- A generalized multivariate linear mixed model : fixed abiotic and random biotic effects.
- Variational estimation algorithm (PLNmodels, Chiquet et al. (2018))

Network inference from incomplete counts

General model

- Assume a random tree dependency structure T

- Dependence structure in Gaussian layer $\mathbf{Z}$

- Distribution for counts $\mathbf{Y}$ accounting for covariates/offsets

T

$\downarrow$

$\mathbf{Z}$

$\downarrow$

$\mathbf{Y}$

- Matrix Tree Theorem

- Gaussian Graphical Model

- Poisson log-normal model

$P\ell N$ model with tree-shaped Gaussian parameters

$$\begin{cases} T \sim \prod_{kl \in T} \beta_{kl}/B, \\[2mm] \boldsymbol{Z}_i \mid T \sim \mathcal{N}(0, \boldsymbol{\Omega}_T) \\[2mm] Y_{ij} \mid \boldsymbol{Z}_i \sim \mathcal{P}(\exp(o_{ij} + \boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{\theta}_j + Z_{ij})). \end{cases}$$
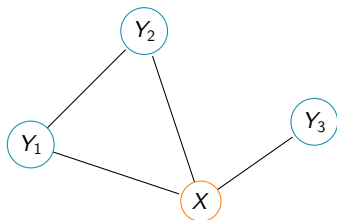
Gaussian mixture with p^{p-2} components:

$$p(\boldsymbol{Z}) = \sum_{T \in \mathcal{T}} p(T)\mathcal{N}(\boldsymbol{Z} \mid T; 0, \boldsymbol{\Omega}_T).$$

Decomposition of the likelihood:

$$p(\boldsymbol{Y}, \boldsymbol{Z}, T) = p_{\boldsymbol{\beta}}(T)\, p_{\Omega_T}(\boldsymbol{Z} \mid T)\, p_{\boldsymbol{\theta}}(\boldsymbol{Y} \mid \boldsymbol{Z}).$$

Marginalization of graphs

Complete graph:

$\Longrightarrow$

Marginal graph:
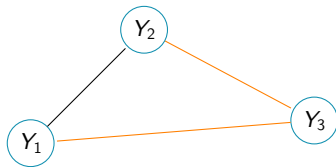
Spurious edges leading to wrong interpretation

X is a missing actor.

Added hidden Gaussian parameters

$$\boldsymbol{Z} \mid T \sim \mathcal{N}(0, \boldsymbol{\Omega}_T^{-1})$$

$\boldsymbol{Z}$: $n \times p$
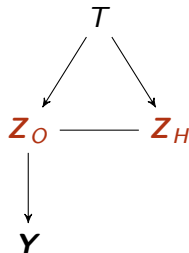
$$(\boldsymbol{Z}_O, \boldsymbol{Z}_H) \mid T \sim \mathcal{N}(0, \boldsymbol{\Omega}_T^{-1})$$

$\boldsymbol{Z}_O$: $n \times p$
$\boldsymbol{Z}_H$: $n \times r$ $\qquad p' = p + r.$

Variational EM algorithm

Finding distribution $q(\boldsymbol{H}) \approx p(\boldsymbol{H} \mid \boldsymbol{Y})$:

- Restricting the search space to a family Q,
- Choosing q with smallest distance to $p(\boldsymbol{H} \mid \boldsymbol{Y})$.
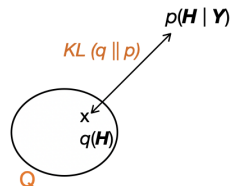
Doing so maximizes a lower-bound of the log-likelihood:

$$\mathcal{J}(\Theta; q) = \log p_\Theta(\boldsymbol{Y}) - KL(q(\boldsymbol{H}) \parallel p_\Theta(\boldsymbol{H} \mid \boldsymbol{Y})).$$

Variational EM algorithm

VE step: $q^{t+1} = \mathrm{argmax}_{q \in Q} \left\{ \mathcal{J}(\Theta^t; q^t) \right\} = \mathrm{argmin}_{q \in Q} \left\{ KL(q^t \parallel p_{\Theta^t}) \right\}$

M step: $\Theta^{t+1} = \mathrm{argmax}_\Theta \left\{ \mathcal{J}(\Theta; q^{t+1}) \right\}$

Variational distribution

Two hidden variables: $\boldsymbol{Z} = (\boldsymbol{Z}_O, \boldsymbol{Z}_H)$ and T.

$$q(\boldsymbol{Z}, T) = h(\boldsymbol{Z}) \, g(T).$$

$h(\boldsymbol{Z})$: Product (independence of samples i) of Gaussians:

$$h(\boldsymbol{Z}) = \prod_i \mathcal{N}_{p+r}(\boldsymbol{Z}_i; \widetilde{\boldsymbol{m}}_i, \widetilde{\boldsymbol{s}}_i)$$

$g(T)$: Mean-field approximation:

$$g(T) \propto \exp\{\mathbb{E}_h[\underbrace{\log p_{\boldsymbol{\beta}}(T) + \log p_{\boldsymbol{\Omega}}(\boldsymbol{Z} \mid T)}_{\text{Factorizes on the edges of T}}]\}$$

$$g(T) \propto \prod_{kl \in T} \widetilde{\beta}_{kl}$$

Variational distribution

Two hidden variables: $\boldsymbol{Z} = (\boldsymbol{Z}_O, \boldsymbol{Z}_H)$ and T.
$$q(\boldsymbol{Z}, T) = h(\boldsymbol{Z})\, g(T).$$

$h(\boldsymbol{Z})$: Product (independence of samples i) of Gaussians:
$$h(\boldsymbol{Z}) = \prod_i \mathcal{N}_{p+r}(\boldsymbol{Z}_i; \widetilde{\boldsymbol{m}}_i, \widetilde{\boldsymbol{s}}_i)$$

$g(T)$: Mean-field approximation:
$$g(T) \propto \exp\{\mathbb{E}_h[\underbrace{\log p_{\boldsymbol{\beta}}(T) + \log p_{\Omega}(\boldsymbol{Z} \mid T)}_{\text{Factorizes on the edges of T}}]\}$$

$$g(T) \propto \prod_{kl \in T} \widetilde{\beta}_{kl}$$

Variational parameters: $\quad \widetilde{\boldsymbol{M}} = (\widetilde{\boldsymbol{M}}_O, \widetilde{\boldsymbol{M}}_H), \quad \widetilde{\boldsymbol{S}} = (\widetilde{\boldsymbol{S}}_O, \widetilde{\boldsymbol{S}}_H), \quad \widetilde{\boldsymbol{\beta}}$
$\qquad\qquad\qquad\qquad\quad n \times p', \qquad\qquad\quad n \times p', \qquad\qquad p'^2$

Proposed algorithm

PLNmodels: Parameters regarding the observed part: $\widehat{\boldsymbol{\theta}}$, $\widetilde{\boldsymbol{M}}_O$, $\widetilde{\boldsymbol{S}}_O$

- Fixed for further computations.

VE step: Update variational parameters: $\widetilde{\boldsymbol{M}}_H^{t+1}$, $\widetilde{\boldsymbol{S}}_H^{t+1}$, $\widetilde{\boldsymbol{\beta}}^{t+1}$

- Given by shapes of g and h distributions.

M step: Update model parameters: $\boldsymbol{\Omega}_T^{t+1}$, $\boldsymbol{\beta}^{t+1}$

- $\boldsymbol{\Omega}_T$: adaptation of ML estimators (Lauritzen, 1996).
- β_{jk}: Kirshner (2008); Meilă and Jaakkola (2006) with numerical control.

M step: $\mathbf{\Omega}_T$

$\{\mathbf{\Omega}_T, T \in \mathcal{T}\}$ involves a very large number of parameters:

$$\text{(size of } \mathcal{T}) \times \text{(size of } \mathbf{\Omega}_T) = \underbrace{p'^{p'-2} \times p'(p'-1)/2}_{>10^{25} \text{ for 20 nodes}}.$$

Using Lauritzen's ML estimator: $p'(p'-1)/2$ estimators (only one matrix!).

Lauritzen's ML estimator

In a GGM with a chordal graph **G** (cliques $\mathcal{C}$, separators $\mathcal{S}$ with multiplicities $\nu(S)$), SSD the sum of squares matrix.

General Lauritzen's MLE

$$\widehat{\Omega}_{\mathbf{G}}^{MLE} = n\Big(\sum_{C \in \mathcal{C}} [(SSD_C)^{-1}]^{p'} - \sum_{S \in \mathcal{S}} \nu(S)[(SSD_S)^{-1}]^{p'} \Big)$$

- The general SSD matrix do not depend on **G**.
- The estimator uses SSD according to the graph structure.

Lauritzen's ML estimator

In a GGM with a chordal graph **G** (cliques $\mathcal{C}$, separators $\mathcal{S}$ with multiplicities $\nu(S)$), SSD the sum of squares matrix.

General Lauritzen's MLE

$$\widehat{\mathbf{\Omega}}_{\mathbf{G}}^{MLE} = n\Big(\sum_{C \in \mathcal{C}}[(SSD_C)^{-1}]^{p'} - \sum_{S \in \mathcal{S}}\nu(S)[(SSD_S)^{-1}]^{p'}\Big)$$

If **G** is a tree $T \in \mathcal{T}$:

- T is chordal.
- Cliques are edges: inverses of 2×2 matrices.
- Separators are nodes: $\mathcal{S} = \{1, ..., p'\}$.
- $\nu(k) = deg(k) - 1$.

Update of $\mathbf{\Omega}_T$

We define:

$$SSD = \mathbb{E}_h[\mathbf{Z}^\intercal \mathbf{Z} \mid \mathbf{Y}] = \widetilde{\mathbf{M}}^\intercal \widetilde{\mathbf{M}} + diag\big(\sum_i \widetilde{\mathbf{s}}_i\big).$$

Tree simplification of Lauritzen's formula:

$$\omega_{Tjk}^{t+1} = \mathbb{1}\{jk \in T\} \left(\frac{-ssd_{jk}^t/n}{1 - (ssd_{jk}^t/n)^2} \right),$$

$$\omega_{Tkk}^{t+1} = 1 - \sum_j (ssd_{jk}^t/n) \times \omega_{Tjk}^{t+1}.$$

The estimates ω_{Tjk} are common to all trees sharing the edge jk:
estimating $\{\mathbf{\Omega}_T, T \in \mathcal{T}\}$ amounts to estimating $p'(p'-1)/2$ quantities.

M step: β

We derive the log-likelihood to get a closed form:

$$\widehat{\beta}_{jk} = \frac{\mathbb{P}_g\{jk \in T\}}{M(\beta)_{jk}}$$

- $\mathbb{P}_g\{jk \in T\} = \displaystyle\sum_{\substack{T \in \mathcal{T} \\ jk \in T}} g(T)$ (in $\mathcal{O}(p'^3)$ thanks to Kirshner (2008)).
- $M(\beta)$ is a $p' \times p'$ matrix defined in Meilă and Jaakkola (2006) as a function of the inverse Laplacian minor $(\mathbf{Q}(\beta)^{11})^{-1}$.

This fixed-point problem is solved using optimization, with a gradient ascent procedure.

Numerical stability and the Matrix Tree Theorem

$\sum_{T \in \mathcal{T}} \prod_{jk \in T} \beta_{jk}$ *computable for any p':*

- Upper and lower bounds for $\boldsymbol{\beta}$, which depend on p' and the machine precision.

If $\boldsymbol{\beta}$ has too many high values, $\mathbf{Q}(\beta)^{11}$ can become numerically non positive-definite (conditioning$< 1e - 16$):

- The mean value is controlled with a sum constraint.
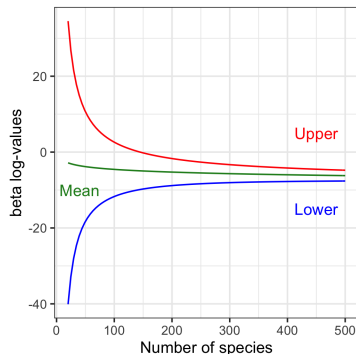
These constraints on the optimization improve the algorithm's numerical stability and allows larger networks.

nestor (Network inference from Species counTs with missing actORs)

This VEM algorithm is implemented in the R package `nestor`.

Sensitive point: choosing a set of initial neighbors for the missing actor(s). Several implemented propositions: sparse PCA, SBM (blockmodels), mclust...
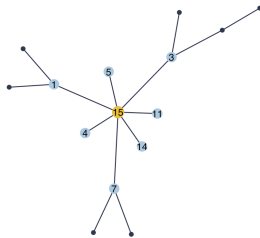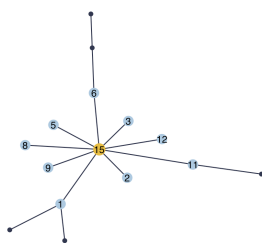
Interesting outputs :

- Matrix of edges probabilities P
- Completed matrices of means and variances M and S

Simulation design

Count datasets:

- 300 scale-free graphs with 15 nodes, their highest degree node is hidden ($r = 1$, $p = 14$).
- Count datasets are simulated under the PLN model.
- Cases are separated by influence of the missing actor:
 Major (deg ≥ 8), Medium ($5 <$ deg ≤ 7) and Minor (deg ≤ 5).
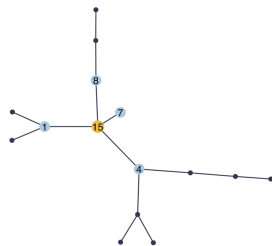
Experiment

Initialization: A set of four initial cliques is proposed, which rely on sparse PCA. Nestor is run with each one and the best run (best lower bound) is kept.

Measures:

- Global inference: AUC compares P to **G**.
- Position of the missing actor: Precision and Recall of the inferred neighbors ($\mathbb{1}\{P_{H\bullet} \geq 0.5\}$ vs. $\mathbf{G}_{H\bullet}$)
- Reconstruction of the missing actor: $Cor(M_H, Z_H)$
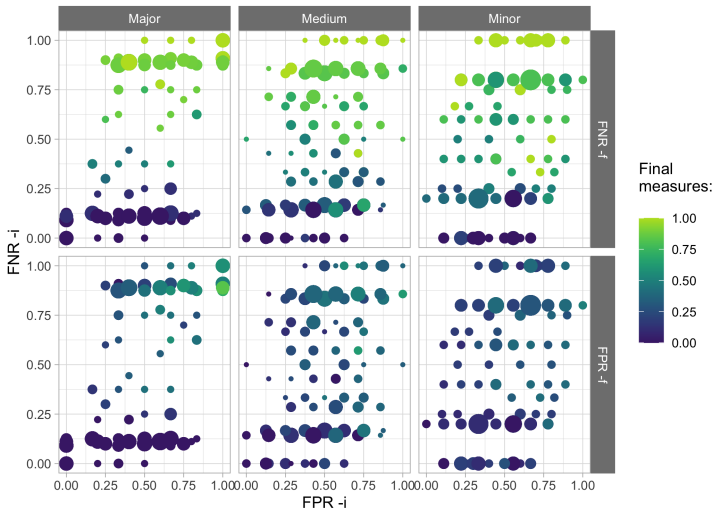
Reconstruction of the missing actor
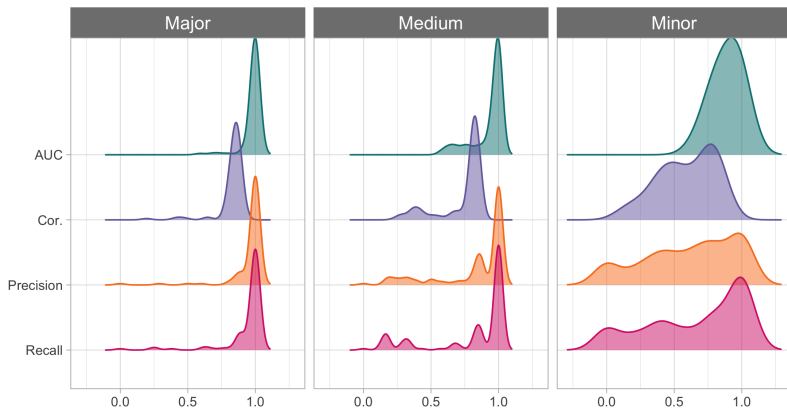
The decrease in performance is actually due to poorer initialization in *Minor* cases.
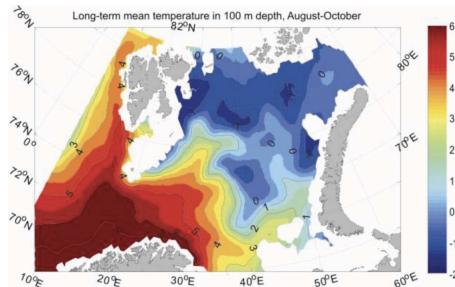
Initialize with more potential neighbors
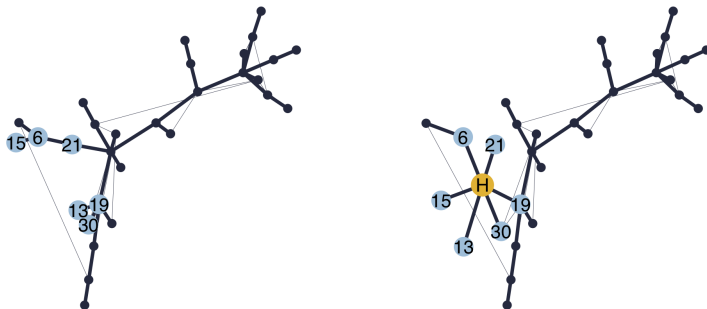
$FNR= FN/P$
$FPR=FP/N$

Barent's sea fishes

- **Y**: abundances of 30 fish species in 89 sites,

- **X**: latitude, longitude, depth and temperature,

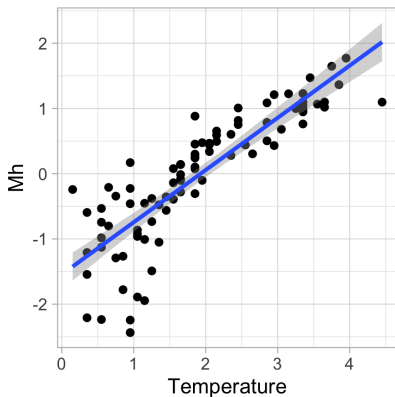- **O**: total detections per site.

Stiansen et al. (2009)

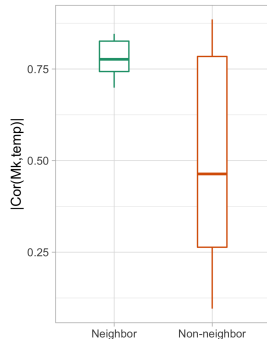$\Rightarrow$ Fit with no covariates.

Barent's fishes networks

Left: observed network (3.3 mins). *Right*: network inferred with one missing actor: H (5.0 mins).

Relationship with temperature

$Cor(Mh, Temp) = 0.85$.

Direct neighbors are more linked to the temperature than other species.

Conclusion

Model
- A probabilistic model for the inference of conditional dependency networks from incomplete abundance data.
- Accounts for covariates, offsets and missing actors.

Inference
- A variational EM algorithm which combines the GGM framework flexibility and spanning trees algebraic properties.
- Outputs edges probabilities and insights on the missing actor's values

Contributions

Articles

- Momal R., Robin S., and Ambroise C. . *"Tree-based inference of species interaction networks from abundance data."* Methods in Ecology and Evolution 11.5 (2020): 621-632.

- Momal R., Robin S., and Ambroise C. . *"Accounting for missing actors in interaction network inference from abundance data."* arXiv preprint arXiv:2007.14299 (2020).

R packages

- `EMtree`: `https://rmomal.github.io/EMtree/`.

- `nestor` (Network inference from Species counTs with missing actORs): `https://rmomal.github.io/nestor`.

Perspectives

New PostDoc position at MetGenoPolis, INRAe.

Direct
- Compute the partial correlations.
- Improve the scalability.
- Simulations with other data models and dependency structures
- Model selection method (best r, and best probability threshold).

Mid-term
- Robustness to new data (error quantification).
- Microbial guildes clustering.

Long-term Network comparison.

Thank you!

raphaelle.momal@agroparistech (expires end of December)
raphaelle.momal@inrae.fr (will soon work)

References I

Aitchison, J. and Ho, C. (1989). The multivariate Poisson-log normal distribution. *Biometrika*, 76(4):643–653.

Chaiken, S. and Kleitman, D. J. (1978). Matrix tree theorems. *Journal of combinatorial theory, Series A*, 24(3):377–381.

Chiquet, J., Mariadassou, M., and Robin, S. (2018). Variational inference for probabilistic poisson pca. *The Annals of Applied Statistics*, 12(4):2674–2698.

Kirshner, S. (2008). Learning with tree-averaged densities and distributions. In *Advances in Neural Information Processing Systems*, pages 761–768.

Lauritzen, S. L. (1996). *Graphical Models*. Oxford Statistical Science Series. Clarendon Press.

Meilă, M. and Jaakkola, T. (2006). Tractable bayesian learning of tree belief networks. *Statistics and Computing*, 16(1):77–92.

Popovic, G. C., Warton, D. I., Thomson, F. J., Hui, F. K. C., and Moles, A. T. (2019). Untangling direct species associations from indirect mediator species effects with graphical models. *Methods in Ecology and Evolution*, 10(9):1571–1583.

Stiansen, J. E., Korneev, O., Titov, O. V., Arneberg, P., Filin, A., Hansen, J., Høines, Å. S., and Marasaev, S. (2009). Joint norwegian-russian environmental status 2008: report on the barents sea ecosystem, part i-short version.
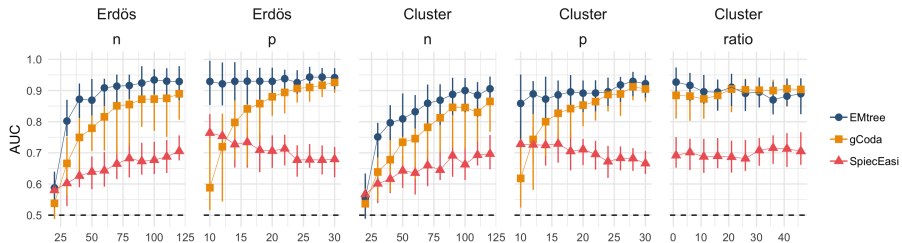
Network inference methods comparison
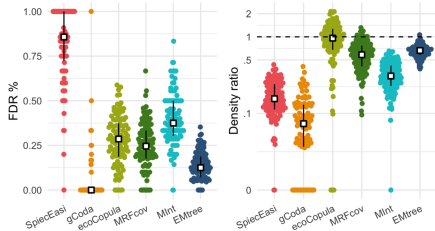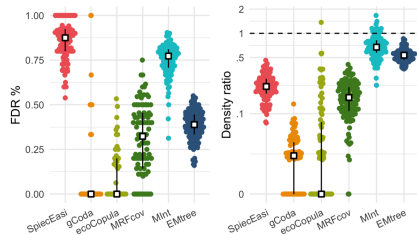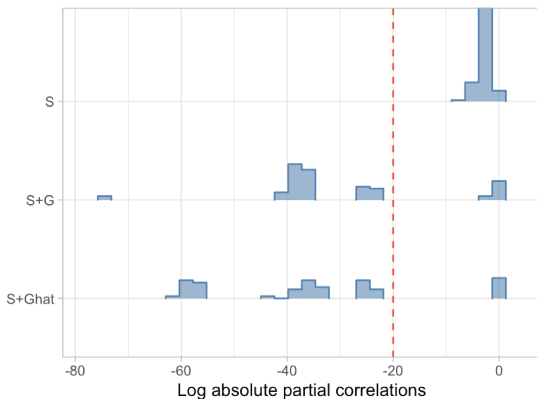
Edges scoring comparison

Signs and strengths of interactions

$$\rho_{jk} = \frac{-\omega_{jk}}{\sqrt{\omega_{kk}\omega_{jj}}}$$

S: sample covariance matrix of $\mathbf{Z}$.
$\widehat{S}$: fitted covariance matrix (ggm R package)

Log absolute partial correlations

$\widehat{S} = S$:

	-1	0	1
-1	5	45	0
1	0	48	7

$\widehat{S} = f(S, \mathbf{G})$:

	-1	0	1
-1	5	0	0
0	0	93	0
1	0	0	7

$\widehat{S} = f(S, \widehat{\mathbf{G}})$:

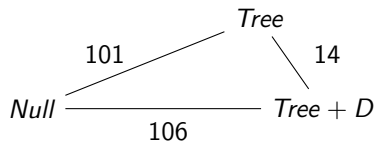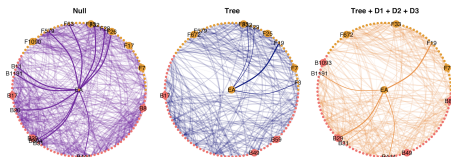|	-1	0	1
-1	4	0	0
0	1	93	2
1	0	0	5

Signs and strengths of interactions
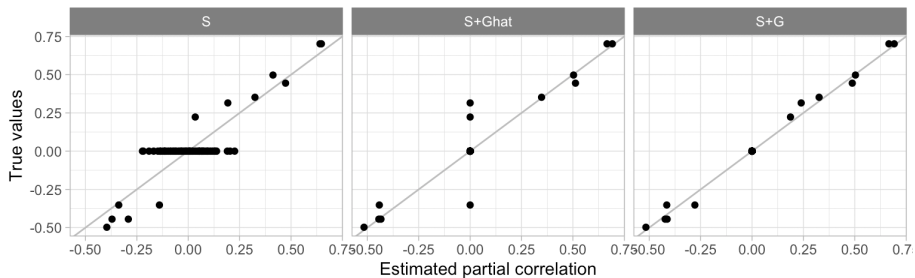
Network comparison

$$D(p_{\beta^A}, p_{\beta^B}) = \frac{1}{2}\left[KL\left(p_{\beta^B} \parallel p_{\beta^A}\right) + KL\left(p_{\beta^A} \parallel p_{\beta^B}\right)\right]$$

$$= \sum_{kl} \log(\beta_{kl}^A/\beta_{kl}^B)\left(\frac{P_{kl}^A - P_{kl}^B}{2}\right)$$

Oak dataset:

Lauritzen's notation

For any square matrix **A**:

$$([\mathbf{A}_B]^p)_{ij} = \left\{ \begin{array}{cl} a_{ij} & \text{if } \{i,j\} \in B, \\ 0 & \text{if } \{i,j\} \notin B. \end{array} \right.$$

$$\mathbf{A} = \begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \end{pmatrix} \quad \Rightarrow \quad [\mathbf{A}_{\{2,3\}}]^3 = \begin{pmatrix} 0 & * & * \\ 0 & * & * \\ 0 & 0 & 0 \end{pmatrix}$$

The M matrix

Lemma (Meilă and Jaakkola, 2006)

$\mathbf{Q}^{pp}$ is the Laplacian matrix $\mathbf{Q}$ to which the the last column and row were removed. M is then defined as follows:

$$[M]_{jk} = \begin{cases} [(\mathbf{Q}^{pp})^{-1}]_{jj} + [(\mathbf{Q}^{pp})^{-1}]_{kk} - 2[(\mathbf{Q}^{pp})^{-1}]_{jk} & 1 \leq j, k < p \\ [(\mathbf{Q}^{pp})^{-1}]_{jj} & k = p, 1 \leq j < p \\ 0 & k = j \end{cases}$$

Prevent numerical issues

The Laplacian matrix $\mathbf{Q}$ must be positive definite, which calls for some numerical control of the weights $\boldsymbol{\beta}$ and $\widetilde{\boldsymbol{\beta}}$.
Weights $\boldsymbol{\beta}$ are controlled with bounds and sum constraints. The same cannot be done for the variational weights as they depend on the number of available samples n.
We define a tempering parameter α:

$$\log \widetilde{\beta}_{kl} = \log \beta_{kl} - \alpha(\frac{n}{2} \log |\widehat{\boldsymbol{R}}_{Tkl}| + \widehat{\omega}_{Tkl}[M^\mathsf{T} M]_{kl}).$$