

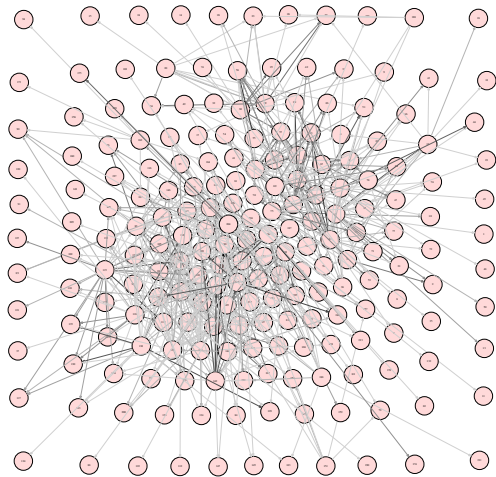
# Modèle de graphes à espace latent continu de type SBM

Jean-Benoist LEGER, BioGeCo, INRA/Université de Bordeaux

29 septembre 2015

Avec :

- David HUNTER dpt of stats, PennState.
- Cyril DUTECH BioGeCo, INRA/Université de Bordeaux.



Matting network of oaks.

Problem:

- Connectivity behavior not in the categorical form.
- More generality than OSBM (Latouche *et al.*, 2009).
- Same model as Mixed Membership BM (Airoldi *et al.* 2008).

Model:

- Latent layer 1:

$$\mathbf{W}_i \stackrel{iid}{\sim} \mathcal{D}(\alpha)$$



- Latent layer 2:

$$\mathbf{Z}_{ij} | \mathbf{W}_i \stackrel{ind}{\sim} \mathcal{M}(1; \mathbf{W}_i)$$

- Observed layer:

$$\mathbf{X}_{ij} | \mathbf{Z}_{ijq} \mathbf{Z}_{ijl} = 1 \stackrel{ind}{\sim} \mathcal{F}_{ql}^{y_{ij}}$$



Problem:

- Connectivity behavior not in the categorical form.
- More generality than OSBM (Latouche *et al.*, 2009).
- Same model as Mixed Membership BM (Airoldi *et al.* 2008).

Model:

- Latent layer 1:

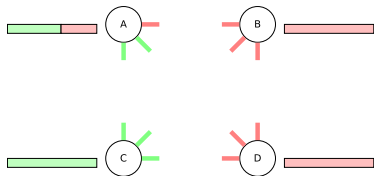
$$\mathbf{W}_i \stackrel{iid}{\sim} \mathcal{D}(\alpha)$$

- Latent layer 2:

$$\mathbf{z}_{ij} | \mathbf{W}_i \stackrel{ind}{\sim} \mathcal{M}(1; \mathbf{W}_i)$$

- Observed layer:

$$\mathbf{X}_{ij} | \mathbf{z}_{ijq} \mathbf{z}_{ijl} = 1 \stackrel{ind}{\sim} \mathcal{F}_{ql}^{y_{ij}}$$



Problem:

- Connectivity behavior not in the categorical form.
- More generality than OSBM (Latouche *et al.*, 2009).
- Same model as Mixed Membership BM (Airoldi *et al.* 2008).

Model:

- Latent layer 1:

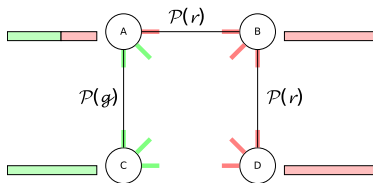
$$\mathbf{W}_i \stackrel{iid}{\sim} \mathcal{D}(\alpha)$$

- Latent layer 2:

$$\mathbf{z}_{ij} | \mathbf{W}_i \stackrel{ind}{\sim} \mathcal{M}(1; \mathbf{W}_i)$$

- Observed layer:

$$\mathbf{x}_{ij} | \mathbf{z}_{ijq} \mathbf{z}_{ijl} = 1 \stackrel{ind}{\sim} \mathcal{F}_{ql}^{y_{ij}}$$



Problem:

- Connectivity behavior not in the categorical form.
- More generality than OSBM (Latouche *et al.*, 2009).
- Same model as Mixed Membership BM (Airoldi *et al.* 2008).

Model:

- Latent layer 1:

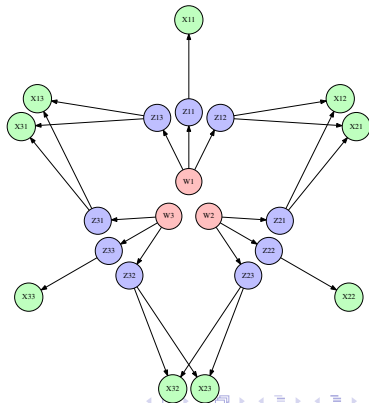
$$\mathbf{W}_i \stackrel{iid}{\sim} \mathcal{D}(\alpha)$$

- Latent layer 2:

$$\mathbf{z}_{ij} | \mathbf{W}_i \stackrel{ind}{\sim} \mathcal{M}(1; \mathbf{W}_i)$$

- Observed layer:

$$\mathbf{X}_{ij} | \mathbf{z}_{ijq} \mathbf{z}_{ijl} = 1 \stackrel{ind}{\sim} \mathcal{F}_{ql}^{y_{ij}}$$



- Complete likelihood :

$$-nB(\alpha) + \sum_{iq} (\alpha_q - 1) \log(W_{iq}) + \sum_{ijq} Z_{ijq} \log(W_{ijq}) + \sum_{ijql} Z_{ijq} Z_{jil} \log f_{ql}^{y_{ij}}(X_{ij})$$

- Variational approximation:

- $(\mathbf{W}_i | X) \sim \mathcal{D}(\gamma_i)$
- $(\mathbf{Z}_i | X) \sim \mathcal{M}(1, \tau_{ij})$
- With complete factorization.
- Parameters:  $\gamma \in \mathbf{R}_+^{n \times Q}$  et  $\tau \in [0, 1]^{n \times n \times Q}$  under the constraint  $\forall ij, \sum_q \tau_{ijq} = 1$

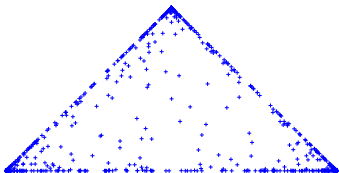
- Optimized criterion:

$$H_1(\gamma) + H_2(\tau) + PL_1(\alpha, \gamma) + PL_2(\gamma, \tau) + PL_o(\tau, \theta)$$

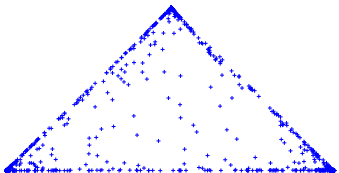
- E-step: Maximization of the criterion in respect to  $(\gamma, \tau)$ .  $n^2(Q - 1) + nQ$  parameters. L-BFGS.
- M-step: direct, or BFGS.
- Initialization method: ASC.



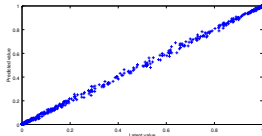
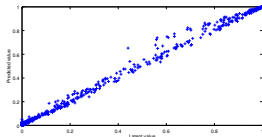
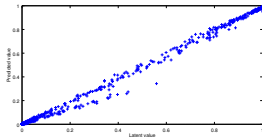
# Simulation results



Simulated latent layer 1.



Prediction (MAP) of the first latent layer.



Predicted value vs. simulated value, by component.

With Cyril Dutech, BioGeCo, INRA/Université de Bordeaux.

Data:

- Fugus: Cryphonestria, 198 individuals.
- Data on 16 locus.
- Spacial data of each individual.

Biological information:

- Clones are present..
- Hybrids between clones are present, some seem to be absent.
- Clonality and sexual reproduction seem to have different typical distance.

Question:

- What is the influence of each reproduction type in network similarity topology?
- Hybrids proportion and type can be explain only by spacial distribution with sexual and asexual reproduction?

Notations:

- $X_{ij}$ : Similarity between individual  $i$  and individual  $j$
- $d_{ij}$ : Spatial distance between individual  $i$  and individual  $j$

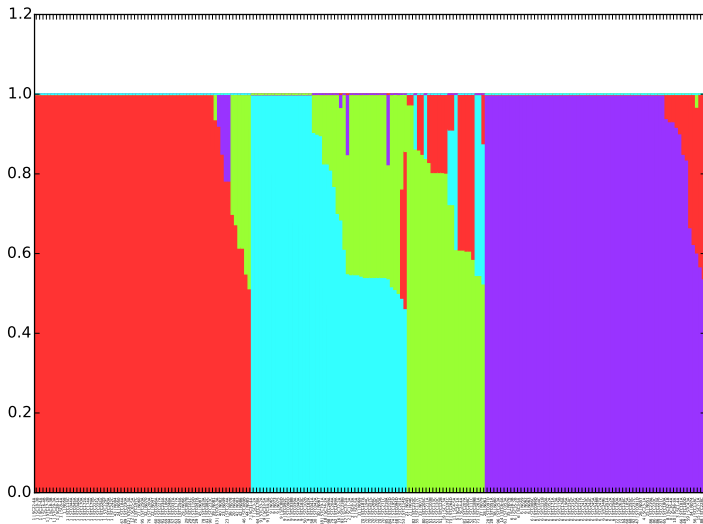
Model:

- $W_i \sim \mathcal{D}(\alpha)$
- $Z_{ij} | W_i \sim \mathcal{M}(1, W_i)$
- $X_{ij} | Z_{ijq} Z_{jil} \sim \mathcal{N}(\mu_{ql} + \beta k(d_{ij}), \sigma^2)$

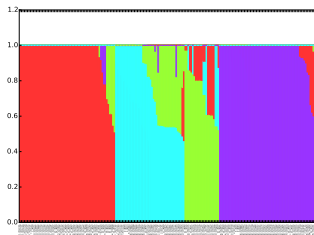
Under the constraint:

- $\forall q, \mu_{qq} = 1$

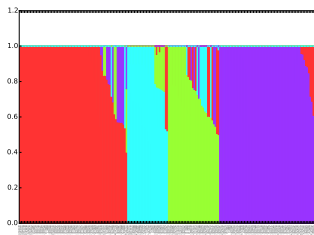
Without using the spacial distance:



Without using the spacial distance:



Using the spacial distance (typical distance use in the kernel: 300m),  
 $\hat{\beta} = -.14$



- We have typical individuals. Are they first in colonization?
- Are hybrids the results of sexual reproduction?
- We don't have enough information for computing the genetic similarity.

Work in progress.

Estimation procedure implemented in `python`

- For using numeric computation library.
- For using *nd-array* of `numpy`.
- Available on the internet in the future (on demand now).