An exchangeable model for (ecological) bipartite networks

S. Robin

Sorbonne université

joint work with S. Ouadah (AgroParisTech), P. Latouche (université Paris Cité)

and T. Le Minh, S. Donnet (INRAE), F. Massol (CNRS, Lille)

WG on Risk, ESSEC, Mar. 2022

Bipartite networks

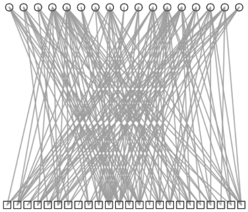
Bipartite networks describe the connections between two set of entities

- authors / publications,
- actors / movies,
- hosts / parasites,
- plants / pollinators,
- etc.

Bipartite networks

Bipartite networks describe the connections between two set of entities

- authors / publications,
- actors / movies,
- hosts / parasites,
- plants / pollinators,
- etc.



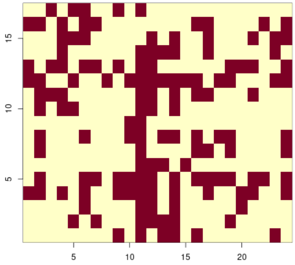
Example:

Top: Plants (○)

Bottom: Pollinators (□)

Zackenberg network from [OBEJ08,SROB16,CRR+18]:

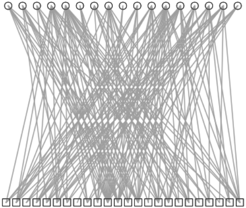
$$m = 17, \quad n = 24$$

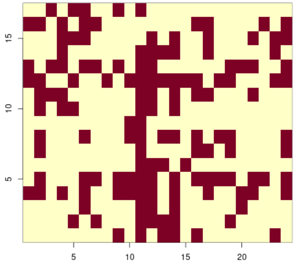


Species exchangeability

Aim. Describe the global organization ('topology') of the network to

- better understand the behavior of the ecosystem,
- or predict its response to a change

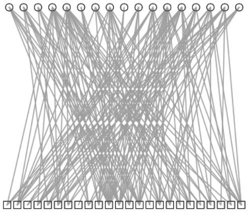


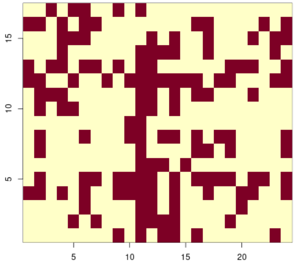


Species exchangeability

Aim. Describe the global organization ('topology') of the network to

- better understand the behavior of the ecosystem,
- or predict its response to a change



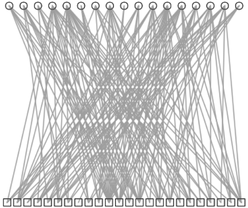


Descriptors ('metrics'): network density, number of connected components, number of 'clusters', 'nestedness', ...

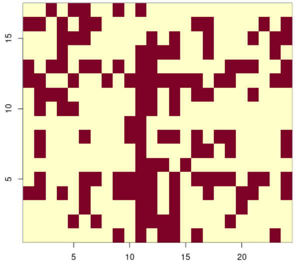
Species exchangeability

Aim. Describe the global organization ('topology') of the network to

- better understand the behavior of the ecosystem,
- or predict its response to a change



Descriptors ('metrics'): network density, number of connected components, number of 'clusters', 'nestedness', ...



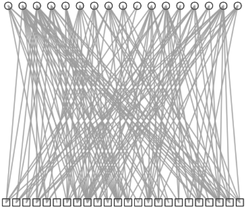
Species exchangeability.

- Most descriptors remain unchanged when relabeling the species.
- Species are not each considered *per se*.

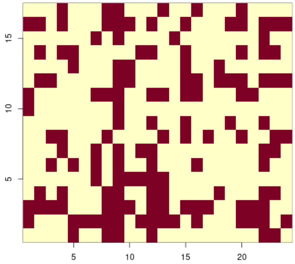
Species exchangeability

Aim. Describe the global organization ('topology') of the network to

- better understand the behavior of the ecosystem,
- or predict its response to a change



Descriptors ('metrics'): network density, number of connected components, number of 'clusters', 'nestedness', ...



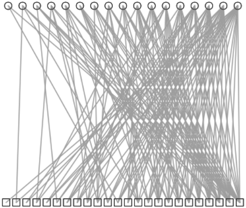
Species exchangeability.

- Most descriptors remain unchanged when relabeling the species.
- Species are not each considered *per se*.

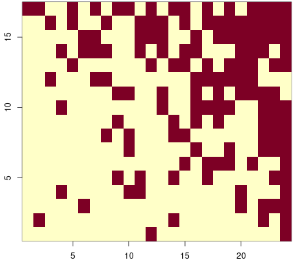
Species exchangeability

Aim. Describe the global organization ('topology') of the network to

- better understand the behavior of the ecosystem,
- or predict its response to a change



Descriptors ('metrics'): network density, number of connected components, number of 'clusters', 'nestedness', ...



Species exchangeability.

- Most descriptors remain unchanged when relabeling the species.
- Species are not each considered *per se*.

Outline

Probabilistic formulation

Network. Can be seen either as

- a graph $\mathcal{G} = (\mathcal{V} = \text{set of vertices}, \mathcal{E} = \text{set of edges} \subset \mathcal{V} \times \mathcal{V})$, or
- an $m \times n$ adjacency matrix $Y = [Y_{ij}]$ where

$$Y_{ij} = 1 \text{ iff the node } i \text{ is connected with the node } j$$

Probabilistic formulation

Network. Can be seen either as

- a graph $\mathcal{G} = (\mathcal{V} = \text{set of vertices}, \mathcal{E} = \text{set of edges} \subset \mathcal{V} \times \mathcal{V})$, or
- an $m \times n$ adjacency matrix $Y = [Y_{ij}]$ where

$$Y_{ij} = 1 \text{ iff the node } i \text{ is connected with the node } j$$

Probabilistic framework. The observed network is seen as a realization of a random graph ruled by some joint distribution $p(y)$, that is

$$p(y) = \mathbb{P}\{Y = y\} = \mathbb{P}\{Y_{11} = 1, Y_{12} = 0, \dots Y_{ij} = 1, \dots, Y_{mn} = 0\}$$

Probabilistic formulation

Network. Can be seen either as

- a graph $\mathcal{G} = (\mathcal{V} = \text{set of vertices}, \mathcal{E} = \text{set of edges} \subset \mathcal{V} \times \mathcal{V})$, or
- an $m \times n$ adjacency matrix $Y = [Y_{ij}]$ where

$$Y_{ij} = 1 \text{ iff the node } i \text{ is connected with the node } j$$

Probabilistic framework. The observed network is seen as a realization of a random graph ruled by some joint distribution $p(y)$, that is

$$p(y) = \mathbb{P}\{Y = y\} = \mathbb{P}\{Y_{11} = 1, Y_{12} = 0, \dots Y_{ij} = 1, \dots, Y_{mn} = 0\}$$

Network analysis. Classical framework in statistical modelling:

- Questions of (ecological) interest have to be translated into
- Questions about the distribution p.

Row-column exchangeability (RCE)

Exchangeability assumption. Species exchangeability means

- Plants can be exchanged with plants, insects can be exchanged with insects,
- That is: for any permutation σ_R of the rows and any permutation σ_C of the columns:

$$\mathbb{P}(Y = y) = \mathbb{P}(Y = y_{\sigma_R,\sigma_C}).$$

Row-column exchangeability (RCE)

Exchangeability assumption. Species exchangeability means

- Plants can be exchanged with plants, insects can be exchanged with insects,
- That is: for any permutation σ_R of the rows and any permutation σ_C of the columns:

$$\mathbb{P}(Y = y) = \mathbb{P}(Y = y_{\sigma_R,\sigma_C}).$$

Aldous-Hoover representation. A random graph $Y \sim p$ is RCE iff there exist a determistic function f such that

$$Y = f(T, U_1, \dots U_m, V_1, \dots, V_n, W_{11}, \dots W_{mn})$$

$$\text{with} \left\{ \begin{array}{ll} T, & \text{(whole graph)} \\ U_1, \dots U_m, & \text{(rows)} \\ V_1, \dots, V_n, & \text{(columns)} \\ W_{11}, \dots W_{mn} & \text{(edges)} \end{array} \right\} \text{iid} \ \sim \mathcal{U}[0,1].$$

Bipartite w-graph

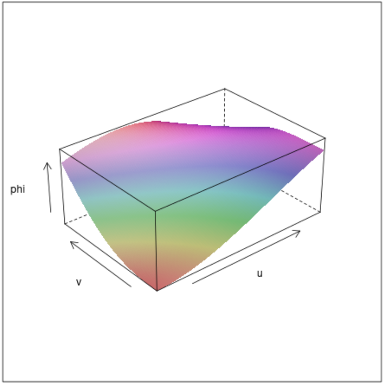
Model.

- Consider a 'graphon' function

$$\phi : [0,1] \times [0,1] \mapsto [0,1],$$

- draw $(U_1, \dots, U_m)$ iid $\sim \mathcal{U}[0,1]$,
- draw $(V_1, \dots, V_n)$ iid $\sim \mathcal{U}[0,1]$,
- draw $(Y_{11}, \dots, Y_{mn})$ independently:

$$\mathbb{P}\{Y_{ij} = 1\} = \phi(U_i, V_j).$$



Bipartite w-graph

Model.

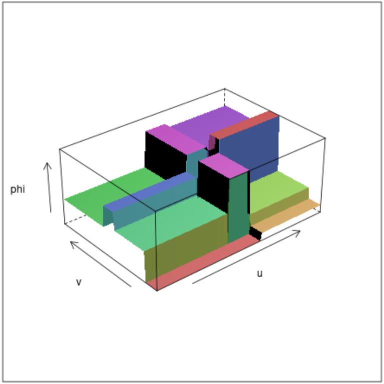
- Consider a 'graphon' function

$$\phi : [0,1] \times [0,1] \mapsto [0,1],$$

- draw $(U_1, \dots, U_m)$ iid $\sim \mathcal{U}[0,1]$,
- draw $(V_1, \dots, V_n)$ iid $\sim \mathcal{U}[0,1]$,
- draw $(Y_{11}, \dots, Y_{mn})$ independently:

$$\mathbb{P}\{Y_{ij} = 1\} = \phi(U_i, V_j).$$

Latent block-model [GN05]. Block-wise constant function ϕ.



Bipartite w-graph

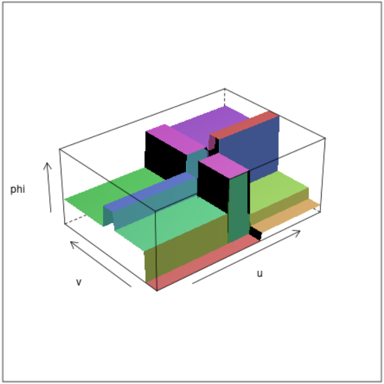
Model.

- Consider a 'graphon' function

$$\phi : [0,1] \times [0,1] \mapsto [0,1],$$

- draw $(U_1, \dots, U_m)$ iid $\sim \mathcal{U}[0,1]$,
- draw $(V_1, \dots, V_n)$ iid $\sim \mathcal{U}[0,1]$,
- draw $(Y_{11}, \dots, Y_{mn})$ independently:

$$\mathbb{P}\{Y_{ij} = 1\} = \phi(U_i, V_j).$$



Latent block-model [GN05]. Block-wise constant function ϕ.

Dissociated models. No 'whole graph' random term T:

- Non-overlapping blocks of the adjacency matrix Y are independent.

Expected degree distribution model

A 'null' model. Product form w-graph:

$$\phi(u, v) = \rho\, f(u)\, g(v)$$

- $f : [0, 1] \mapsto \mathbb{R}^+$: row imbalance (generalist vs specialist plants),
- $g : [0, 1] \mapsto \mathbb{R}^+$: column imbalance (generalist vs specialist insects),
- $\rho \in [0, 1]$: network density[1].

[1] $\int f(u)\, \mathrm{d}u = 1, \quad \int g(v)\, \mathrm{d}v = 1, \quad \rho \leqslant 1/(\max(f)\max(g))$

Expected degree distribution model

A 'null' model. Product form w-graph:

$$\phi(u, v) = \rho\, f(u)\, g(v)$$

- $f : [0,1] \mapsto \mathbb{R}^+$: row imbalance (generalist vs specialist plants),
- $g : [0,1] \mapsto \mathbb{R}^+$: column imbalance (generalist vs specialist insects),
- $\rho \in [0,1]$: network density[1].

Expected degrees.

$$\begin{aligned} D_i^\circ &= \text{ degree of top node } i: & \mathbb{E}(D_i^\circ \mid U_i = u) &= n\,\rho\, f(u), \\ D_j^\square &= \text{ degree of bottom node } j: & \mathbb{E}(D_j^\square \mid V_j = v) &= m\,\rho\, g(v). \end{aligned}$$

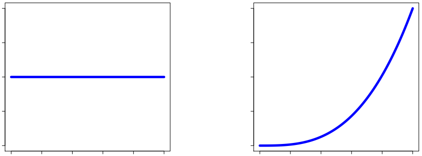
→ Bipartite version of the expected degree distribution (EDD) model [CL02].

[1] $\int f(u)\, \mathrm{d}u = 1, \quad \int g(v)\, \mathrm{d}v = 1, \quad \rho \leqslant 1/(\max(f)\max(g))$

BEDD model

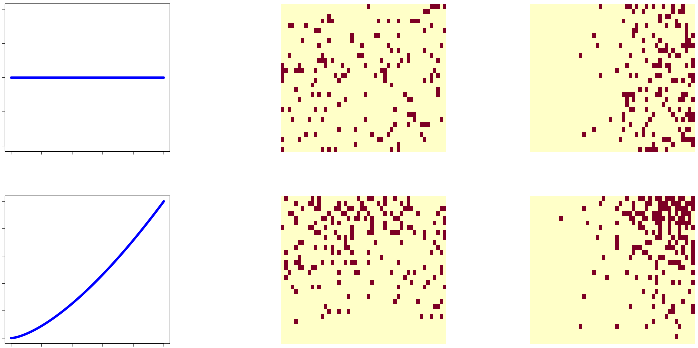
- Only individual effects matter: $f(U_i)$, $g(V_j)$.
- No specific interaction (U_i, V_j)

$g =$ $g =$



$f =$

$f =$



Outline

Bipartite network motifs

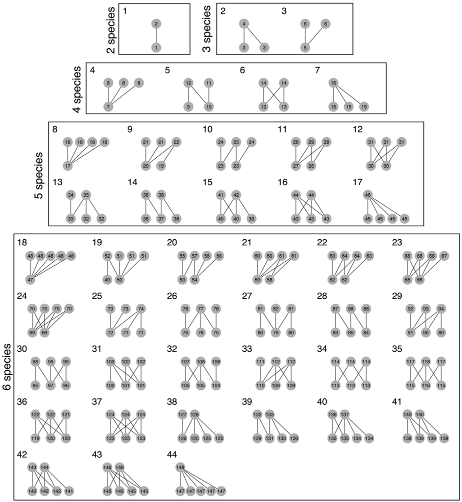
'Meso-scale' analysis. [SCB+19]

- Motifs ='building-blocks'
- between local (several nodes) and global (sub-graph)

Interest.

- Generic description of a network
- Enables network comparison
- Even when the nodes are different

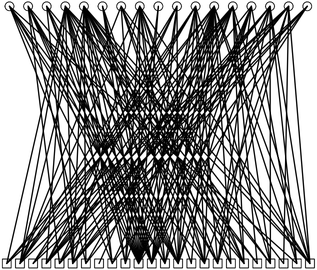
(+ 'species-role': see later?)



Existing tool. `bmotif` package [SSS+19]: counts motif occurrences

Example

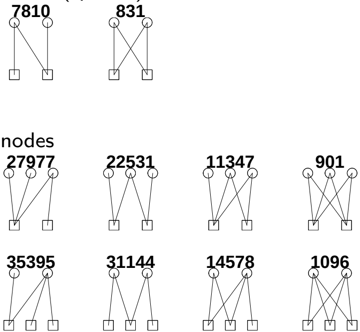
Zackenberg network.



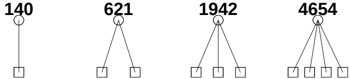
Motif counts.

4 nodes (species)

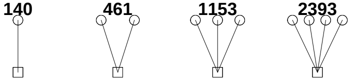
5 nodes



top 'stars' (plants)



bottom 'stars' (pollinators)



Counting motifs

Number of positions.

- Choose p nodes among m
- Choose q nodes among n
- Try all *automorphisms*

$$c_s := \binom{m}{p} \times \binom{n}{q} \times r_s$$

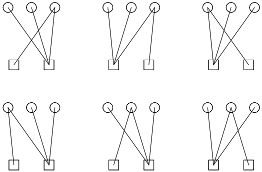
Counting motifs

Number of positions.

- Choose p nodes among m
- Choose q nodes among n
- Try all *automorphisms*

$$c_s := \binom{m}{p} \times \binom{n}{q} \times r_s$$

Automorphisms = non-redundant permutations



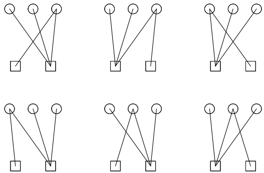
Counting motifs

Number of positions.

- Choose p nodes among m
- Choose q nodes among n
- Try all *automorphisms*

$$c_s := \binom{m}{p} \times \binom{n}{q} \times r_s$$

Automorphisms = non-redundant permutations



Motif count. Try all positions $\alpha = 1, \dots c_s$, define

$$Y_{s\alpha} = 1 \text{ if match}, \qquad 0 \text{ otherwise},$$

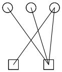
then count the number of matches:

$$N_s = \sum_\alpha Y_{s\alpha}$$

$\rightarrow$ Motif frequency: $F_s := N_s / c_s$

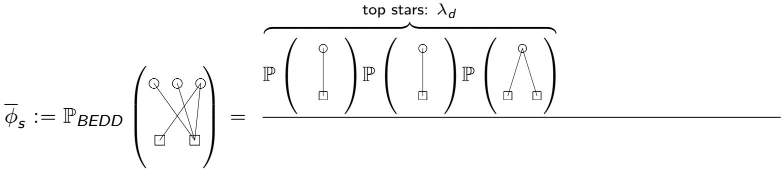
Motif probability

Occurrence probability $\overline{\phi}_s = \mathbb{P}\{Y_{s\alpha} = 1\}$. Under the BEDD model:

$\overline{\phi}_s := \mathbb{P}_{BEDD} \Big($  $\Big) = \frac{\quad}{\quad}$

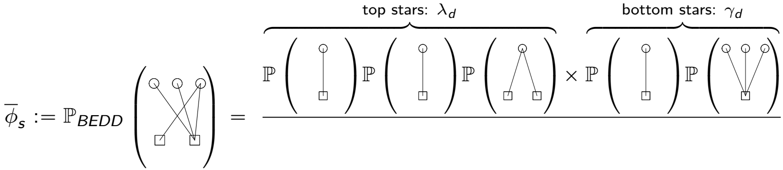
Motif probability

Occurrence probability $\overline{\phi}_s = \mathbb{P}\{Y_{s\alpha} = 1\}$. Under the BEDD model:



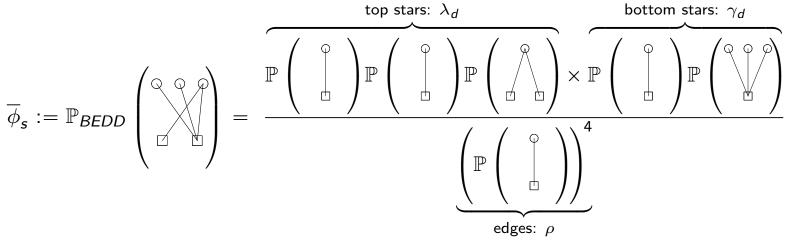
Motif probability

Occurrence probability $\overline{\phi}_s = \mathbb{P}\{Y_{s\alpha} = 1\}$. Under the BEDD model:



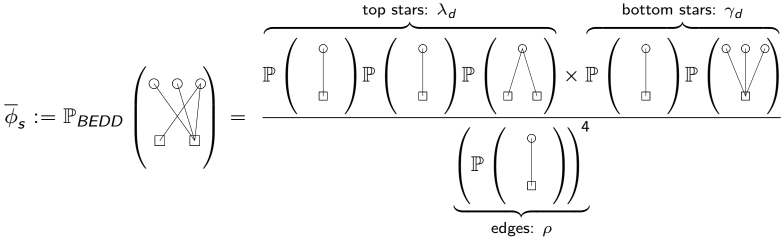
Motif probability

Occurrence probability $\overline{\phi}_s = \mathbb{P}\{Y_{s\alpha} = 1\}$. Under the BEDD model:



Motif probability

Occurrence probability $\overline{\phi}_s = \mathbb{P}\{Y_{s\alpha} = 1\}$. Under the BEDD model:

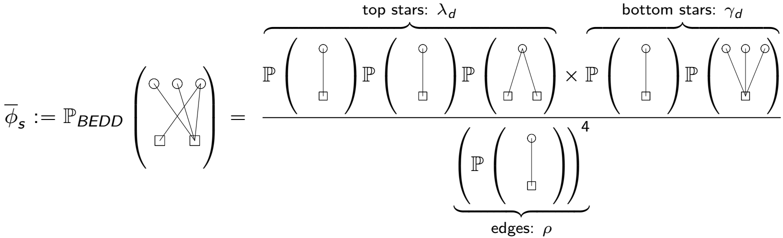


$$= \frac{\rho^2\lambda_2\rho\gamma_3}{\rho^4} \qquad = \frac{\lambda_2\gamma_3}{\rho}$$

where $\lambda_d = \rho^d \int f^d(u)\, \mathrm{d}u, \qquad \gamma_d = \rho^d \int g^d(v)\, \mathrm{d}v$

Motif probability

Occurrence probability $\overline{\phi}_s = \mathbb{P}\{Y_{s\alpha} = 1\}$. Under the BEDD model:



$$= \frac{\rho^2 \lambda_2 \rho \gamma_3}{\rho^4} \qquad = \frac{\lambda_2 \gamma_3}{\rho} \qquad \text{where } \lambda_d = \rho^d \int f^d(u) \, \mathrm{d}u, \qquad \gamma_d = \rho^d \int g^d(v) \, \mathrm{d}v$$

Estimated probability.

$$\overline{\phi}_s := \frac{\lambda_2 \gamma_3}{\rho} \qquad \rightarrow \qquad \overline{F}_s := \frac{\Lambda_2 \Gamma_3}{R}$$

where Λ_2, Γ_3, R = empirical frequencies of top stars, bottom stars and edges.

Distribution of the count

Moments:

- Mean: $\mathbb{E}_{BEDD}(N_s) = c_s \times \overline{\phi}_s$

- Variance: Requires to evaluate $\mathbb{E}_{BEDD}(N_s^2) = \mathbb{E}_{BEDD}\left(\sum_\alpha Y_{s\alpha}\right)^2$

  $\rightarrow$ Need to account for overlap between positions (*super-motifs*: [PDK$^+$08])

- Covariance: Same game to compute $\mathbb{C}\text{ov}(N_s, N_{s'})$

Distribution of the count

Moments:

- Mean: $\mathbb{E}_{BEDD}(N_s) = c_s \times \overline{\phi}_s$
- Variance: Requires to evaluate $\mathbb{E}_{BEDD}(N_s^2) = \mathbb{E}_{BEDD}\left(\sum_\alpha Y_{s\alpha}\right)^2$

  $\rightarrow$ Need to account for overlap between positions (*super-motifs*: [PDK+08])
- Covariance: Same game to compute $\mathbb{C}\text{ov}(N_s, N_{s'})$

Asymptotic normality for non-star motifs [OLR22]. Under BEDD (and sparsity conditions):

$$(F_s - \overline{F}_s) \Big/ \sqrt{\widehat{\mathbb{V}}(F_s)} \quad \xrightarrow{m,n\to\infty} \quad \mathcal{N}(0,1)$$

Distribution of the count

Moments:

- Mean: $\mathbb{E}_{BEDD}(N_s) = c_s \times \overline{\phi}_s$
- Variance: Requires to evaluate $\mathbb{E}_{BEDD}(N_s^2) = \mathbb{E}_{BEDD}\left(\sum_\alpha Y_{s\alpha}\right)^2$

  $\rightarrow$ Need to account for overlap between positions (*super-motifs*: [PDK+08])

- Covariance: Same game to compute $\mathbb{C}\text{ov}(N_s, N_{s'})$

Asymptotic normality for non-star motifs [OLR22]. Under BEDD (and sparsity conditions):

$$(F_s - \overline{F}_s) \Big/ \sqrt{\widehat{\mathbb{V}}(F_s)} \quad \xrightarrow{m,n\to\infty} \quad \mathcal{N}(0,1)$$

Proof:

- decompose

$$F_s - \overline{F}_s = \underbrace{(F_s - \phi_s)}_{\text{random fluctuations}} + \underbrace{(\phi_s - \overline{\phi}_s)}_{\text{null under BEDD}} + \underbrace{(\overline{\phi}_s - \overline{F}_s)}_{\text{estimation error }\to 0},$$

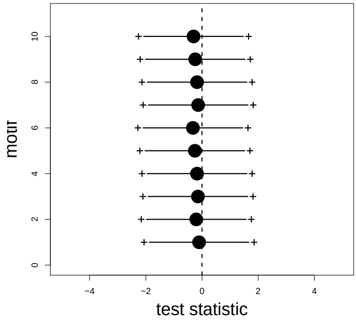
- construct a counting martingale [GL17] for $(F_s - \phi_s)$ + consistent estimate of $\widehat{\mathbb{V}}(F_s)$.

Goodness-of-fit (GOF) of the BEDD model

Raw test statistic:

$$T_s = \frac{N_s - \widehat{\mathbb{E}} N_s}{\sqrt{\widehat{\mathbb{V}} N_s}}$$

Zackenberg network.



[2] $\Sigma = P \Lambda P^\intercal$, $\Sigma = P \Lambda^{-1/2} P^\intercal$

Goodness-of-fit (GOF) of the BEDD model

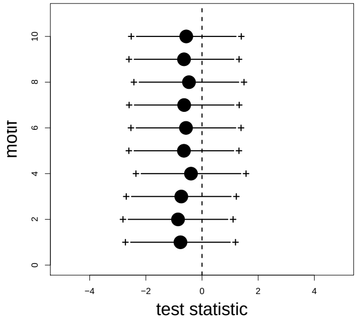
Raw test statistic:

$$T_s = \frac{N_s - \widehat{\mathbb{E}} N_s}{\sqrt{\widehat{\mathbb{V}} N_s}}$$

Corrected statistic: accounts for the estimation error in $\widehat{\mathbb{E}} N$

$$T'_s = \frac{N_s - (\widehat{\mathbb{E}} N_s - \widehat{\mathbb{B}}(\widehat{\mathbb{E}} N_s))}{\sqrt{\widehat{\mathbb{V}}(N_s - \widehat{\mathbb{E}} N_s)}}$$

Zackenberg network.



[2] $\Sigma = P \Lambda P^\intercal$, $\Sigma = P \Lambda^{-1/2} P^\intercal$

Goodness-of-fit (GOF) of the BEDD model

Raw test statistic:

$$T_s = \frac{N_s - \widehat{\mathbb{E}} N_s}{\sqrt{\widehat{\mathbb{V}} N_s}}$$

Corrected statistic: accounts for the estimation error in $\widehat{\mathbb{E}} N$

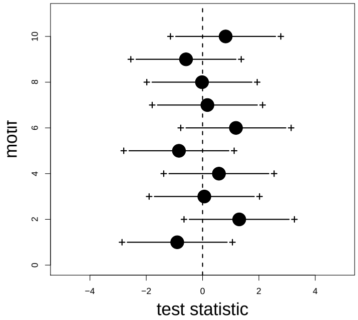
$$T'_s = \frac{N_s - (\widehat{\mathbb{E}} N_s - \widehat{\mathbb{B}}(\widehat{\mathbb{E}} N_s))}{\sqrt{\widehat{\mathbb{V}}(N_s - \widehat{\mathbb{E}} N_s)}}$$

Cholevski [2] transformation: accounts for the correlation between the counts

$$\Sigma_{s,s'} = \mathbb{C}\text{ov}(N_s - \widehat{\mathbb{E}} N_s, N_{s'} - \widehat{\mathbb{E}} N_{s'})$$

$$T'' = \widehat{\Sigma}^{-1/2} \left[N_s - (\widehat{\mathbb{E}} N_s - \widehat{\mathbb{B}}(\widehat{\mathbb{E}} N_s)) \right]$$

Zackenberg network.



[2] $\Sigma = P \Lambda P^\intercal$, $\Sigma = P \Lambda^{-1/2} P^\intercal$

Network comparison

Same degree imbalance for top nodes.

- Consider two networks $G^A \sim BEDD(\rho^A, f^A, g^A)$ and $G^B \sim BEDD(\rho^B, f^B, g^B)$ and

$$H_0 = \{f^A = f^B\}$$

Network comparison

Same degree imbalance for top nodes.

- Consider two networks $G^A \sim BEDD(\rho^A, f^A, g^A)$ and $G^B \sim BEDD(\rho^B, f^B, g^B)$ and

$$H_0 = \{f^A = f^B\}$$

- Under H_0 [OLR22]

$$W_s = \frac{\left(F_s^A - \widehat{\mathbb{E}}_0(F_s^A)\right) - \left(F_s^B - \widehat{\mathbb{E}}_0(F_s^B)\right)}{\sqrt{\widehat{\mathbb{V}}_0(F_s^A) + \widehat{\mathbb{V}}_0(F_s^B)}} \xrightarrow{D} \mathcal{N}(0,1)$$

Network comparison

Same degree imbalance for top nodes.

- Consider two networks $G^A \sim BEDD(\rho^A, f^A, g^A)$ and $G^B \sim BEDD(\rho^B, f^B, g^B)$ and

$$H_0 = \{f^A = f^B\}$$

- Under H_0 [OLR22]

$$W_s = \frac{\left(F_s^A - \widehat{\mathbb{E}}_0(F_s^A)\right) - \left(F_s^B - \widehat{\mathbb{E}}_0(F_s^B)\right)}{\sqrt{\widehat{\mathbb{V}}_0(F_s^A) + \widehat{\mathbb{V}}_0(F_s^B)}} \xrightarrow{D} \mathcal{N}(0,1)$$

Plant-pollinator vs Plant-seed disperser.

- $G^A = 546 \times 1044$ plant-pollinator network, $G^B = 207 \times 110$ plant-seed disperser network
- Is there the same degree imbalance between plants in the two networks ?

Network comparison

Same degree imbalance for top nodes.

- Consider two networks $G^A \sim BEDD(\rho^A, f^A, g^A)$ and $G^B \sim BEDD(\rho^B, f^B, g^B)$ and

$$H_0 = \{f^A = f^B\}$$

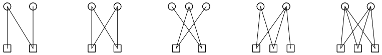
- Under H_0 [OLR22]

$$W_s = \frac{\left(F_s^A - \widehat{\mathbb{E}}_0(F_s^A)\right) - \left(F_s^B - \widehat{\mathbb{E}}_0(F_s^B)\right)}{\sqrt{\widehat{\mathbb{V}}_0(F_s^A) + \widehat{\mathbb{V}}_0(F_s^B)}} \xrightarrow{D} \mathcal{N}(0,1)$$

Plant-pollinator vs Plant-seed disperser.

- $G^A = 546 \times 1044$ plant-pollinator network, $G^B = 207 \times 110$ plant-seed disperser network
- Is there the same degree imbalance between plants in the two networks ?

Results.



W'_s : -2.71 -1.90 -1.76 -1.34 -0.96

Outline

U-statistics

U-statistics.

- Consider a symmetric function (*kernel*) $h : \mathbb{R}^k \mapsto \mathbb{R}$:

$$h(y_1, \dots y_k) = h(y_{\sigma(1)}, \dots y_{\sigma(k)}) \qquad \text{for any permutation } \sigma$$

- Consider $(Y_1, \dots Y_n)$,
- Define

$$U_n = \binom{n}{k}^{-1} \sum_{1 \leqslant i_1 < \dots < i_k \leqslant n} h(Y_{i_1}, \dots Y_{i_k}),$$

- Asymptotic normality conditions for U_n when the $(Y_1, \dots Y_n)$ are iid [Hoe48] or exchangeable [NS63].

U-statistics for bipartite networks

U-statistics of order 2×2.

- Consider row-column symmetric kernel $h : \mathbb{R}^4 \mapsto \mathbb{R}$:

$$h(y_{11}, y_{12}, y_{21}, y_{22}) = h(y_{21}, y_{22}, y_{11}, y_{12}) = h(y_{12}, y_{11}, y_{22}, y_{21}),$$

- Consider $Y = [Y_{ij}]_{1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n}$,
- Define

$$U_{m,n} = \binom{m}{2}^{-1} \binom{n}{2}^{-1} \sum_{1 \leqslant i_1 < i_2 \leqslant m} \sum_{1 \leqslant j_1 < j_2 \leqslant n} h(Y_{i_1 j_1}, Y_{i_1 j_2}, Y_{i_2 j_1}, Y_{i_2 j_2}).$$

- Tam Le Minh's PhD [LM21]: asymptotic normality of $U_{m,n}$ when Y is RCE dissociated[3], but not only with a product-form (i.e. includes *w*-graph).

[3] plus technical conditions

Weighted BEDD model

Weighted networks.

- Interactions Y_{ij} may be valued ('weighted').
- Example: Y_{ij} = number of visits from insect j to plant i within a given period of time.

[4] $\int f(u) \, \mathrm{d}u = 1, \quad \int g(v) \, \mathrm{d}v = 1$

Weighted BEDD model

Weighted networks.

- Interactions Y_{ij} may be valued ('weighted').
- Example: Y_{ij} = number of visits from insect j to plant i within a given period of time.

Weighted BEDD model: Poisson BEDD.

- $f : [0,1] \mapsto \mathbb{R}^+$: row imbalance (generalists vs specialists),
- $g : [0,1] \mapsto \mathbb{R}^+$: column imbalance (generalists vs specialists),
- $\lambda \in \mathbb{R}^+$: mean interaction intensity[4],

[4] $\int f(u)\, \mathrm{d}u = 1, \quad \int g(v)\, \mathrm{d}v = 1$

Weighted BEDD model

Weighted networks.

- Interactions Y_{ij} may be valued ('weighted').
- Example: Y_{ij} = number of visits from insect j to plant i within a given period of time.

Weighted BEDD model: Poisson BEDD.

- $f : [0,1] \mapsto \mathbb{R}^+$: row imbalance (generalists vs specialists),
- $g : [0,1] \mapsto \mathbb{R}^+$: column imbalance (generalists vs specialists),
- $\lambda \in \mathbb{R}^+$: mean interaction intensity[4],
- Draw

$$(U_1, \dots, U_m) \text{ iid} \sim \mathcal{U}[0,1], \qquad (V_1, \dots, V_n) \text{ iid} \sim \mathcal{U}[0,1],$$

[4] $\int f(u) \, \mathrm{d}u = 1, \quad \int g(v) \, \mathrm{d}v = 1$

Weighted BEDD model

Weighted networks.

- Interactions Y_{ij} may be valued ('weighted').
- Example: Y_{ij} = number of visits from insect j to plant i within a given period of time.

Weighted BEDD model: Poisson BEDD.

- $f : [0,1] \mapsto \mathbb{R}^+$: row imbalance (generalists vs specialists),
- $g : [0,1] \mapsto \mathbb{R}^+$: column imbalance (generalists vs specialists),
- $\lambda \in \mathbb{R}^+$: mean interaction intensity[4],
- Draw
$$(U_1, \dots, U_m) \text{ iid} \sim \mathcal{U}[0,1], \qquad (V_1, \dots, V_n) \text{ iid} \sim \mathcal{U}[0,1],$$
- Draw $(Y_{11}, \dots, Y_{mn})$ independently conditionally on $(U_1, \dots, U_m), (V_1, \dots, V_n)$,
$$Y_{ij} \mid U_i, V_j \sim \mathcal{P}(\lambda f(U_i) g(V_j)).$$

[4] $\int f(u)\,\mathrm{d}u = 1, \quad \int g(v)\,\mathrm{d}v = 1$

Some kernels

Mean intensity.

$$h_1 = \frac{1}{4}(Y_{11} + Y_{12} + Y_{21} + Y_{22}) \qquad \Rightarrow \qquad \mathbb{E}_{PBEDD} h_1 = \lambda,$$

$$h_3 = \frac{1}{2}(Y_{11} Y_{22} + Y_{21} Y_{12}) \qquad \Rightarrow \qquad \mathbb{E}_{PBEDD} h_3 = \lambda^2.$$

Some kernels

Mean intensity.

$$h_1 = \frac{1}{4}(Y_{11} + Y_{12} + Y_{21} + Y_{22}) \qquad \Rightarrow \qquad \mathbb{E}_{PBEDD} h_1 = \lambda,$$
$$h_3 = \frac{1}{2}(Y_{11}Y_{22} + Y_{21}Y_{12}) \qquad \Rightarrow \qquad \mathbb{E}_{PBEDD} h_3 = \lambda^2.$$

Row imbalance. Denoting $F_2 = \int f^2(u)\, \mathrm{d}u$,

$$h_2 = \frac{1}{2}(Y_{11}Y_{12} + Y_{21}Y_{22}) \quad \Rightarrow \quad \mathbb{E}_{PBEDD} h_2 = \lambda^2 F_2.$$

- Test $H_0 = \{F_2 = 1\}$ = 'no imbalance among rows'
- Test $H_0 = \{F_2^A = F_2^B\}$ = 'same degree of row imbalance in networks A and B'

Some kernels

Mean intensity.

$$h_1 = \frac{1}{4}(Y_{11} + Y_{12} + Y_{21} + Y_{22}) \quad \Rightarrow \quad \mathbb{E}_{PBEDD} h_1 = \lambda,$$

$$h_3 = \frac{1}{2}(Y_{11} Y_{22} + Y_{21} Y_{12}) \quad \Rightarrow \quad \mathbb{E}_{PBEDD} h_3 = \lambda^2.$$

Row imbalance. Denoting $F_2 = \int f^2(u)\, \mathrm{d}u$,

$$h_2 = \frac{1}{2}(Y_{11} Y_{12} + Y_{21} Y_{22}) \quad \Rightarrow \quad \mathbb{E}_{PBEDD} h_2 = \lambda^2 F_2.$$

- Test $H_0 = \{F_2 = 1\}$ = 'no imbalance among rows'
- Test $H_0 = \{F_2^A = F_2^B\}$ = 'same degree of row imbalance in networks A and B'

Over-dispersion (wrt Poisson):

$$h_4 = \frac{1}{4}(Y_{11}^2 + Y_{12}^2 + Y_{21}^2 + Y_{22}^2) \quad \Rightarrow \quad \mathbb{E}_{PBEDD} h_4 = \lambda + \lambda^2.$$

- Test $H_0 = \{\mathbb{V}(Y_{ij} \mid U_i, V_j) = \mathbb{E}(Y_{ij} \mid U_i, V_j)\}$ = 'no over-dispersion'

Variance degeneracy

Technical conditions in [LM21] impose that $\mathbb{V} U_{m,n}$ is controlled by the 'leading' covariances

$$\mathbb{C}\text{ov}(h(Y_{11}, Y_{12}, Y_{21}, Y_{22}), h(Y_{13}, Y_{14}, Y_{33}, Y_{33})) \qquad \text{(one common row)}$$

$$\text{and} \quad \mathbb{C}\text{ov}(h(Y_{11}, Y_{12}, Y_{21}, Y_{22}), h(Y_{31}, Y_{33}, Y_{41}, Y_{43})) \qquad \text{(one common column)}$$

(if not: wrong scaling for the TCL).

Variance degeneracy

Technical conditions in [LM21] impose that $\mathbb{V}U_{m,n}$ is controlled by the 'leading' covariances

$$\mathbb{C}\text{ov}(h(Y_{11}, Y_{12}, Y_{21}, Y_{22}), h(Y_{13}, Y_{14}, Y_{33}, Y_{33})) \qquad \text{(one common row)}$$

$$\text{and} \quad \mathbb{C}\text{ov}(h(Y_{11}, Y_{12}, Y_{21}, Y_{22}), h(Y_{31}, Y_{33}, Y_{41}, Y_{43})) \qquad \text{(one common column)}$$

(if not: wrong scaling for the TCL).

Not only a technical issue. To test $H_0 = \{F_2 = 1\}$ ('no imbalance among rows'), natural kernel:

$$h = h_2 - h_3 = \frac{1}{2}(Y_{11}Y_{12} + Y_{21}Y_{22} - Y_{11}Y_{22} - Y_{21}Y_{12})$$

$$\Rightarrow \quad \mathbb{E}h = \lambda^2(F_2 - 1) \overset{H_0}{=} 0$$

but then, both leading covariances are 0...

Outline

Future works

Summary.

- RCE models provide a generic framework for the analysis of bipartite networks
- The product-form BEDD model is non-naive null model
- Tests can be derived for goodness-of-fit or network comparisons

Future works

Summary.

- RCE models provide a generic framework for the analysis of bipartite networks
- The product-form BEDD model is non-naive null model
- Tests can be derived for goodness-of-fit or network comparisons

Future for network motifs.

- Unclear connection between motif and alternative H_1
- Species role in networks can be characterized by their precise position in each motif
- 'Phenologic' motif: time labeled motifs to investigate mutualistic interactions along time

Future works

Summary.

- RCE models provide a generic framework for the analysis of bipartite networks
- The product-form BEDD model is non-naive null model
- Tests can be derived for goodness-of-fit or network comparisons

Future for network motifs.

- Unclear connection between motif and alternative H_1
- Species role in networks can be characterized by their precise position in each motif
- 'Phenologic' motif: time labeled motifs to investigate mutualistic interactions along time

Future for network U-statistics.

- Better understand variance degeneracy: define a relevant Hoeffding decomposition?
- Motif counts are actually U-statistics: easier way to prove joint normality?

References I

F. Chung and L. Lu. Connected components in random graphs with given expected degree sequences. *Annals of Combinatorics*, 6(2):125–145, 2002.

A. R Cirtwill, T. Roslin, C. Rasmussen, J. M. Olesen, and D. B Stouffer. Between-year changes in community composition shape species' roles in an arctic plant–pollinator network. *Oikos*, 127(8):1163–1176, 2018.

C. Gao and J. Lafferty. Testing for global network structure using small subgraph statistics. Technical Report 1704.06742, arXiv, 2017.

G. Govaert and M. Nadif. An EM algorithm for the block mixture model. *IEEE Trans. Pattern Anal. Machine Intel.*, 27(4):643–7, 2005.

W. Hoeffding. A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, pages 293–325, 1948.

T. Le Minh. Weak convergence of U-statistics on a row-column exchangeable matrix. Technical Report 2103.12597, arXiv, 2021.

H.K Nandi and P.K Sen. On the properties of U-statistics when the observations are not independent: Part two unbiased estimation of the parameters of a finite population. *Calcutta Statistical Association Bulletin*, 12(4):124–148, 1963.

J. M Olesen, J. Bascompte, H. Elberling, and P. Jordano. Temporal dynamics in a pollination network. *Ecology*, 89(6):1573–1582, 2008.

S. Ouadah, P. Latouche, and S. Robin. Motif-based tests for bipartite networks. *Electronic Journal of Statistics*, 16(1):293 – 330, 2022.

F. Picard, J.-J. Daudin, M. Koskas, S. Schbath, and S. Robin. Assessing the exceptionality of network motifs,. *J. Comp. Biol.*, 15(1):1–20, 2008.

References II

B.I. Simmons, A. Cirtwill, N. Baker, L.V. Dicks, D.B. Stouffer, and W.J. Sutherland. Motifs in bipartite ecological networks: uncovering indirect interactions. *Oikos*, 128(2):154–170, 2019.

S. Saavedra, R. P Rohr, J. M Olesen, and J. Bascompte. Nested species interactions promote feasibility over stability during the assembly of a pollinator community. *Ecology and evolution*, 6(4):997–1007, 2016.

B. I Simmons, M. JM Sweering, M. Schillinger, L. V Dicks, W. J Sutherland, and R. Di Clemente. bmotif: A package for motif analyses of bipartite networks. *Methods in Ecology and Evolution*, 10(5):695–701, 2019.

Overlapping motifs

$\lvert\alpha^t \cap \beta^t\rvert$	0	0	1	1
$\lvert\alpha^b \cap \beta^b\rvert$	1	2	0	1

Power study: goodness-of-fit

Alternative: 2-block stochastic block model

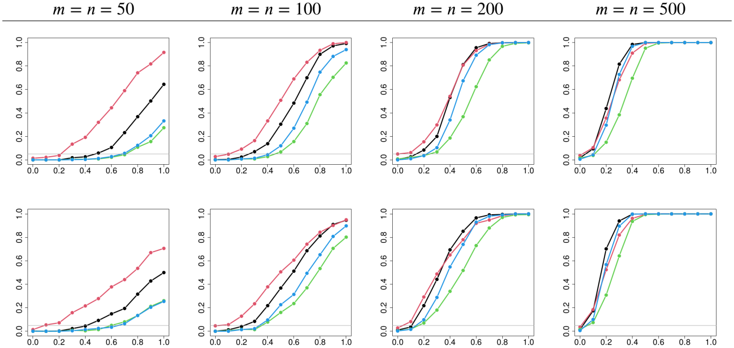


Figure 5: Empirical power of the goodness-of-fit tests, averaged over $S = 500$ simulations. Top: scenario I (easy: $\gamma_{\max} = 0.95$); bottom: scenario II (hard: $\gamma_{\max} = 0.5$). From left to right: $m = n = 50, 100, 200, 500$. Color = motif: black=5, red=6, green=10, blue=15.

Power study: network comparison

Alternative: $f^A(u) = 2u, \quad f^* \equiv 1, \quad f^B = (1-\alpha)f^A + \alpha f^*$

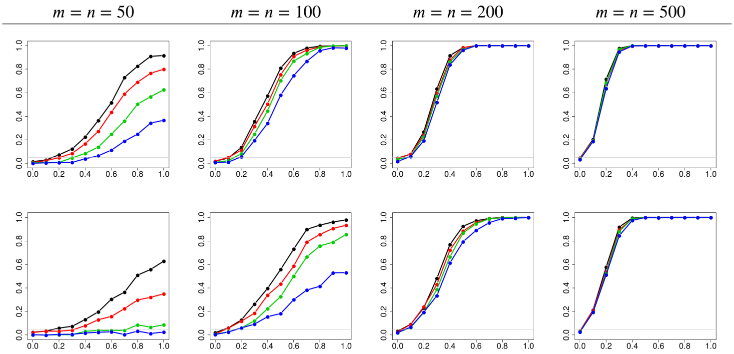


Figure 6: Empirical power of the network comparison test for $H_0 = \{g^A = g^B\}$, averaged over $S = 500$ simulations. Top: scenario I (easy); bottom: scenario II (hard). From left to right: $m = n = 50, 100, 200, 500$. Color = motif: same legend as Figure 5.