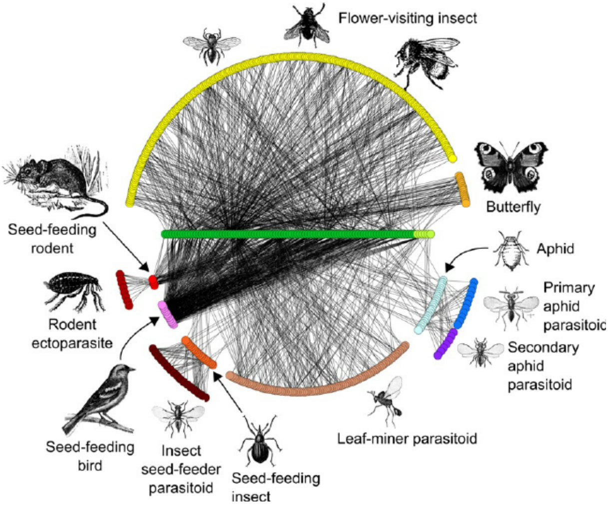
Inference of species interaction networks from incomplete data

Raphaëlle Momal
Supervision: S. Robin[1] and C. Ambroise[2]

[1]UMR AgroParisTech / INRA MIA-Paris
[2]LaMME, Evry

November 12th, 2020

Network example in ecology



Pocock et. al 2012

- Tool to better understand species interactions, eco-systems organizations
- Allows for resilience analyses, pathogens control, ecosystem comparison, response prediction...

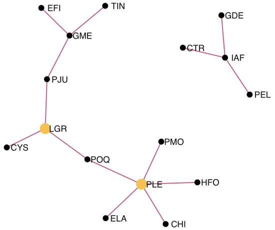
Aim of network inference from abundance data

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
⋮	⋮	⋮	⋮

(a) species abundances $\mathbf{Y}$

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

(b) covariates $\mathbf{X}$



(c) $\mathbf{G}$

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

- Unknown underlying structure.
- Unobserved interaction data.

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
$\vdots$	$\vdots$	$\vdots$	$\vdots$

(a) incomplete abundances $\mathbf{Y}$

date	site
apr93	km03
apr93	km03
apr93	km03
apr93	km03
apr93	km17
apr93	km17
$\vdots$	$\vdots$

(b) incomplete $\mathbf{X}$

$\Longrightarrow$?

(c) $\mathbf{G}$

Mathematical framework

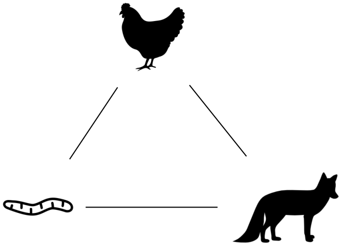
i Graphical Models

ii Graph exploration with trees

iii Poisson log-Normal model

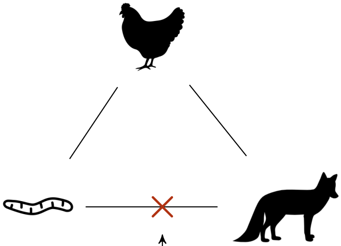
Which statistical link?

Dependence?



Which statistical link?

Dependence?

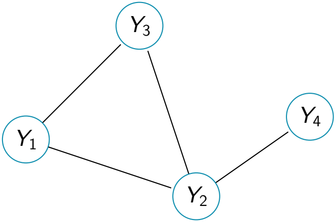


Spurious dependence

Conditional dependence:

- Only direct links: less links.
- Probabilistic background $p(a, b \mid c) = p(a \mid c)\, p(b \mid c)$.
- Possible to 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:

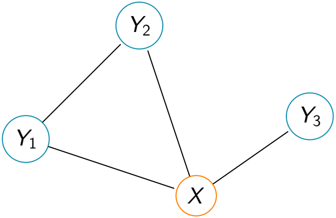
f global Markov $\iff f(\boldsymbol{Y}) = \prod_{c \in \mathcal{C}} \psi(Y_c)$.

Here $\mathcal{C} = \big\{\{1,2,3\},\{2,4\}\big\}$:

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

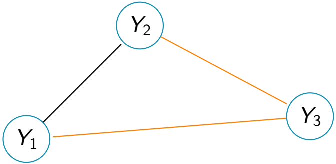
Marginalization of graphs

Complete graph:



$\Longrightarrow$

Marginal graph:



Spurious edges leading to wrong interpretation

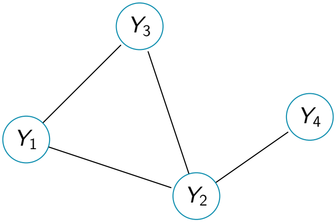
X is a covariate or a species unaccounted for in the model.

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:

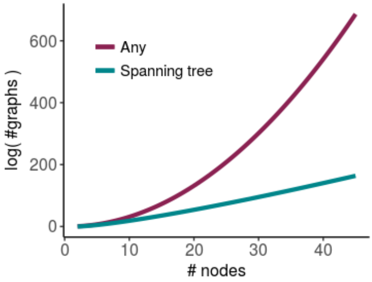


$$\iff \qquad \boldsymbol{\Omega} = \begin{pmatrix} * & * & * & 0 \\ * & * & * & * \\ * & * & * & 0 \\ 0 & * & 0 & * \end{pmatrix}$$

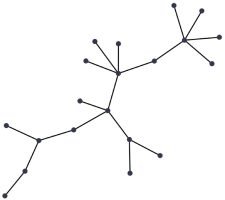
Exploring the graph space

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:

$$[\mathbf{Q}]_{jk} = \begin{cases} \sum_k w_{jk} & \text{if } j = k \\ -w_{jk} & \text{otherwise} \end{cases}$$

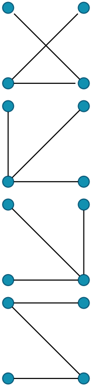
Matrix-tree Theorem (Chaiken and Kleitman, 1978)

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



$p(t_1) = 0.12$

$p(t_2) = 0.51$

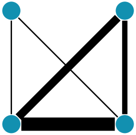
$p(t_3) = 0.02$

$p(t_4) = 0.3$

$\vdots$

Network inference

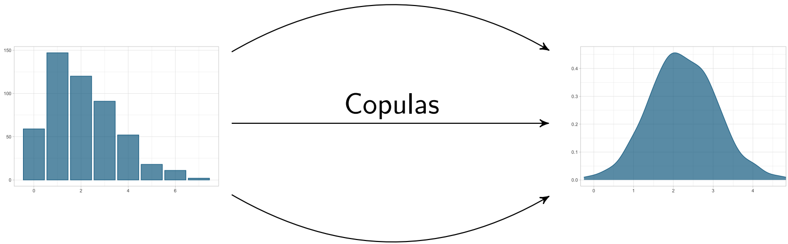
= edge probabilities:



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

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

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:

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

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

- 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 counts

i Model
ii Inference
iii Illustration

General model

- Assume a random tree dependency structure T
- Dependence structure in Gaussian layer $\boldsymbol{Z}$
- Distribution for counts $\boldsymbol{Y}$ accounting for covariates/offsets

- Matrix Tree Theorem
- Gaussian Graphical Model
- Poisson log-normal model

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

$$T \sim \prod_{kl\in T} \beta_{kl}/B,$$

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

$$T \sim \prod_{kl \in T} \beta_{kl} / B,$$

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

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

$$\begin{cases} T \sim \prod_{kl\in T} \beta_{kl}/B, \\ \boldsymbol{Z}_i \mid T \sim \mathcal{N}(0, \boldsymbol{\Omega}_T) \\ Y_{ij} \mid \boldsymbol{Z}_i \sim \mathcal{P}(\exp(o_{ij} + \boldsymbol{x}_i^\intercal \boldsymbol{\theta}_j + Z_{ij})). \end{cases}$$

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:

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

Two-step procedure

EM algorithm (Dempster et al., 1977)

Maximizes the likelihood in presence of latent variables:

E step: Compute $\mathbb{E}[\log p_{\Theta^t}(\boldsymbol{Y}, \boldsymbol{Z}, T) \mid \boldsymbol{Y}]$

M step: $\Theta^{t+1} = \operatorname{argmax}_{\Theta} \left\{\mathbb{E}[\log p_{\Theta^t}(\boldsymbol{Y}, \boldsymbol{Z}, T) \mid \boldsymbol{Y}]\right\}$

Two-step procedure

EM algorithm (Dempster et al., 1977)

Maximizes the likelihood in presence of latent variables:

$$\begin{aligned} &\text{E step:} && \text{Compute } \mathbb{E}[\log p_{\Theta^t}(\boldsymbol{Y}, \boldsymbol{Z}, T) \mid \boldsymbol{Y}] \\ &\text{M step:} && \Theta^{t+1} = \mathrm{argmax}_{\Theta} \left\{ \mathbb{E}[\log p_{\Theta^t}(\boldsymbol{Y}, \boldsymbol{Z}, T) \mid \boldsymbol{Y}] \right\} \end{aligned}$$

1. `PLNmodels` (Chiquet et al., 2018) gives $\widehat{\boldsymbol{\theta}}$ and approximates of $\boldsymbol{Z} \mid \boldsymbol{Y}$ sufficient statistics.
2. EM algorithm to get $\widehat{\boldsymbol{\beta}}$.

Actually: $\tilde{\mathbb{E}}[\log p_{\beta}(\boldsymbol{Y}, \boldsymbol{Z}, T) \mid \boldsymbol{Z}] = \tilde{\mathbb{E}}[\log p_{\beta}(\boldsymbol{Z}, T) \mid \boldsymbol{Z}] + cst.$

Factorization on the edges

Tree structure factorization:

$$p_{\mathbf{\Omega}_T}(\boldsymbol{Z} \mid T) = \prod_k p(\boldsymbol{Z}_k) \prod_{kl \in T} \frac{p(\boldsymbol{Z}_k, \boldsymbol{Z}_l)}{p(\boldsymbol{Z}_k)\, p(\boldsymbol{Z}_l)}$$

Only the 1^{rst} and 2^{nd} order moments of $\boldsymbol{Z} \mid \boldsymbol{Y}$ are required, replaced by their variational approximation from step 1.

Expression of the surrogate

$$\tilde{\mathbb{E}}[\log p_\beta(\boldsymbol{Z}, T) \mid \boldsymbol{Z}] = \sum_{j<k} P_{jk} \log\left(\beta_{jk} \widehat{\psi}_{jk}\right) - \log B + cst,$$

where $\widehat{\psi}_{jk} = (1 - \widehat{\rho}_{jk}^2)^{-n/2}$ and $P_{jk} = \mathbb{P}\{jk \in T \mid \boldsymbol{Z}\}$.

Proposed EM algorithm

The M matrix is built from the inverse of a Laplacian matrix (Meilă and Jaakkola, 2006).

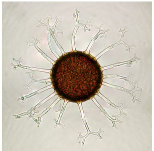
E step: $p(T \mid \boldsymbol{Z})$ factorizes on the edges.
Using the weight matrix $\mathbf{W} = \boldsymbol{\beta} \odot \widehat{\psi}$, all probabilities can be computed at once:

$$P_{jk} = w_{jk} M(\mathbf{W})_{jk} \text{ (Kirshner, 2008)}$$

M step: Requires the computation of $\partial_{\beta_{jk}}(\sum_{T \in \mathcal{T}} \prod_{jk \in T} \beta_{jk})$.
Closed form is available:

$$\beta_{jk}^{t+1} = \frac{P_{jk}^t}{M(\boldsymbol{\beta}^t)_{jk}}$$

Oak powdery mildew



Pathogen Erysiphe alphitoides (EA).

Oak leaf with powdery mildew.

Metabarcoding of oak tree leaves microbiome Jakuschkin et al. (2016).

- $\boldsymbol{Y}$: 116 sample of 114 microbial species counts (bacteria/fungi)
- $\mathbf{X}$: sampled tree, and 3 quantitative covariates
- $\mathbf{O}$: Different read depth for bacteria and fungi

Edge selection frequencies

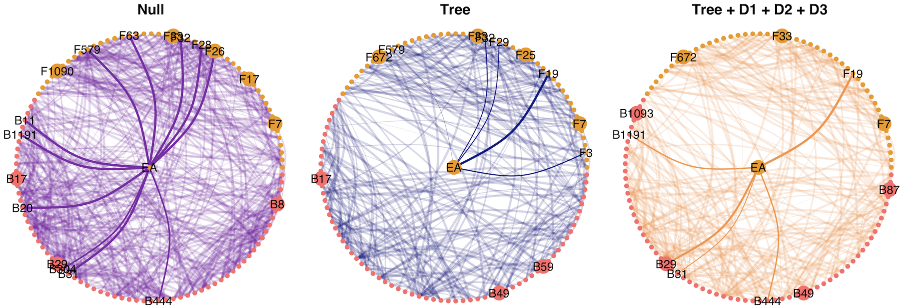
1. Create S random sub-samples using 80% of input abundance data
2.

s	edges probabilities				
1	2e-04	0.0024	0.0414	0.2507	
2	1e-04	0.0013	0.0004	0.0574	
3	2e-04	0.0013	0.0008	0.0127	…
⋮	⋮	⋮	⋮	⋮	

3. Apply average probability $2/p$ threshold on all resampled probabilities
4. $$F_{jk} = \sum_{s=1}^{S} \mathbb{1}\{P_{jk}^{s} \geq 2/p\}/S$$

Edges selection frequencies: 0.000 0.0381 0.0190 0.7048 …

Oak mildew networks

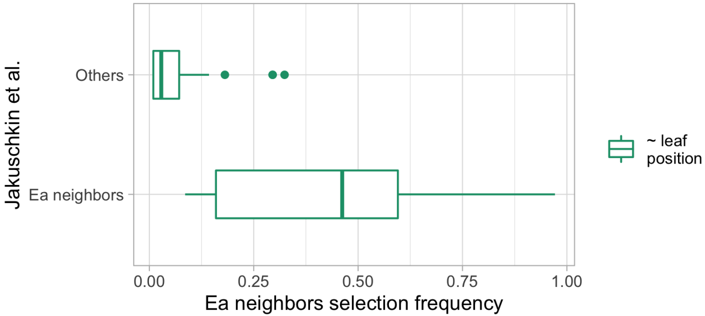


Frequencies above 90%.

6.5s: average running time for one model.

Ea neighbors: previous study

On the 39 infected samples:



Comparison with Jakuschkin et al. (2016)

20s: average running time.

Inference from incomplete counts

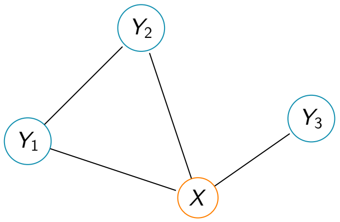
i Model

ii Inference

iii Illustration

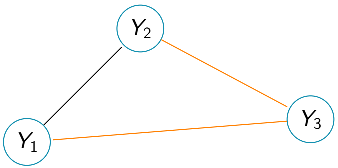
Marginalization of graphs

Complete graph:



$\Longrightarrow$

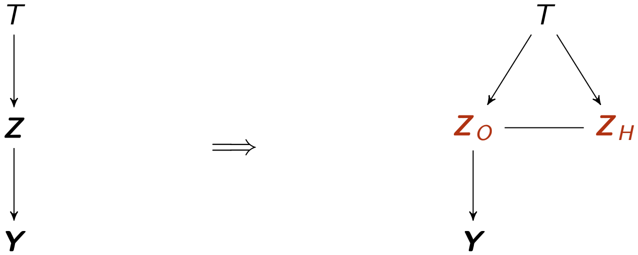
Marginal graph:



Spurious edges leading to wrong interpretation

X is a missing actor.

Added hidden Gaussian parameters



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

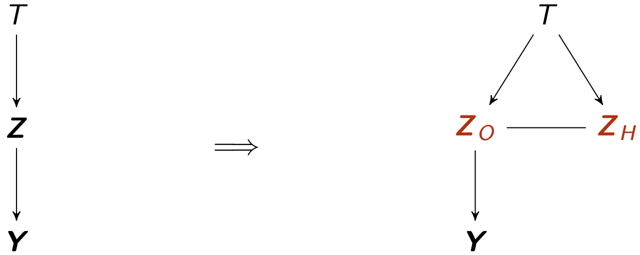
$(\boldsymbol{Z}_O, \boldsymbol{Z}_H) \mid T \sim \mathcal{N}(0, \boldsymbol{\Omega}_T^{-1})$

$\boldsymbol{Z}$: $n \times p$

$\boldsymbol{Z}_O$: $n \times p$

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

Added hidden Gaussian parameters



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

$\boldsymbol{Z}$: $n \times p$

$\boldsymbol{Z}_O$: $n \times p$

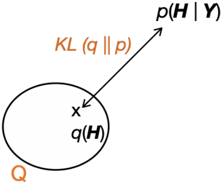
$\boldsymbol{Z}_H$: $n \times r$ $\qquad p' = p + r.$

- Same model with r additional dimensions
- Need access to sufficient statistics regarding Z_H

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:

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

Variational EM algorithm

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

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

Variational distribution

Two hidden variables: $\boldsymbol{Z} = (\boldsymbol{Z}_O, \mathbf{Z}_H)$ 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: $\boldsymbol{Z} = (\mathbf{Z}_O, \mathbf{Z}_H)$ 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 parameters: $\underset{n \times p',}{\widetilde{\boldsymbol{M}} = (\widetilde{\boldsymbol{M}}_O, \widetilde{\boldsymbol{M}}_H)}$, $\underset{n \times p',}{\widetilde{\boldsymbol{S}} = (\widetilde{\boldsymbol{S}}_O, \widetilde{\boldsymbol{S}}_H)}$, $\underset{p'^2}{\widetilde{\boldsymbol{\beta}}}$

Proposed algorithm

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

- 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: $\boldsymbol{\Omega}_T^{t+1}$, $\boldsymbol{\beta}^{t+1}$

- $\beta_{jk} = P_{jk}/M(\boldsymbol{\beta})_{jk}$ with $P_{jk} = \sum_{T \in \mathcal{T}, T \ni jk} g(T)$,
- $\boldsymbol{\Omega}_T$: adaptation of ML estimators (Lauritzen, 1996).

Proposed algorithm

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

- 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: $\boldsymbol{\Omega}_T^{t+1}$, $\boldsymbol{\beta}^{t+1}$

- $\beta_{jk} = P_{jk} / M(\boldsymbol{\beta})_{jk}$ with $P_{jk} = \sum_{T \in \mathcal{T}, T \ni jk} g(T)$,
- $\boldsymbol{\Omega}_T$: adaptation of ML estimators (Lauritzen, 1996). $p'^{p'-2} \times p'^2/2$ parameters $\Rightarrow$ $p'^2/2$ estimators.

Lauritzen's ML estimator

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

General Lauritzen's MLE

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

- The general *SSD* matrix do not depend on $\mathbf{G}$.
- 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

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

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:

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

Tree simplification of Lauritzen's formula:

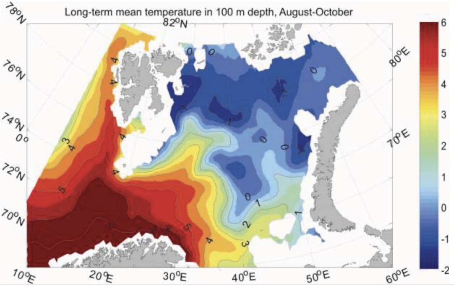
$$\omega_{Tjk}^{t+1} = \mathbb{1}\{jk \in T\} \left(\frac{-ssd_{jk}^t/n}{1-(ssd_{jk}^t/n)^2} \right),$$

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

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

Barent's sea fishes

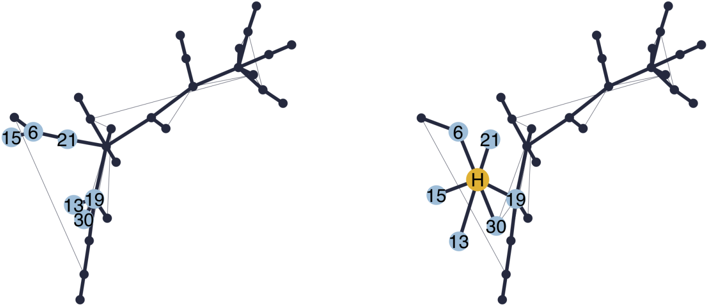
- ***Y***: abundances of 30 fish species in 89 sites,
- **X**: latitude, longitude, depth and temperature,
- **O**: total detections per site.



Stiansen et al. (2009)

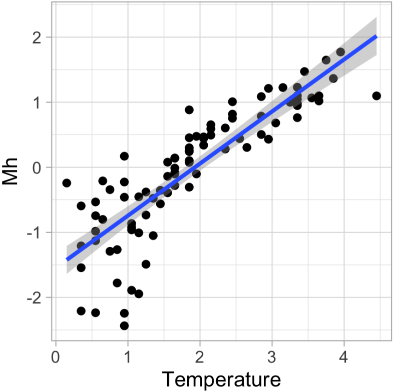
⇒ Fit with no covariates.

Barent's fishes networks

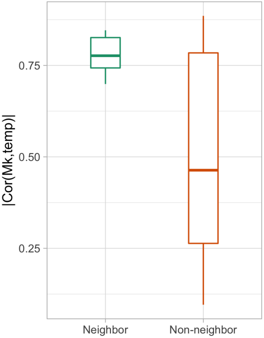


Left: observed network (3.3 mins). *Right*: network inferred with one missing actor: H (5.0 mins).

Relationship with temperature



$Cor(Mh, Temp) = 0.85$.



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

Conclusion and Perspectives

Conclusion

A probabilistic model for:

- Inferring conditional dependency network from abundance data.
- Accounting for covariates, offsets and missing actors.

An inference which:

- Takes advantage of the Gaussian framework flexibility
- Uses spanning trees algebraic properties to rely on determinants and inverses of graph Laplacian matrices.

Methods are implemented in R and available.

Extensions

Network analysis:

- Compare networks with the estimated tree distributions.
- Study interactions sign and strength available by computing partial correlations.

Ecological specifics:

- Different emission law (presence/absence), provided there is a Gaussian latent layer of parameters.
- Account for spatial dependencies within the Gaussian covariance structure.

Direct model:

- Graphical model on counts with tree averaging

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`.

References I

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

Beal, M. and Ghahramani, Z. (2003). The variational bayesian em algorithm for incomplete data: with application to scoring graphical model structures. In *Bayesian Statistics 7: Proceedings of the Seventh Valencia International Meeting*, volume 7, pages 453–464. Oxford University Press.

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.

Chow, C. and Liu, C. (1968). Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, 14(3):462–467.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Statist. Soc., series B*, 39:1–38.

Jakuschkin, B., Fievet, V., Schwaller, L., Fort, T., Robin, C., and Vacher, C. (2016). Deciphering the pathobiome: Intra- and interkingdom interactions involving the pathogen erysiphe alphitoides. *Microb Ecol*, 72(4):870–880.

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.

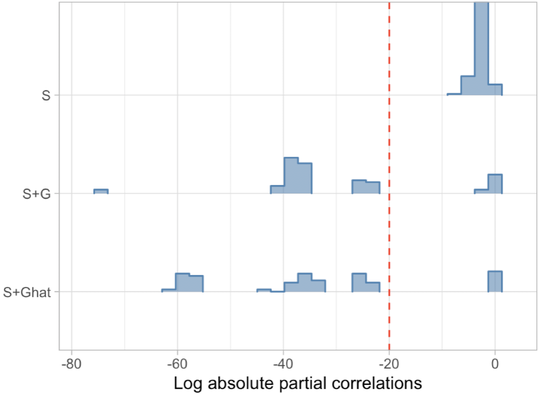
Meilă, M. and Jaakkola, T. (2006). Tractable bayesian learning of tree belief networks. *Statistics and Computing*, 16(1):77–92.

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.

Signs and strengths of interactions

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

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



$\widehat{S} = S$:

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

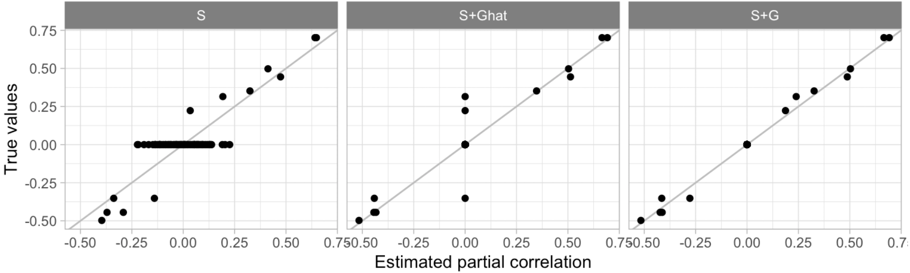
$\widehat{S} = f(S, \mathbf{G})$:

	-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

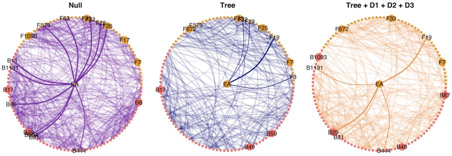
Signs and strengths of interactions

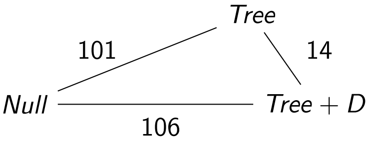


Network comparison

$$D(p_{\boldsymbol{\beta}^A}, p_{\boldsymbol{\beta}^B}) = \frac{1}{2}\left[KL\left(p_{\boldsymbol{\beta}^B} \,||\, p_{\boldsymbol{\beta}^A}\right) + KL\left(p_{\boldsymbol{\beta}^A} \,||\, p_{\boldsymbol{\beta}^B}\right)\right]$$
$$= \sum_{kl} \log(\beta^A_{kl}/\beta^B_{kl})\left(\frac{P^A_{kl} - P^B_{kl}}{2}\right)$$

Oak dataset:





A different emission law

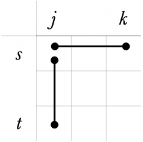
$$\begin{cases} T \sim \prod_{kl \in T} \beta_{kl} / B, \\ \boldsymbol{Z}_i \mid T \sim \mathcal{N}(0, \boldsymbol{\Omega}_T), \\ Y_{ij} \mid \boldsymbol{Z}_i \sim \mathcal{F}_j(o_{ij}, \boldsymbol{x}_i, Z_{ij}). \end{cases}$$

$\mathcal{F}_j : \mathcal{B}, \mathcal{P}, ...$

Account for spatial dependencies

Separate dependencies: $\Gamma = (\Gamma_{st})_{1\leq s,t\leq n}$, $\Sigma_T = (\sigma_{jk})_{1\leq j,k\leq p}$.

$$\begin{cases} \mathbb{C}\mathrm{ov}\left(Z_{sj}, Z_{sk}\right) & = \gamma_{ss}\sigma_{jk} \\ \mathbb{C}\mathrm{ov}\left(Z_{sj}, Z_{tj}\right) & = \sigma_{jj}\gamma_{st} \end{cases}$$



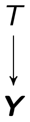
Defining $Vec(\boldsymbol{Z}) = (Z_{11}, ..., Z_{1p}, Z_{21}, ..., Z_{np}) \in \mathbb{R}^{n\times p}$, we obtain:

$$Vec(\boldsymbol{Z}) \sim \mathcal{N}(0, \Gamma \otimes \Sigma_T).$$

Γ as a function of $|s-t|$ reduces the number of parameters.

Network inference from counts

With any marginal and bivariate distribution for counts:

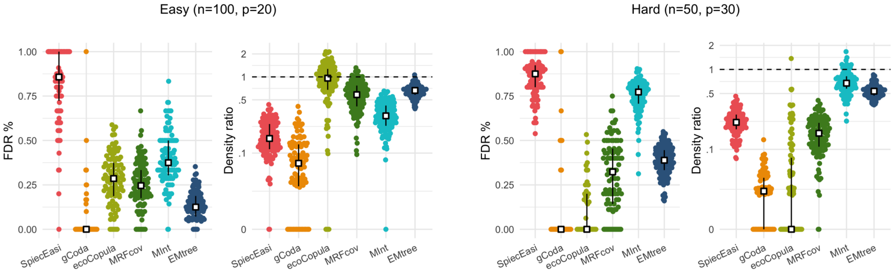


$$p_\theta(\boldsymbol{Y}_i \mid T) = \prod_{j=1}^{p} p_\theta(Y_{ij}) \prod_{jk \in T} \frac{p_\theta(Y_{ij}, Y_{ik})}{p_\theta(Y_{ij}) p_\theta(Y_{ik})}.$$

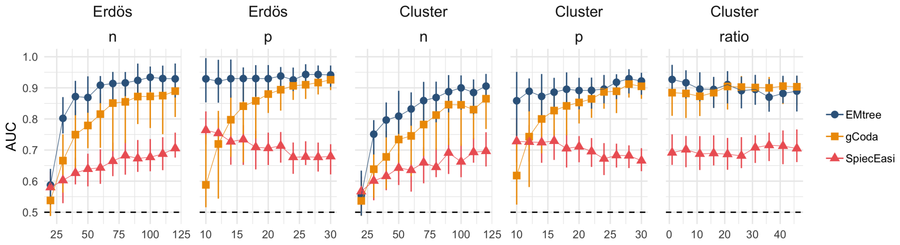
The joint distribution of counts would be a mixture on spanning trees:

$$p_{\beta,\theta}(\boldsymbol{Y}) = \sum_{T \in \mathcal{T}} p_\beta(T) p_\theta(\boldsymbol{Y} \mid T).$$

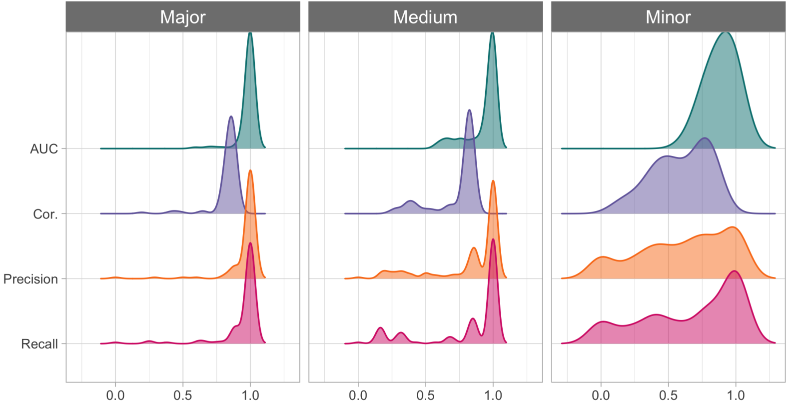
Network inference methods comparison



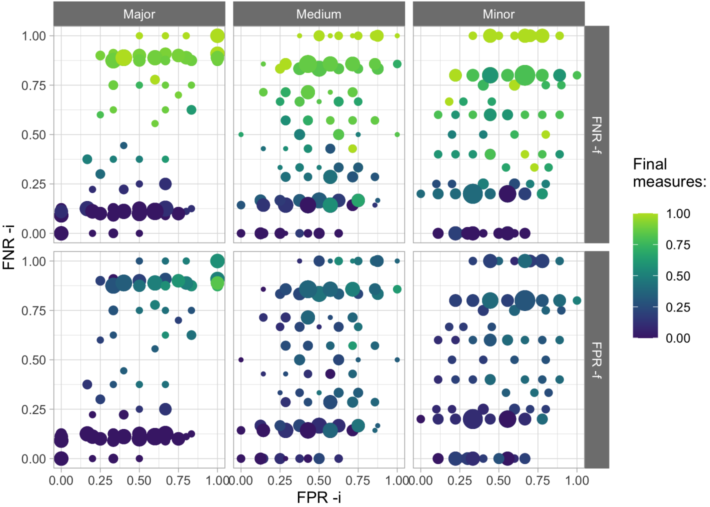
Edges scoring comparison



Reconstruction of the missing actor



Initialize with more potential neighbors



Lauritzen's notation

For any square matrix **A**:

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

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

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 **Q** must be positive definite, which calls for some numerical control of the weights β:

- centering in log scale
- sum constraint

Variational weights depend on the number of available samples *n*. Tempering parameter α:

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