# SMC sampling from deterministic approximations:

# Application to the Poisson stochastic block-model


S. Robin

Joint work with S. Donnet

INRA / AgroParisTech / univ. Paris Saclay


CIRM, November 2018, Marseille

# Outline

Motivating example

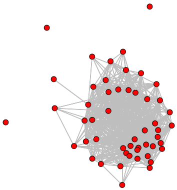
Variational EM inference

SMC sampling

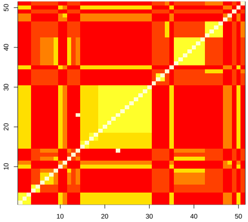
Illustrations

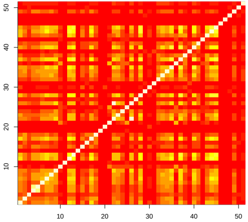
# Motivating example

Interaction network



Edge covariates

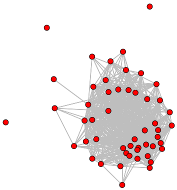


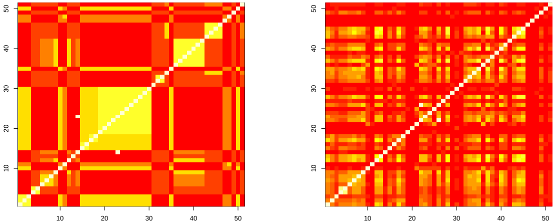

$Y_{ij}$ = number of interactions between nodes $i$ and $j$ (count)

# Motivating example

Interaction network



Edge covariates




$Y_{ij}$ = number of interactions between nodes $i$ and $j$ (count)

## Questions.

- Is there some structure in the network?
- Do the covariates contribute to explain it?
- Do they explain all of the structure? Is there some 'residual' structure?

# Stochastic block-model (SBM)

Proposed model. Poisson SBM, including covariates [MRV10]

# Stochastic block-model (SBM)

**Proposed model.** Poisson SBM, including covariates [MRV10]

**Frequentist version.**

$n$ nodes $(1 \leq i, j \leq n)$

$\{Z_i\}_i$ iid $\sim \mathcal{M}_K(1, \pi)$

$\{Y_{ij}\}_{i<j}$ independent $\mid \{Z_i\}$

$Y_{ij} \mid (Z_i = k, Z_j = \ell) \sim \mathcal{P}(\exp(\alpha_{k\ell} + x_{ij}^\intercal \beta))$

**Latent variables $Z$, parameter $\theta = (\pi, \alpha, \beta)$.**

| | | | |
|---|---|---|---|
| $\{Z_i\}$ = | node memberships | $\pi$ = | group proportions |
| $\alpha$ = | between group interactions | $\beta$ = | effects of the covariates |

# Stochastic block-model (SBM)

Proposed model. Poisson SBM, including covariates [MRV10]

| Frequentist version. | Bayesian version. |
| --- | --- |
| $n$ nodes ($1 \leq i, j \leq n$) | |
| $\{Z_i\}_i$ iid $\sim \mathcal{M}_K(1, \pi)$ | $\pi \sim \mathcal{D}_K$ |
| $\{Y_{ij}\}_{i<j}$ independent $\mid \{Z_i\}$ | |
| $Y_{ij} \mid (Z_i = k, Z_j = \ell) \sim \mathcal{P}(\exp(\alpha_{k\ell} + x_{ij}^\intercal \beta))$ | $\gamma = (\alpha, \beta) \sim \mathcal{N}$ |

Latent variables $Z$, parameter $\theta = (\pi, \alpha, \beta)$.

| | | | |
| --- | --- | --- | --- |
| $\{Z_i\}$ = | node memberships | $\pi$ = | group proportions |
| $\alpha$ = | between group interactions | $\beta$ = | effects of the covariates |

# Inference of SBM

- Bayesian inference using MCMC: time consuming + convergence issues
- Frequenstist inference via maximum likelihood (ML): intractable
- Variational approximation of ML (VEM): efficient, but with no statistical guaranty
- No easy-to-handle variational Bayes approximation (no conjugacy)

# Inference of SBM

- Bayesian inference using MCMC: time consuming + convergence issues
- Frequenstist inference via maximum likelihood (ML): intractable
- Variational approximation of ML (VEM): efficient, but with no statistical guaranty
- No easy-to-handle variational Bayes approximation (no conjugacy)

Aim.

- Design a posterior sampling taking advantage of the efficiency of (frequentist) VEM

# Outline

# EM and VEM

SBM = incomplete data model

Maximum likelihood. Most popular way: EM

$$\log p_\theta(Y) = \mathbb{E}\left(\left(\log p_\theta(Y, Z) \mid Y\right) - \mathbb{E}\left(\log p_\theta(Z \mid Y) \mid Y\right)\right.$$

$\rightarrow$ Requires to determine (some moments of) $p_\theta(Z \mid Y)$, which is intractable.

# EM and VEM

SBM = incomplete data model

Maximum likelihood. Most popular way: EM

$$\log p_\theta(Y) = \mathbb{E}\left((\log p_\theta(Y, Z) \mid Y\right) - \mathbb{E}\left(\log p_\theta(Z \mid Y) \mid Y\right)$$

$\rightarrow$ Requires to determine (some moments of) $p_\theta(Z \mid Y)$, which is intractable.

Variational approximation. When $p_\theta(Z \mid Y)$ is intractable, rather maximize

$$\begin{aligned} J(\theta, q) &= \log p_\theta(Y) - KL\left(q(Z) \| p_\theta(Z \mid Y)\right) \\ &= \mathbb{E}_q \log p_\theta(Y, Z) - \mathbb{E}_q \log q(Z) \end{aligned}$$

taking $q \in \mathcal{Q}$.

Mean field. Typical choice for SBM: $\mathcal{Q} = \{q : q(Z) = \prod_i q_i(Z_i)\}$.

# Approximate posterior from Louis approximation

Louis formulas [Lou82]. Compute Fisher information using EM side-products:

$$\partial^2_{\theta^2} \log p_\theta(Y) = \mathbb{E}\left(\partial^2_{\theta^2} \log p_\theta(Y, Z) \mid Y\right) + \mathbb{V}\left(\partial_\theta \log p_\theta(Y, Z) \mid Y\right)$$

# Approximate posterior from Louis approximation

Louis formulas [Lou82]. Compute Fisher information using EM side-products:

$$\partial^2_{\theta^2} \log p_\theta(Y) = \mathbb{E}\left(\partial^2_{\theta^2} \log p_\theta(Y, Z) \mid Y\right) + \mathbb{V}\left(\partial_\theta \log p_\theta(Y, Z) \mid Y\right)$$

Approximate variance for VEM estimates. Denote $(\widetilde{\theta}, \widetilde{q}) = \arg\max_{\theta, q \in \mathcal{Q}} J(\theta, q)$ and take

$$\begin{aligned}\mathbb{V}(\widetilde{\theta})^{-1} &\simeq \mathbb{E}_{\widetilde{q}}\left(\partial^2_{\theta^2} \log p_\theta(Y, Z)\right) + \mathbb{V}_{\widetilde{q}}\left(\partial_\theta \log p_\theta(Y, Z)\right) \\ &\simeq \mathbb{E}_{\widetilde{q}}\left(\partial^2_{\theta^2} \log p_\theta(Y, Z)\right) && \text{as } \mathbb{V}_{\widetilde{q}}(\cdot) \simeq 0.\end{aligned}$$

# Approximate posterior from Louis approximation

Louis formulas [Lou82]. Compute Fisher information using EM side-products:

$$\partial^2_{\theta^2} \log p_\theta(Y) = \mathbb{E}\left(\partial^2_{\theta^2} \log p_\theta(Y, Z) \mid Y\right) + \mathbb{V}\left(\partial_\theta \log p_\theta(Y, Z) \mid Y\right)$$

Approximate variance for VEM estimates. Denote $(\widetilde{\theta}, \widetilde{q}) = \arg\max_{\theta, q \in \mathcal{Q}} J(\theta, q)$ and take

$$\begin{aligned} \mathbb{V}(\widetilde{\theta})^{-1} &\simeq \mathbb{E}_{\widetilde{q}}\left(\partial^2_{\theta^2} \log p_\theta(Y, Z)\right) + \mathbb{V}_{\widetilde{q}}\left(\partial_\theta \log p_\theta(Y, Z)\right) \\ &\simeq \mathbb{E}_{\widetilde{q}}\left(\partial^2_{\theta^2} \log p_\theta(Y, Z)\right) && \text{as } \mathbb{V}_{\widetilde{q}}(\cdot) \simeq 0. \end{aligned}$$

Approximate posterior. Define the approximate posterior $\widetilde{p}$

$$\widetilde{p}(\pi) = \mathcal{D}(\widetilde{a}), \qquad \widetilde{p}(\gamma) = \mathcal{N}(\gamma; \widetilde{\mu}, \widetilde{\Sigma})$$

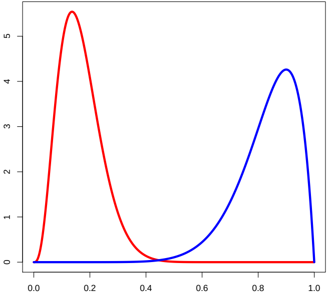
setting

$$\mathbb{E}_{\widetilde{p}}(\theta) = \widetilde{\theta}, \qquad \mathbb{V}_{\widetilde{p}}(\theta) = \mathbb{V}(\widetilde{\theta})$$

# Outline

# Sampling principle

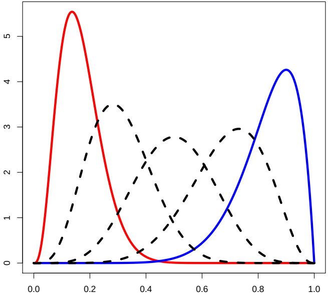
- $\widetilde{p}$ = proposal, $p^*$ = target



# Sampling principle

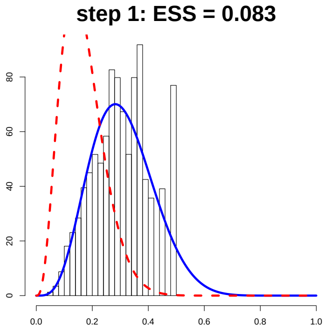
- $\widetilde{p}$ = proposal, $p^*$ = target
- Intermediate distributions

$$\widetilde{p} = p_0, p_1, \dots, p_H = p^*$$



# Sampling principle

- $\widetilde{p}$ = proposal, $p^*$ = target
- Intermediate distributions
$$\widetilde{p} = p_0, p_1, \dots, p_H = p^*$$
- Iteratively:
use $p_h$ to get a sample from $p_{h+1}$



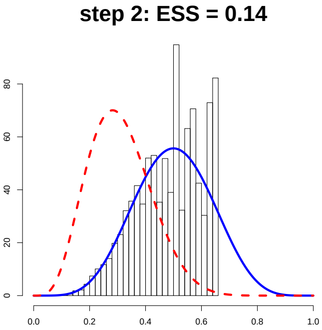
# Sampling principle

- $\widetilde{p}$ = proposal, $p^*$ = target

- Intermediate distributions

$$\widetilde{p} = p_0, p_1, \dots, p_H = p^*$$

- Iteratively:
  use $p_h$ to get a sample from $p_{h+1}$

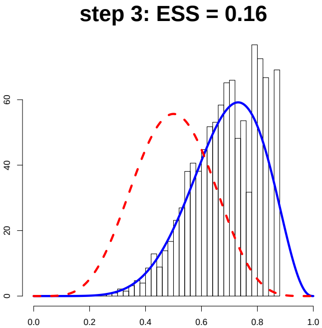


# Sampling principle

- $\widetilde{p}$ = proposal, $p^*$ = target
- Intermediate distributions

$$\widetilde{p} = p_0, p_1, \dots, p_H = p^*$$

- Iteratively:
  use $p_h$ to get a sample from $p_{h+1}$

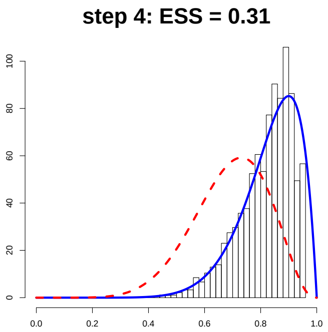


# Sampling principle

- $\widetilde{p}$ = proposal, $p^*$ = target
- Intermediate distributions

$$\widetilde{p} = p_0, p_1, \dots, p_H = p^*$$

- Iteratively:
  use $p_h$ to get a sample from $p_{h+1}$



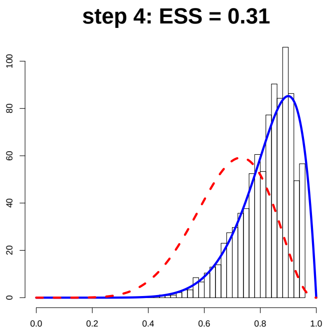
# Sampling principle

- $\widetilde{p}$ = proposal, $p^*$ = target

- Intermediate distributions

$$\widetilde{p} = p_0, p_1, \dots, p_H = p^*$$

- Iteratively:
  use $p_h$ to get a sample from $p_{h+1}$




Here. Take $p_0 = \widetilde{p}$ (rather than $p_0 = \pi$ = prior), $p^* = p(\cdot \mid Y)$

# Sequential importance sampling scheme

Denote

$$U = (\theta, Z), \qquad \pi = \text{ prior}, \qquad \ell = \text{ likelihood}$$

Distribution path: set $0 = \rho_0 < \rho_1 < \cdots < \rho_{H-1} < \rho_H = 1$,

$$\begin{aligned} p_h(U) &\propto \widetilde{p}(U)^{1-\rho_h} \times p(U|Y)^{\rho_h} \\ &\propto \widetilde{p}(U) \times r(U)^{\rho_h}, \qquad\qquad r(U) = \frac{\pi(U)\ell(Y|U)}{\widetilde{p}(U)} \end{aligned}$$

# Sequential importance sampling scheme

Denote

$$U = (\theta, Z), \qquad \pi = \text{ prior}, \qquad \ell = \text{ likelihood}$$

Distribution path: set $0 = \rho_0 < \rho_1 < \cdots < \rho_{H-1} < \rho_H = 1$,

$$p_h(U) \propto \widetilde{p}(U)^{1-\rho_h} \times p(U|Y)^{\rho_h}$$

$$\propto \widetilde{p}(U) \times r(U)^{\rho_h}, \qquad\qquad r(U) = \frac{\pi(U)\ell(Y|U)}{\widetilde{p}(U)}$$

Sequential sampling. At each step $h$, provides

$$\mathcal{E}_h = \{(U_h^m, w_h^m)\}_m = \text{ weighted sample of } p_h$$

# Sequential importance sampling scheme

Denote

$$U = (\theta, Z), \qquad \pi = \text{ prior}, \qquad \ell = \text{ likelihood}$$

Distribution path: set $0 = \rho_0 < \rho_1 < \cdots < \rho_{H-1} < \rho_H = 1$,

$$p_h(U) \propto \widetilde{p}(U)^{1-\rho_h} \times p(U|Y)^{\rho_h}$$

$$\propto \widetilde{p}(U) \times r(U)^{\rho_h}, \qquad\qquad r(U) = \frac{\pi(U)\ell(Y|U)}{\widetilde{p}(U)}$$

Sequential sampling. At each step $h$, provides

$$\mathcal{E}_h = \{(U_h^m, w_h^m)\}_m = \text{ weighted sample of } p_h$$

Question. How to tune $\{\rho_h\}$ or $H$ to keep each sampling step efficient?

# Proposed algorithm

Init.: Sample $(U_0^m)_m$ iid $\sim \widetilde{p}$, $w_0^m = 1$

[1] To avoid degeneracy. Weights set to 1 after it.

[2] $K_h$ has stationary distribution $p_h$ (e.g. Gibbs sampler). Only propagation: no convergence needed

# Proposed algorithm

Init.: Sample $(U_0^m)_m$ iid $\sim \widetilde{p}$, $w_0^m = 1$

Step $h$: Using the previous sample $\mathcal{E}_{h-1} = \{(U_{h-1}^m, w_{h-1}^m)\}$

[1] To avoid degeneracy. Weights set to 1 after it.

[2] $K_h$ has stationary distribution $p_h$ (e.g. Gibbs sampler). Only propagation: no convergence needed

# Proposed algorithm

Init.: Sample $(U_0^m)_m$ iid $\sim \widetilde{p}$, $w_0^m = 1$

Step $h$: Using the previous sample $\mathcal{E}_{h-1} = \{(U_{h-1}^m, w_{h-1}^m)\}$

1. set $\rho_h$ such that $cESS(\mathcal{E}_{h-1}; p_{h-1}, p_h) = \tau_1$

[1] To avoid degeneracy. Weights set to 1 after it.

[2] $K_h$ has stationary distribution $p_h$ (e.g. Gibbs sampler). Only propagation: no convergence needed

# Proposed algorithm

Init.: Sample $(U_0^m)_m$ iid $\sim \widetilde{p}$, $w_0^m = 1$

Step $h$: Using the previous sample $\mathcal{E}_{h-1} = \{(U_{h-1}^m, w_{h-1}^m)\}$

1. set $\rho_h$ such that $cESS(\mathcal{E}_{h-1}; p_{h-1}, p_h) = \tau_1$
2. compute $w_h^m = w_{h-1}^m \times (r_h^m)^{\rho_h - \rho_{h-1}}$

[1] To avoid degeneracy. Weights set to 1 after it.

[2] $K_h$ has stationary distribution $p_h$ (e.g. Gibbs sampler). Only propagation: no convergence needed

# Proposed algorithm

Init.: Sample $(U_0^m)_m$ iid $\sim \widetilde{p}$, $w_0^m = 1$

Step $h$: Using the previous sample $\mathcal{E}_{h-1} = \{(U_{h-1}^m, w_{h-1}^m)\}$

1. set $\rho_h$ such that $cESS(\mathcal{E}_{h-1}; p_{h-1}, p_h) = \tau_1$
2. compute $w_h^m = w_{h-1}^m \times (r_h^m)^{\rho_h - \rho_{h-1}}$
3. ([1]) if $ESS_h = \overline{w}_h^2 / \overline{w_h^2} < \tau_2$, resample the particles

[1] To avoid degeneracy. Weights set to 1 after it.

[2] $K_h$ has stationary distribution $p_h$ (e.g. Gibbs sampler). Only propagation: no convergence needed

# Proposed algorithm

Init.: Sample $(U_0^m)_m$ iid $\sim \widetilde{p}$, $w_0^m = 1$

Step $h$: Using the previous sample $\mathcal{E}_{h-1} = \{(U_{h-1}^m, w_{h-1}^m)\}$

1. set $\rho_h$ such that $cESS(\mathcal{E}_{h-1}; p_{h-1}, p_h) = \tau_1$
2. compute $w_h^m = w_{h-1}^m \times (r_h^m)^{\rho_h - \rho_{h-1}}$
3. ([1]) if $ESS_h = \overline{w}_h^2 / \overline{w_h^2} < \tau_2$, resample the particles
4. ([2]) propagate the particles $U_h^m \sim K_h(U_h^m | U_{h-1}^m)$

[1] To avoid degeneracy. Weights set to 1 after it.

[2] $K_h$ has stationary distribution $p_h$ (e.g. Gibbs sampler). Only propagation: no convergence needed

# Proposed algorithm

Init.: Sample $(U_0^m)_m$ iid $\sim \widetilde{p}$, $w_0^m = 1$

Step $h$: Using the previous sample $\mathcal{E}_{h-1} = \{(U_{h-1}^m, w_{h-1}^m)\}$

1. set $\rho_h$ such that $cESS(\mathcal{E}_{h-1}; p_{h-1}, p_h) = \tau_1$
2. compute $w_h^m = w_{h-1}^m \times (r_h^m)^{\rho_h - \rho_{h-1}}$
3. ([1]) if $ESS_h = \overline{w}_h^2 / \overline{w_h^2} < \tau_2$, resample the particles
4. ([2]) propagate the particles $U_h^m \sim K_h(U_h^m | U_{h-1}^m)$

Stop: When $\rho_h$ reaches 1.

[1] To avoid degeneracy. Weights set to 1 after it.

[2] $K_h$ has stationary distribution $p_h$ (e.g. Gibbs sampler). Only propagation: no convergence needed

# Proposed algorithm

Init.: Sample $(U_0^m)_m$ iid $\sim \widetilde{p}$, $w_0^m = 1$

Step $h$: Using the previous sample $\mathcal{E}_{h-1} = \{(U_{h-1}^m, w_{h-1}^m)\}$

1. set $\rho_h$ such that $cESS(\mathcal{E}_{h-1}; p_{h-1}, p_h) = \tau_1$
2. compute $w_h^m = w_{h-1}^m \times (r_h^m)^{\rho_h - \rho_{h-1}}$
3. ([1]) if $ESS_h = \overline{w}_h^2 / \overline{w_h^2} < \tau_2$, resample the particles
4. ([2]) propagate the particles $U_h^m \sim K_h(U_h^m | U_{h-1}^m)$

Stop: When $\rho_h$ reaches 1.

Justification [DDJ06]. At each step $h$, construct a distribution for the whole particle path with marginal $p_h$.

[1] To avoid degeneracy. Weights set to 1 after it.

[2] $K_h$ has stationary distribution $p_h$ (e.g. Gibbs sampler). Only propagation: no convergence needed

# Adaptive step size

Conditional ESS: efficiency of sample $\mathcal{E}$ from $q$ for distribution $p$

$$cESS(\mathcal{E}; q, p) = \frac{M\left(\sum_m W^m a^m\right)^2}{\sum_m W^m (a^m)^2}, \qquad a^m = \frac{p(U^m)}{q(U^m)}$$

$\rightarrow$ Step 1: find next $p_h$ s.t. sample $\mathcal{E}_{h-1}$ is reasonably efficient.

# Adaptive step size

Conditional ESS: efficiency of sample $\mathcal{E}$ from $q$ for distribution $p$

$$cESS(\mathcal{E}; q, p) = \frac{M \left(\sum_m W^m a^m\right)^2}{\sum_m W^m (a^m)^2}, \qquad a^m = \frac{p(U^m)}{q(U^m)}$$

→ Step 1: find next $p_h$ s.t. sample $\mathcal{E}_{h-1}$ is reasonably efficient.

Update formula of the weights

$$cESS(\mathcal{E}_{h-1}; p_{h-1}, p_h) = \frac{M \left[\sum_m W_{h-1}^m \left(r_{h-1}^m\right)^{\rho_h - \rho_{h-1}}\right]^2}{\sum_m W_{h-1}^m \left(r_{h-1}^m\right)^{2\rho_h - 2\rho_{h-1}}}$$

→ can be computed for any $\rho_h$ before sampling.

→ $\rho_h$ tuned to meet $\tau_1$, which controls the step size $\rho_h - \rho_{h-1}$ (and $H$)

# Marginal likelihood

Denote

$$\gamma_h(U) = \widetilde{p}(U)\alpha(U)^{\rho_h}, \qquad Z_h = \int \gamma_h(U) \, \mathrm{d}U, \qquad p_h = \gamma_h(U)/Z_h$$

The marginal likelihood is given by

$$p(Y) = \int \pi(U)\ell(Y|U) \, \mathrm{d}U = \int \gamma_H(U) \, \mathrm{d}U = Z_H$$

which can be estimated with

$$\widehat{\left(\frac{Z_H}{Z_0}\right)} = \prod_{h=1}^{H} \widehat{\left(\frac{Z_h}{Z_{h-1}}\right)} \qquad \text{where} \qquad \widehat{\left(\frac{Z_h}{Z_{h-1}}\right)} = \sum_m W_h^m (\alpha_h^m)^{\rho_h - \rho_{h-1}}$$

# Outline

Motivating example

Variational EM inference

SMC sampling

Illustrations

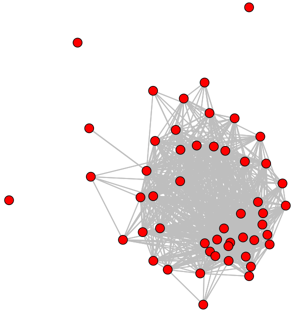
# Tree network

From [VPDL08].

$n = 51$ tree species

3 covariates (distances):
taxonomy, geography, genetics

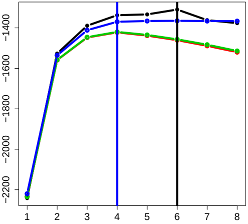
$Y_{ij}$ = number of shared fungal parasites



# Sampling path & choice of $K$

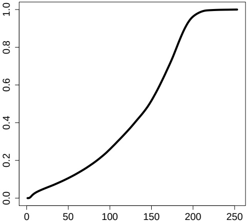
Full model. All covariates

$\widehat{p}(K \mid Y)$



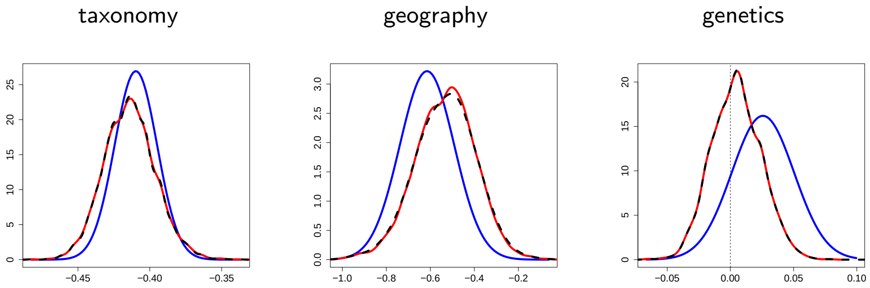

$J_K$, $\widetilde{BIC}$, $\widetilde{ICL}$

Sampling path: $\rho_h$



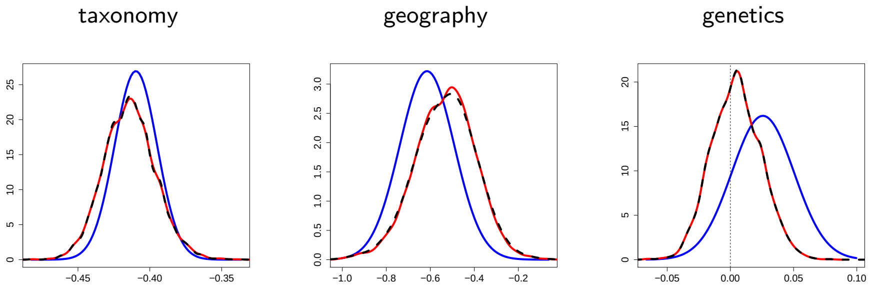

$\widehat{K} = \arg\max_K \widehat{p}(K \mid Y)$

# Posterior distribution of $\beta$




$$\color{blue}{\widetilde{p}(\beta \mid \widehat{K})}, \quad \color{red}{\widehat{p}(\beta \mid Y, \widehat{K})}, \quad \color{black}{\widehat{p}(\beta \mid Y) = \sum_K \widehat{p}(K \mid Y)\widehat{p}(\beta \mid Y, K)}$$

# Posterior distribution of $\beta$




$$\widetilde{p}(\beta \mid \widehat{K}), \quad \widehat{p}(\beta \mid Y, \widehat{K}), \quad \widehat{p}(\beta \mid Y) = \sum_K \widehat{p}(K \mid Y)\widehat{p}(\beta \mid Y, K)$$

| Correlation between estimates. | $(\beta_1, \beta_2)$ | $(\beta_1, \beta_3)$ | $(\beta_2, \beta_3)$ |
|---|---|---|---|
| $\widetilde{p}(\beta)$ | −0.012 | 0.021 | 0.318 |
| $\widehat{p}(\beta \mid Y)$ | −0.274 | −0.079 | −0.088 |

Model selection. $\widehat{P}\{x = (\text{taxo., geo.}) \mid Y\} \simeq 70\%$, $\widehat{P}\{x = (\text{taxo.}) \mid Y\} \simeq 30\%$

# Residual structure

Between group interactions $(\alpha_{k\ell})$ = 'residuals' = not explained by the covariates.

[3]with increasing marginal $\overline{\phi}(u) = \int \phi(u, v) \, dv$ to ensure identifiability.

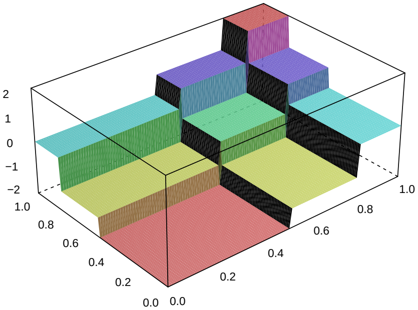
# Residual structure

Between group interactions $(\alpha_{k\ell})$ = 'residuals' = not explained by the covariates.

'Graphon' representation. [LRO17]
Group interactions encoded as

$$\phi : [0,1]^2 \mapsto \mathbb{R}$$

- symmetric[3],
- block-wise constant,
- block width = $\pi_k$
- block height = $\alpha_{k\ell}$




[3]with increasing marginal $\overline{\phi}(u) = \int \phi(u,v)\, dv$ to ensure identifiability.

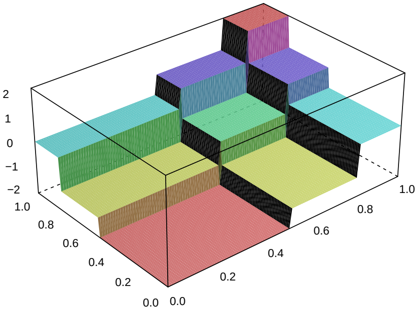
# Residual structure

Between group interactions $(\alpha_{k\ell})$ = 'residuals' = not explained by the covariates.

'Graphon' representation. [LRO17]
Group interactions encoded as

$$\phi : [0,1]^2 \mapsto \mathbb{R}$$

- symmetric[3],
- block-wise constant,
- block width $= \pi_k$
- block height $= \alpha_{k\ell}$




Same representation for all $K$. $Y_{ij} | (U_i, U_j) \sim \mathcal{P}\left(\exp(\phi(U_i, U_j) + x_{ij}^\top \beta\right)$

[3] with increasing marginal $\overline{\phi}(u) = \int \phi(u, v) \, dv$ to ensure identifiability.

# Tree network residual structure

Residual graphon.
Each particle $\theta^m$ provides an estimate of $\phi^m(u, v)$

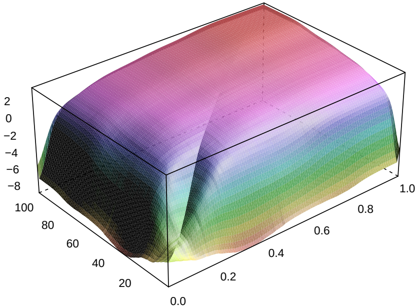
All estimates can be averaged (over both $m$ and $K$)

# Tree network residual structure

Residual graphon.

Each particle $\theta^m$ provides an estimate of $\phi^m(u, v)$

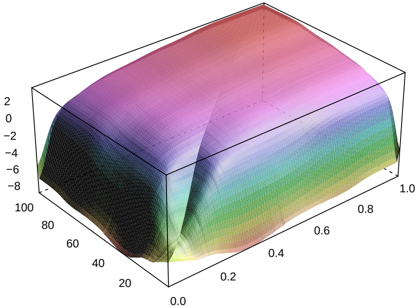
All estimates can be averaged (over both $m$ and $K$)



# Tree network residual structure

**Residual graphon.**
Each particle $\theta^m$ provides an estimate of $\phi^m(u, v)$

All estimates can be averaged (over both $m$ and $K$)




**Interpretation.**

- A remaining individual effect (some species interact more than other in average)
- A small fraction of species interact much less than expected.

# Social network of equid species

2 datasets [RSF+15].

- $n = 28$ zebras, $n = 29$ onagers
- sex and age (juvenile / adult) recorded

# Social network of equid species

2 datasets [RSF+15].

- $n = 28$ zebras, $n = 29$ onagers
- sex and age (juvenile / adult) recorded

Model comparison.

Zebras:

$$\widehat{P}(x = (\text{sex}) \mid Y) \simeq 1$$

Onagers:

$$\widehat{P}(x = (\text{sex} \times \text{age}) \mid Y) \simeq 1$$

# Social network of equid species

**2 datasets** [RSF+15].

- $n = 28$ zebras, $n = 29$ onagers
- sex and age (juvenile / adult) recorded

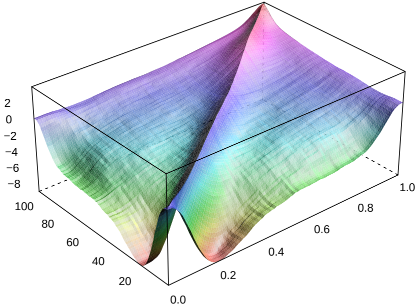
**Model comparison.**

Zebras:

$$\widehat{P}(x = (\text{sex}) \mid Y) \simeq 1$$

Onagers:

$$\widehat{P}(x = (\text{sex} \times \text{age}) \mid Y) \simeq 1$$




Onager network: residual structure

# Discussion

## Rational.

- Frequentist VEM side-product can be used to define an approximate posterior
- SMC sampling can start from there to the sample from the posterior

# Discussion

## Rational.

- Frequentist VEM side-product can be used to define an approximate posterior
- SMC sampling can start from there to the sample from the posterior

## Open problems. (About dig data...)

- Louis approximate prior $\widetilde{p}$ is not that bad. Still, numerous steps are needed to reach the posterior
  ... because of the large dimension of $U = (\theta, Z)$
- Especially true for (uselessly) large $K$
  ... but VEM inference can not be trusted to choose it

# References


P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B*, 68(3):411–436, 2006.

S. Donnet and S. Robin. Using deterministic approximations to accelerate SMC for posterior sampling. Technical Report 1612.06928, arXiv, 2017.

T. A. Louis. Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society. Series B*, pages 226–233, 1982.

P. Latouche, S. Robin, and S. Ouadah. Goodness of fit of logistic regression models for random graphs. *Journal of Computational and Graphical Statistics*, (just-accepted), 2017.

M. Mariadassou, S. Robin, and C. Vacher. Uncovering latent structure in valued graphs: a variational approach. *The Annals of Applied Statistics*, pages 715–742, 2010.

D. I Rubenstein, S. R Sundaresan, I. R Fischhoff, C. Tantipathananandh, and T. Y Berger-Wolf. Similar but different: dynamic social network analysis highlights fundamental differences between the fission-fusion societies of two equid species, the onager and Grevy's zebra. *PloS one*, 10(10):e0138645, 2015.

C. Vacher, D. Piou, and M.-L. Desprez-Loustau. Architecture of an antagonistic tree/fungus network: The asymmetric influence of past evolutionary history. *PLoS ONE*, 3(3):1740, 2008. `e1740`. `doi:10.1371/journal.pone.0001740`.

# Theoretical justification

At each step $h$, [DDJ06] construct a distribution for the whole particle path with marginal $p_h$.

- $\overline{p}_h(\theta_{0:h})$ distribution of the particle path

$$\overline{p}_h(\theta_{0:h}) \propto p_h(\theta_h) \prod_{k=1}^{h} L_k(\theta_{k-1}|\theta_k)$$

- $L_h$ = backward kernel

$$L_h(\theta_{h-1}|\theta_h) = K_h(\theta_h|\theta_{h-1}) p_h(\theta_{h-1}) / p_h(\theta_h)$$

- Update for the weights

$$w_h(\theta_{0:h}) = w_{h-1}(\theta_{0:h-1}) \alpha(\theta_h)^{\rho_h - \rho_{h-1}}$$

# Some comments

**Resampling (optional step 3).**

- avoids degeneracy
- set weights $w_h^m = 1$ after resampling

**Propagation kernel $K_h$ (step 4).**

- with stationary distribution $p_h$ (e.g. Gibbs sampler)
- just propagation: does not change the distribution $\rightarrow$ no convergence needed

# Some comments

**Resampling (optional step 3).**

- avoids degeneracy
- set weights $w_h^m = 1$ after resampling

**Propagation kernel $K_h$ (step 4).**

- with stationary distribution $p_h$ (e.g. Gibbs sampler)
- just propagation: does not change the distribution $\rightarrow$ no convergence needed

**Theoretical justification:** [DDJ06]. At each step $h$, construct a distribution for the whole particle path with marginal $p_h$.