Goodness of fit for graph logistic regression:
a (variational) Bayes approach

S. Robin

Joint work with S. Donnet P. Latouche, S. Ouadah.

Bayes Club, Dec. 2017, Leiden
Motivating example

Tree interaction network

Data.

- $n$ tree species
- network

\[ i \sim j \iff \text{species } i \text{ and } j \text{ share at least one parasite} \]

- $d$ covariates: genetic, phylogenetic and geographic distance between each pair of species

Questions.

- Do covariates explain why species interact?
- Do the covariates explain the whole topology of the network?
- If not, how to depict the residual structure of the network?
Motivating example

Tree network: the data

Genetic dist.

Geographic dist.

Taxonomic dist.

Network

Network
Outline

Motivating example

SBM-reg model

Variational Bayes inference

Bridge sampling

Some simulations

Back to the original problem

Discussion
Statistical model

Aim
1. Account for the effect of (edge) covariates
2. Account for some network-structured residual

Idea: combine two classical models
1. Logistic regression: \((Y_{ij})_{i < j}\) indep.,
   \[\text{logit} \ P(Y_{ij} = 1) = x_{ij}^\top \beta\]

2. \textit{W}-graph: \((U_i)_i\) iid \(\sim \mathcal{U}_{[0,1]}\), \((Y_{ij})_{i < j}\) indep. \(\mid (U_i)_i\),
   \[\text{logit} \ P(Y_{ij} = 1 \mid U_i, U_j) = \phi(U_i, U_j)\]
   \(\phi = (\text{logit-})\text{graphon function}\)
$W$-graph model \cite{LS06}

- Latent variables:
  \[
  (U_i) \text{ iid } \sim U_{[0,1]},
  \]

- Graphon function:
  \[
  \phi(u, v) : [0, 1]^2 \rightarrow [0, 1]
  \]

Edges:
\[
P(Y_{ij} = 1 | U_i, U_j) = \phi(U_i, U_j)
\]
SBM-reg model

Stochastic block-model [HL79,NS01]

- Latent variables:
  \[(Z_i) \text{ iid } \sim \mathcal{M}(1, \pi)\]

- Blockwise constant graphon:
  \[
  \phi(z, z') = \sum_{k, \ell} \alpha_{k\ell} \mathbb{1}\{z \in B_k, z' \in B_\ell\}
  \]

- Edges:
  \[
  P(Y_{ij} = 1|Z_i, Z_j) = \phi(Z_i, Z_j)
  \]
Stochastic block-model [HL79,NS01]

- Latent variables:
  \[(Z_i) \text{ iid } \sim \mathcal{M}(1, \pi)\]

- Blockwise constant graphon:
  \[
  \phi(z, z') = \sum_{k, \ell} \alpha_{k\ell} \mathbb{1}\{z \in B_k, z' \in B_\ell\}
  \]

- Edges:
  \[
  P(Y_{ij} = 1|Z_i, Z_j) = \phi(Z_i, Z_j)
  \]
Proposed model [LRO17]

'Ideal' model: \((U_i)_i \text{iid} \sim \mathcal{U}[0,1], \quad (Y_{ij})_{i<j} \text{ indep.} \mid (U_i)_i,\)

\[
\text{logit } P(Y_{ij} = 1 \mid U_i, U_j) = \underbrace{x_{ij}^T \beta}_{\text{covariate effect}} + \underbrace{\phi(U_i, U_j)}_{\text{residual structure}}
\]

→ Fully non-parametric estimation of \(\phi\) not easy
Proposed model [LRO17]

'Ideal' model: \((U_i)_i \text{iid} \sim \mathcal{U}_{[0,1]}, \quad (Y_{ij})_{i<j} \text{ indep.} \mid (U_i)_i,\)

\[
\text{logit} \, P(Y_{ij} = 1 \mid U_i, U_j) = x_{ij}^\top \beta + \phi(U_i, U_j)
\]

\(\rightarrow\) Fully non-parametric estimation of \(\phi\) not easy

SBM-reg model: Block-wise constant approximation of \(\phi\) [LR16]

\((Z_i)_i \text{iid} \sim \mathcal{M}_K(1, \pi), \quad (Y_{ij})_{i<j} \text{ indep.} \mid (Z_i)_i,\)

\[
\text{logit} \, P(Y_{ij} = 1 \mid Z_i, Z_j) = x_{ij}^\top \beta + \alpha_{Z_i, Z_j}
\]
Full Bayesian model

- $\beta \sim \mathcal{N}(\beta)$: regression coefficients
- $K \sim P(K)$: number of blocks
- $\pi \sim \mathcal{D}(\pi | K)$: block proportions
- $Z \sim \mathcal{M}(Z | \pi)$: block memberships
- $\alpha \sim \mathcal{N}(\alpha | K)$: block interactions
- $Y \sim \mathcal{B}(Y | Z, \alpha, \beta)$: network

Notations:
$x = (x_{ij}^\ell), Z = (Z_i), Y = (Y_{ij}), \theta = (\beta, \pi, \alpha)$
Back to the original problem

Questions.

- Do covariates explain why species interact?
  \[ p(\beta \mid Y) \]

- Do the covariates explain the whole topology of the network?
  \[ p(K = 1 \mid Y) \]

- If not, how to depict the residual structure of the network?
  \[ \hat{\phi}(u, v) = \sum_k P(K = k \mid Y) \mathbb{E}[\phi(u, v) \mid Y, K = k] \]
Outline

Motivating example

SBM-reg model

Variational Bayes inference

Bridge sampling

Some simulations

Back to the original problem

Discussion
A reminder on variational Bayes

**Aim:** approximate the posterior

\[ \tilde{p}(\theta) \approx p(\theta \mid Y), \]

typically

\[ \tilde{p}(\cdot) = \arg \min_{q \in Q} KL[q(\cdot) \mid \mid p(\cdot \mid Y)]. \]

**Models with hidden variables** [BG03]: get

\[ \tilde{p}(\theta, Z) \approx p(\theta, Z \mid Y), \]

typically, assume that \( \tilde{p}(\theta, Z) = \tilde{p}_\theta(\theta) \tilde{p}_Z(Z) \rightarrow \text{VB-EM}. \)
Two examples

Logistic regression [JJ00]: \( Q = \mathcal{N}, \)

\[
\beta \sim \mathcal{N}(m, S), \quad Y_i \mid \beta \sim \mathcal{B}\left(1 \div \frac{1}{1 - e^{x_i^T \beta}}\right): \quad \tilde{p}(\beta) = \mathcal{N}(\tilde{m}, \tilde{S})
\]

SBM [DPR08,GDR12,LR16]: \( Q = \{\text{factorized distribution}\} \)

\[
\tilde{p}(Z) = \prod_i \tilde{p}_i(Z_i)
\]

→ mean-field approximation (R package \texttt{mixer} on CRAN)

Both perform well (both empirical and theoretical arguments)
VB inference for SBM-reg

**VB-EM:** a VB-EM algorithm can be designed [LRO17] to get

\[
\tilde{p}^K(\theta, Z) = \tilde{p}_\theta^K(\theta) \tilde{p}_Z^K(Z) \approx p(\theta, Z|Y, K)
\]

**Posterior of \(K\):** VB approximation also holds to get [VMMR12]

\[
\tilde{p}(K) \approx p(K|Y)
\]

**Bayesian model averaging** then applies to get, e.g.

\[
\tilde{p}(\theta) = \sum_k \tilde{P}(K = k) \tilde{p}_\theta^k(\theta)
\]

**R package** GOFnetwork (github.com/platouche/gofNetwork)
Tree network

Covariate effects: $\tilde{p}(\beta)$

<table>
<thead>
<tr>
<th></th>
<th>genet.</th>
<th>geogr.</th>
<th>taxo.</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>4.6 $10^{-5}$</td>
<td>2.3 $10^{-1}$</td>
<td>-9.0 $10^{-1}$</td>
</tr>
<tr>
<td>st. dev.</td>
<td>1.5 $10^{-5}$</td>
<td>2.1 $10^{-1}$</td>
<td>4.2 $10^{-2}$</td>
</tr>
<tr>
<td>ratio</td>
<td>3.1</td>
<td>1.1</td>
<td>-21</td>
</tr>
</tbody>
</table>

Goodness of fit:

$\tilde{P}(K = 1) = 4.83$
Tree network

Covariate effects: $\tilde{p}(\beta)$

<table>
<thead>
<tr>
<th></th>
<th>genet.</th>
<th>geogr.</th>
<th>taxo.</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>4.6 $10^{-5}$</td>
<td>2.3 $10^{-1}$</td>
<td>-9.0 $10^{-1}$</td>
</tr>
<tr>
<td>st. dev.</td>
<td>1.5 $10^{-5}$</td>
<td>2.1 $10^{-1}$</td>
<td>4.2 $10^{-2}$</td>
</tr>
<tr>
<td>ratio</td>
<td>3.1</td>
<td>1.1</td>
<td>-21</td>
</tr>
</tbody>
</table>

Goodness of fit:

$\tilde{P}(K = 1) = 4.83 \, 10^{-153}$ (!!!)
Tree network

Covariate effects: $\tilde{p}(\beta)$

<table>
<thead>
<tr>
<th></th>
<th>genet.</th>
<th>geogr.</th>
<th>taxo.</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>$4.6 \times 10^{-5}$</td>
<td>$2.3 \times 10^{-1}$</td>
<td>$-9.0 \times 10^{-1}$</td>
</tr>
<tr>
<td>st. dev.</td>
<td>$1.5 \times 10^{-5}$</td>
<td>$2.1 \times 10^{-1}$</td>
<td>$4.2 \times 10^{-2}$</td>
</tr>
<tr>
<td>ratio</td>
<td>3.1</td>
<td>1.1</td>
<td>-21</td>
</tr>
</tbody>
</table>

Goodness of fit:

$\tilde{P}(K = 1) = 4.83 \times 10^{-153}$ (!!!)

Covariates partially explain the network topology
Tree network

Covariate effects: $\hat{p}(\beta)$

<table>
<thead>
<tr>
<th></th>
<th>genet.</th>
<th>geogr.</th>
<th>taxo.</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>$4.6 \times 10^{-5}$</td>
<td>$2.3 \times 10^{-1}$</td>
<td>$-9.0 \times 10^{-1}$</td>
</tr>
<tr>
<td>st. dev.</td>
<td>$1.5 \times 10^{-5}$</td>
<td>$2.1 \times 10^{-1}$</td>
<td>$4.2 \times 10^{-2}$</td>
</tr>
<tr>
<td>ratio</td>
<td>3.1</td>
<td>1.1</td>
<td>-21</td>
</tr>
</tbody>
</table>

Goodness of fit:

$\tilde{P}(K = 1) = 4.83 \times 10^{-153}$ (!!!)

Covariates partially explain the network topology
Tree network

Covariate effects: $\hat{p}(\beta)$

<table>
<thead>
<tr>
<th></th>
<th>genet. $4.6 \times 10^{-5}$</th>
<th>geogr. $2.3 \times 10^{-1}$</th>
<th>taxo. $-9.0 \times 10^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>$1.5 \times 10^{-5}$</td>
<td>$2.1 \times 10^{-1}$</td>
<td>$4.2 \times 10^{-2}$</td>
</tr>
<tr>
<td>st. dev.</td>
<td>$3.1$</td>
<td>$1.1$</td>
<td>$-21$</td>
</tr>
<tr>
<td>ratio</td>
<td>$3.1$</td>
<td>$1.1$</td>
<td>$-21$</td>
</tr>
</tbody>
</table>

Goodness of fit:

$\tilde{P}(K = 1) = 4.83 \times 10^{-153}$ (!!!)

Covariates partially explain the network topology
Outline

Motivating example

SBM-reg model

Variational Bayes inference

Bridge sampling

Some simulations

Back to the original problem

Discussion
From VB approximation to the true posterior

**VBEM** only provides an approximation of the true posterior

\[ \tilde{p}(\theta) \approx p(\theta|Y), \]

→ Can we use it to evaluate the true posterior?
From VB approximation to the true posterior

VBEM only provides an approximation of the true posterior

\[ \tilde{p}(\theta) \approx p(\theta | Y), \]

→ Can we use it to evaluate the true posterior?

First idea = Importance sampling (IS): \((\theta^m) \text{ iid } \sim q,\)

\[ \mathbb{E}[f(\theta) | Y] = \sum_m W^m f(\theta^m), \quad w^m = \frac{p(\theta^m | Y)}{q(\theta^m)}, \quad W^m = \frac{w^m}{\sum_\ell w^\ell} \]

→ Use \(\tilde{p}\) as the proposal distribution \(q\).
Importance of the proposal

Effective sample size $= ESS := \overline{w^2}/\overline{w^2}$.
Bridge sampling\(^1\) principle

- \(q = \text{proposal}, \ p^* = \text{target}\)

\(^1\)‘Bridge sampling’ = ‘Sequential importance sampling’ (∈ ‘SMC’)
Bridge sampling$^1$ principle

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

- Define intermediate distributions

\[ q = p_0, p_1, \ldots, p_H = p^* \]

---

$^1$'Bridge sampling' = 'Sequential importance sampling' (∈ 'SMC')
Bridge sampling\(^1\) principle

- \(q = \) proposal, \(p^* = \) target
- Define intermediate distributions

\[ q = p_0, p_1, \ldots, p_H = p^* \]

- Iteratively:
  use IS from \(p_h\) to \(p_{h+1}\)

---

\(^1\)'Bridge sampling' = 'Sequential importance sampling' (∈ 'SMC')
Bridge sampling principle

- \( q = \) proposal, \( p^* = \) target
- Define intermediate distributions
  \[ q = p_0, p_1, \ldots, p_H = p^* \]
- Iteratively:
  use IS from \( p_h \) to \( p_{h+1} \)

---

1 'Bridge sampling' = 'Sequential importance sampling' (∈ 'SMC')
Bridge sampling\(^1\) principle

- \( q = \) proposal, \( p^* = \) target
- Define intermediate distributions
  \[ q = p_0, p_1, \ldots, p_H = p^* \]
- Iteratively:
  use IS from \( p_h \) to \( p_{h+1} \)

\(^1\)‘Bridge sampling’ = ‘Sequential importance sampling’ (∈ ‘SMC’)

S. Robin (INRA / AgroParisTech)  Bayesian GOF for logistic graphs  Dec. 2017, Leiden  19 / 38
Bridge sampling\(^1\) principle

- \( q = \) proposal, \( p^* = \) target
- Define intermediate distributions
  \[ q = p_0, p_1, \ldots, p_H = p^* \]
- Iteratively:
  use IS from \( p_h \) to \( p_{h+1} \)

---

\(^1\)'Bridge sampling' = 'Sequential importance sampling' (∈ 'SMC')
Bridge sampling\(^1\) principle

- \( q = \) proposal, \( p^* = \) target

- Define intermediate distributions
  \[
  q = p_0, p_1, \ldots, p_H = p^*
  \]

- Iteratively:
  use IS from \( p_h \) to \( p_{h+1} \)

Application:

\[
q = \tilde{p}_Y, \quad p^* = p(\cdot | Y)
\]

\(^1\)’Bridge sampling’ = 'Sequential importance sampling' (∈ 'SMC')
Path sampling

Distribution path\(^2\): set \(0 = \rho_0 < \rho_1 < \cdots < \rho_{H-1} < \rho_H = 1\),

\[
p_h(\theta) \propto \tilde{p}_Y(\theta)^{1-\rho_h} \times p(\theta|Y)^{\rho_h}
\]

\[
\propto \tilde{p}_Y(\theta) \times \alpha(\theta)^{\rho_h}, \quad \alpha(\theta) = \frac{\pi(\theta)\ell(Y|\theta)}{\tilde{p}_Y(\theta)}
\]

\(^2\)\cite{Nea01}: \(p_h(\theta) \propto \pi(\theta)\ell(Y|\theta)^{\rho_h}\), i.e. \(\tilde{p}_Y = \pi\)
Path sampling

Distribution path\(^2\): set \(0 = \rho_0 < \rho_1 < \cdots < \rho_{H-1} < \rho_H = 1\),

\[
p_h(\theta) \propto \tilde{p}_Y(\theta)^{1-\rho} \times p(\theta|Y)^{\rho_h}
\]

\[
\propto \tilde{p}_Y(\theta) \times \alpha(\theta)^{\rho_h}, \quad \alpha(\theta) = \frac{\pi(\theta) \ell(Y|\theta)}{\tilde{p}_Y(\theta)}
\]

Aim of bridge sampling: at each step \(h\), provide

\[
\mathcal{E}_h = \{(\theta^m_h, w^m_h)\}_{m} = \text{weighted sample of } p_h
\]

\(^2\text{[Nea01]}\): \(p_h(\theta) \propto \pi(\theta) \ell(Y|\theta)^{\rho_h}\), i.e. \(\tilde{p}_Y = \pi\)
Path sampling

Distribution path\(^2\): set \(0 = \rho_0 < \rho_1 < \cdots < \rho_{H-1} < \rho_H = 1\),

\[
p_h(\theta) \propto \tilde{p}_Y(\theta)^{1-\rho_h} \times p(\theta|Y)^{\rho_h} \\
\propto \tilde{p}_Y(\theta) \times \alpha(\theta)^{\rho_h}, \quad \alpha(\theta) = \frac{\pi(\theta)\ell(Y|\theta)}{\tilde{p}_Y(\theta)}
\]

Aim of bridge sampling: at each step \(h\), provide

\[\mathcal{E}_h = \{(\theta^m_h, w^m_h)\}_m = \text{weighted sample of } p_h\]

Questions

\[\begin{align*}
\text{Step number } H \ ? \\
\text{Step size } \rho_h - \rho_{h-1} \ ? \\
\text{How to actually sample } p_h \text{ from the sample } \mathcal{E}_{h-1} \ ?
\end{align*}\]

\(^2\text{[Nea01]: } p_h(\theta) \propto \pi(\theta)\ell(Y|\theta)^{\rho_h}, \text{ i.e. } \tilde{p}_Y = \pi\]
Bridge sampling algorithm [DR17]

Init.: Sample \((\theta^m_0)_m\) iid \(\sim \tilde{p}_Y\), \(w^m_0 = 1\)
Bridge sampling algorithm [DR17]

Init.: Sample \((\theta^m_0)_m\) iid \(\sim \tilde{p}_Y\), \(w^m_0 = 1\)

Step \(h\): Using the previous sample \(\mathcal{E}_{h-1} = \{(\theta^m_{h-1}, w^m_{h-1})\}\)
Bridge sampling algorithm [DR17]

Init.: Sample $(\theta^m_0)_m \ iid \sim \tilde{p}_Y$, $w^m_0 = 1$

Step $h$: Using the previous sample $\mathcal{E}_{h-1} = \{(\theta^m_{h-1}, w^m_{h-1})\}$

1. set $\rho_h$ such that $cESS(\mathcal{E}_{h-1}; \rho_{h-1}, \rho_h) = \tau_1$
Bridge sampling algorithm [DR17]

Init.: Sample \((\theta_0^m)_m\) iid \(\sim \tilde{p}_Y\), \(w_0^m = 1\)

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

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

2. compute \(w_h^m = w_{h-1}^m \times (\alpha_h^m)^{\rho_h - \rho_{h-1}}\)
Bridge sampling algorithm [DR17]

Init.: Sample \((\theta^m_0)_m\) iid \(\sim \tilde{p}_Y\), \(w^m_0 = 1\)

Step \(h\): Using the previous sample \(\mathcal{E}_{h-1} = \{ (\theta^m_{h-1}, w^m_{h-1}) \} \)

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

2. compute \(w^m_h = w^m_{h-1} \times (\alpha^m_h)^{\rho_h-\rho_{h-1}}\)

3. if \(ESS_h = \frac{\overline{w}_h^2}{\overline{w}_h^2} < \tau_2\), resample the particles
Bridge sampling algorithm [DR17]

Init.: Sample $(\theta_0^m)_m$ iid $\sim \tilde{\rho}_Y$, $w_0^m = 1$

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

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

2. compute $w_h^m = w_{h-1}^m \times (\alpha_h^m)^{\rho_h - \rho_{h-1}}$

3. if $ESS_h = \frac{w_h^2}{\bar{w}_h^2} < \tau_2$, resample the particles

4. propagate the particles $\theta_h^m \sim K_h(\theta_h^m|\theta_{h-1}^m)$
Bridge sampling algorithm [DR17]

Init.: Sample \((\theta_0^m)^m\) iid \(\sim \tilde{\rho}_Y\), \(w_0^m = 1\)

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

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

2. compute \(w_h^m = w_{h-1}^m \times (\alpha_h^m)^{\rho_h-\rho_{h-1}}\)

3. if \(ESS_h = \bar{w}_h^2/\bar{w}_h^2 < \tau_2\), resample the particles

4. propagate the particles \(\theta_h^m \sim K_h(\theta_h^m|\theta_{h-1}^m)\)

Stop: When \(\rho_h\) reaches 1.
Some comments

Adaptive step size (step 1).
- $cESS(\mathcal{E}_{h-1}; p_{h-1}, p_h)$ can be computed for any $\rho_h$ before sampling.
- $\rho_h$ can be tuned to meet $\tau_1$: controls of the step size $\rho_h - \rho_{h-1}$

Propagation kernel $K_h$ (step 4).
- with stationary distribution $p_h$ (e.g. Gibbs sampler)
- only propagation: does not change the distribution $\rightarrow$ no convergence needed

Resampling (optional step 3).
- avoids degeneracy
- set weights $w_{hm} = 1$ after resampling
Theoretical justification \cite{DDJ06}

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

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

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

- $L_h = \text{backward kernel}$

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

- Update for the weights

$$
\omega_h(\theta_{0:h}) = \omega_{h-1}(\theta_{0:h-1}) \alpha(\theta_h)^{\rho_h - \rho_{h-1}}
$$
Marginal likelihood

Denote

\[ \gamma_h(\theta) = \tilde{\rho}_Y(\theta) \alpha(\theta)^{\rho_h}, \quad Z_h = \int \gamma_h(\theta) \, d\theta, \quad p_h = \gamma_h(\theta) / Z_h \]

The marginal likelihood is given by

\[ p(Y) = \int \pi(\theta) \ell(Y|\theta) \, d\theta = \int \gamma_H(\theta) \, d\theta = Z_H \]

which can be estimated without bias with

\[ \left( \frac{Z_H}{Z_0} \right) = \prod_{h=1}^{H} \left( \frac{Z_h}{Z_{h-1}} \right) \quad \text{where} \quad \left( \frac{Z_h}{Z_{h-1}} \right) = \sum_m W_h^m (\alpha_h^m)^{\rho_h - \rho_h - 1} \]
Outline

Motivating example

SBM-reg model

Variational Bayes inference

Bridge sampling

Some simulations

Back to the original problem

Discussion
Logistic regression: Sampling path

\[ \text{SMC: } (\Delta_{VB} = \text{diag}(\Sigma_{VB})) \]

- : \( \tilde{p}_Y = \tilde{p}_{VB} \)
- : \( \tilde{p}_Y = \tilde{p}_{ML} \)
- : variance \( \tilde{p}_Y = \Delta_{VB}/5 \)
- : variance \( \tilde{p}_Y = 10\Delta_{VB} \)
- : \( \tilde{p}_Y = \mathcal{N}(\mu_{VB} + .5, \Delta_{VB}/5) \)

[Nea01]:
- : \( \tilde{p}_Y = \pi \)

\[ \Delta = \text{hybrid} \]
SBM-reg model

Simulation design.

- $n = 20, 50$ nodes, $K^* = 1, 2$ classes, $d = 3$ covariates,
- $M = 1000$ particles, $B = 100$ samples.
- Parameters sampled from the prior.
SBM-reg model

Simulation design.
- $n = 20, 50$ nodes, $K^* = 1, 2$ classes, $d = 3$ covariates,
- $M = 1000$ particles, $B = 100$ samples.
- Parameters sampled from the prior.

Property check. $\theta^* \sim \pi$, $Y \sim \ell(Y|\theta^*)$ and $\{(\theta^m, w^m)\}$ a sample from $q(\theta)$:

$$q(\theta) = p(\theta|Y)$$

$$\Rightarrow \sum_m W_m \mathbb{I}\{\theta^m \leq \theta^*\} \sim \mathcal{U}[0, 1]$$
SBM-reg: $K^*$ known

Posterior distribution of the regression coefficients $\beta_\ell$
SBM-reg: $K^*$ known

Posterior distribution of the regression coefficients $\beta_\ell$

Empirical level of 95%-credibility intervals (CI):

VB: 84.75%, SMC: 93.75%
SBM-reg: Model selection

For each sample, compute

\[ p_{SMC}(K|Y) = \hat{Z}_H, \quad p_{VB}(K|Y) = \tilde{p}_Y(K) \]

\[ \hat{K}_{SMC} = \arg \max_K p_{SMC}(K|Y), \text{ idem } \hat{K}_{VB} \]
SBM-reg: Model selection

For each sample, compute

\[ p_{SMC}(K|Y) = \hat{Z}_H, \quad p_{VB}(K|Y) = \tilde{p}_Y(K) \]

\[ \hat{K}_{SMC} = \arg \max_K p_{SMC}(K|Y), \text{ idem } \hat{K}_{VB} \]

Results.

| n  | g* | \( \hat{K} = K^* \) | \text{mean } p(K^*|Y) |
|-----|----|----------------------|------------------------|
|     |    | VB       | SMC          | VB       | SMC          |
| 20  | 1  | 1.00     | 0.46         | 0.947    | 0.435        |
| 20  | 2  | 0.10     | 0.23         | 0.138    | 0.257        |
| 50  | 1  | 1.00     | 0.60         | 0.982    | 0.562        |
| 50  | 2  | 0.42     | 0.36         | 0.410    | 0.387        |

→ VB as least as good as SMC...
Some simulations

**SBM-reg: Model averaging**

Account for model uncertainty [HMRV99]: consider

\[
p(\theta|Y) = \sum_K p(K|Y) p(\theta|Y, K)
\]

\[
\Rightarrow \quad \nabla(\theta|Y) = \mathbb{E}_{K|Y} [\nabla(\theta|Y, K)] + \nabla_{K|Y} [\mathbb{E}(\theta|Y, K)]
\]

within models \hspace{2cm} between models
Some simulations

SBM-reg: Model averaging

Account for model uncertainty [HMRV99]: consider

\[ p(\theta|Y) = \sum_K p(K|Y)p(\theta|Y,K) \]

\[ \Rightarrow \nabla(\theta|Y) = \mathbb{E}_{K|Y}[\nabla(\theta|Y,K)] + \nabla_{K|Y}[\mathbb{E}(\theta|Y,K)] \]

within models

between models

Results.

Empirical level of 95%-CI:

VB: 85.8%

SMC: 93.25%
Outline

Motivating example

SBM-reg model

Variational Bayes inference

Bridge sampling

Some simulations

Back to the original problem

Discussion
Ecological network

<table>
<thead>
<tr>
<th></th>
<th>genet.</th>
<th>geo.</th>
<th>taxo.</th>
<th>genet.</th>
<th>geo.</th>
<th>taxo.</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>$4.6 \times 10^{-5}$</td>
<td>$2.3 \times 10^{-1}$</td>
<td>$-9.0 \times 10^{-1}$</td>
<td>$4.1 \times 10^{-5}$</td>
<td>$3.6 \times 10^{-1}$</td>
<td>$-9.1 \times 10^{-1}$</td>
</tr>
<tr>
<td>within var.</td>
<td>$2.2 \times 10^{-10}$</td>
<td>$4.3 \times 10^{-2}$</td>
<td>$1.7 \times 10^{-3}$</td>
<td>$1.1 \times 10^{-9}$</td>
<td>$2.2 \times 10^{-1}$</td>
<td>$8.9 \times 10^{-3}$</td>
</tr>
<tr>
<td>between var.</td>
<td>$5.6 \times 10^{-17}$</td>
<td>$1.2 \times 10^{-6}$</td>
<td>$2.4 \times 10^{-7}$</td>
<td>$4.0 \times 10^{-12}$</td>
<td>$1.9 \times 10^{-3}$</td>
<td>$2.8 \times 10^{-3}$</td>
</tr>
<tr>
<td>st. dev.</td>
<td>$1.5 \times 10^{-5}$</td>
<td>$2.1 \times 10^{-1}$</td>
<td>$4.2 \times 10^{-2}$</td>
<td>$3.3 \times 10^{-5}$</td>
<td>$4.7 \times 10^{-1}$</td>
<td>$1.1 \times 10^{-1}$</td>
</tr>
<tr>
<td>ratio</td>
<td>$3.1$</td>
<td>$1.1$</td>
<td>$-21$</td>
<td>$1.2$</td>
<td>$7.6 \times 10^{-1}$</td>
<td>$-8.4$</td>
</tr>
</tbody>
</table>

- Smaller posterior between-model variance with VB
- Smaller posterior variance with VB
- Can affect the conclusions in terms of significance
  → e.g. less obvious effect of the genetic distance
Outline

Motivating example

SBM-reg model

Variational Bayes inference

Bridge sampling

Some simulations

Back to the original problem

Discussion
Conclusion

Summary.

- Generic model to assess the contribution of covariates to the structure of a network
- Fast (but approximate) variational Bayes algorithm
- Bridge sampling post-processing to access the true posterior
Conclusion

Summary.

- Generic model to assess the contribution of covariates to the structure of a network
- Fast (but approximate) variational Bayes algorithm
- Bridge sampling post-processing to access the true posterior

Some limitations.

- Bridge sampling still computationally demanding for large networks
- How to transform node covariate into edge covariates ([HGH08])
References


Some typical graphon functions

'Scale free'

Community

Small world
Political blog network

$n = 196$ blogs ($N = 19110$ pairs), 3 covariates, density $= .075$

Inferred graphon (no covariate)  Residual graphon (3 covariates)

$\tilde{P}(H_0) \approx 10^{-172}$