Goodness of fit of logistic models for random graphs

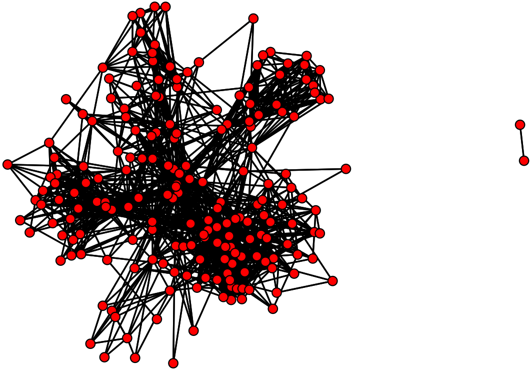
P. Latouche
joint work with S. Robin and S. Ouadah

Université Paris 1 Panthéon-Sorbonne
Laboratoire SAMM

CIRM, Bayesian Statistics and Algorithms, 2016

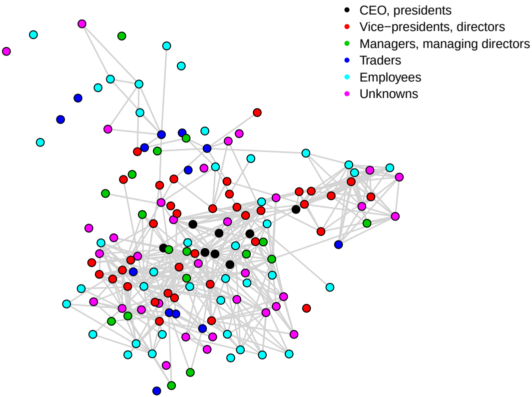
Contents

Outline



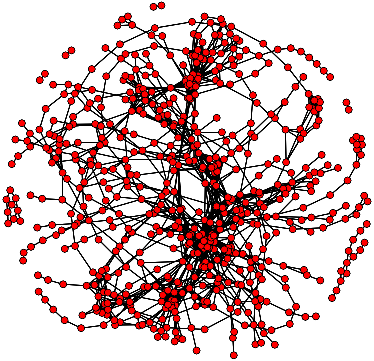
A subset of the French political blogosphere (2006)
(Observatoire Presidentielle project).

In social sciences



Electronic communications between 148 employees of the Enron company, before its bankruptcy in 2001.

In Biology



The metabolic network of bacteria *Escherichia coli* (Lacroix et al., 2006).

Models

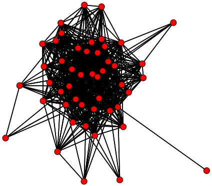
In computer science and physics

$\hookrightarrow$ Modularity, community structures, ...

Random graph models

Erdös-Rényi model (1959)
Latent position model (Hoff et al. 2002)
Latent position cluster model (Handcock et al. 2007)
Stochastic block model (SBM) (Nowicki and Snijders 2001)
Mixed membership SBM (Airoldi et al. 2008)
Overlapping SBM (Latouche et al. 2011)
W-graph model

Taking covariates into account



Interactions between 51 trees (Vacher et al., 2008).

List of covariates for each pair of nodes

- Genetic distance
- Geographic distance
- Taxonomic distance

Models

- Latent position model (Hoff et al. 2002)
- Latent position cluster model (Handcock et al. 2007)
- Extensions for SBM (Mariadassou et al., 2007; Zanghi et al. 2010)
- ...

In practice

Logistic models are often considered :

$$H_0 : \qquad Y_{ij} \sim \mathcal{B}\left[g(x_{ij}^{\intercal}\beta + \alpha)\right],$$

where $\beta \in \mathbb{R}^d$, $\alpha \in \mathbb{R}$ and $g(t) = 1/(1 + \exp(-t))$.

In practice

- Estimate the parameters (β, α) using standard tools for generalized linear models

Question

How to evaluate the fit of the (estimated) model ?

$\hookrightarrow$ rely on two models with *no* covariate

In practice

- Estimate the parameters (β, α) using standard tools for generalized linear models

Question

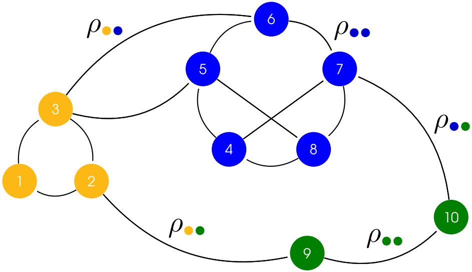
How to evaluate the fit of the (estimated) model ?

$\hookrightarrow$ rely on two models with *no* covariate

Stochastic Block Model (SBM)

- Nowicki and Snijders (2001)
  - Earlier work : Govaert et al. (1977)
- Z_i independent hidden variables :
  - $Z_i \sim \mathcal{M}\Big(1,\ \pi = (\pi_1, \pi_2, \dots, \pi_K)\Big)$
  - $Z_{ik} = 1$: vertex i belongs to class k
- $Y|Z$ edges drawn independently :

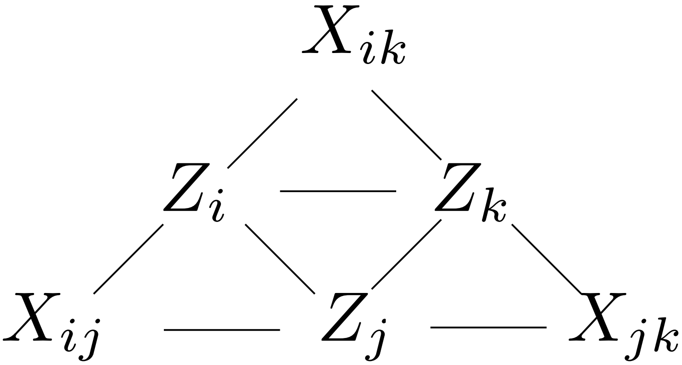
$$Y_{ij}|\{Z_{ik}Z_{jl} = 1\} \sim \mathcal{B}(\rho_{kl})$$



1
2
3
4
5
6
7
8
9
10
ρ
ρ
ρ
ρ
ρ

Issues

- Moral graph
- $p(Z|X, \pi, \rho)$



Moral graph.

Estimation

- VEM (Daudin et al. 2008)
- VBEM (Latouche et al. 2012)
- MCMC (Nowicki and Snijders, 2001)
- Model selection with : “BIC”, ICL, . . . (Daudin et al. 2008)
- With Chinese restaurant process (Kemp et al. 2006)
- With Indian buffet process (Miller et al. 2009)
- Collapsing (McDaid et al. 2012)
- Greedy search (Latouche et al. 2015)

W-graph model

W-graph model $\hookrightarrow$ *graphon* (Borgs et al. 2007)

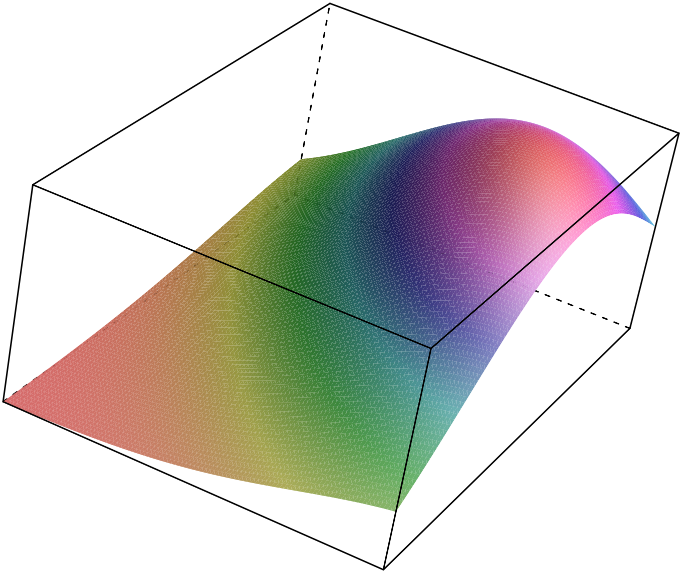
Graphon function

- $W : [0,1]^2 \rightarrow [0,1]$
- $W(u,v)$: probability that nodes (i,j) with coordinates u and v connect

Sampling

- Sample $U_i \sim \mathcal{U}(0,1),\ \forall i$
- Sample edges $Y_{ij}|U_i, U_j \sim \mathcal{B}\left(W(U_i, U_j)\right)$

W-graph model



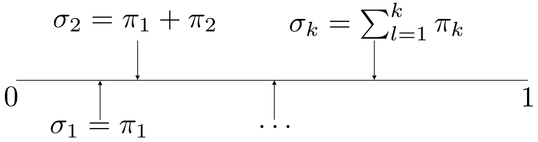
Example of a graphon function.

SBM and W-graph models

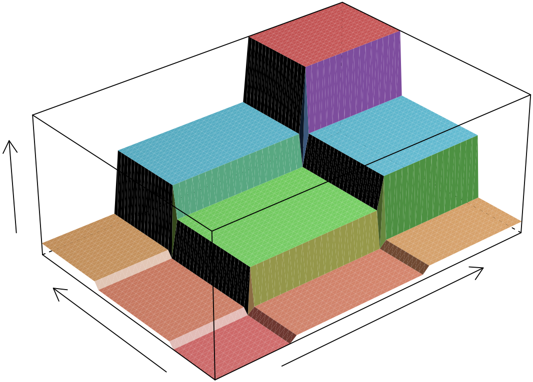
- SBM : special case of a W-graph model
- Recall : SBM : K + $\pi = (\pi_1, \ldots, \pi_K)$ + $\rho = \left(\rho_{kl}\right)_{kl}$

Connection

- Define $\sigma_k = \sum_{l=1}^{k} \pi_l, \ \forall k$
- $C_\pi(u) = 1 + \sum_{k=1}^{K} \mathbb{I}\{\sigma_k \leq u\} \iff u \in \text{class } k$
- $W(u,v) = \rho_{C_\pi(u), C_\pi(v)} \iff W(u,v) = \rho_{kl}$



SBM and W-graph models



Graphon function of a SBM model with $K = 3$ classes.

Estimation of W-graph models

- Not an easy task
- Some approaches : Kallenberg, 1999; Palla et al. 2010; Lloyd et al. 2012; Airoldi et al. 2013; Wolfe and Olhede 2013; Klopp et al. 2015; ...
- Latouche et Robin, 2015
- Alternative approach : Caron and Fox (2015)

Bayesian approach

Estimate the posterior distribution of $W(u, v)$

- $p(M_K|Y)$ (Latouche et al. 2009)
- $p(\pi|Y) \approx q(\pi) = \mathrm{Dir}(\pi; a)$
- $p(\rho_{kl}|Y) \approx q(\rho_{kl}) = Beta(\rho_{kl}; \eta_{kl}, \zeta_{kl})$ with VBEM (Latouche et al. 2012)
- Results of Mariadassou and Matias (2015)
- Simulation study of Gazal et al. (2011)

Estimator

$$W(u, v) = \rho_{C(u),C(v)}$$

Bayesian approach

Estimate the posterior distribution of $W(u, v)$

- $p(M_K|Y)$ (Latouche et al. 2009)
- $p(\pi|Y) \approx q(\pi) = \text{Dir}(\pi; a)$
- $p(\rho_{kl}|Y) \approx q(\rho_{kl}) = Beta(\rho_{kl}; \eta_{kl}, \zeta_{kl})$ with VBEM (Latouche et al. 2012)
- Results of Mariadassou and Matias (2015)
- Simulation study of Gazal et al. (2011)

Estimator

$$W(u, v) = \rho_{C(u),C(v)}$$

Weights

Random variables

- ρ_{kl}
- $C(u), C(v)$

Dirichlet

- $\pi \sim \text{Dir}(a) \implies$

$$\left(\sum_{j=1}^{q} \pi_j, \sum_{j=q+1}^{l} \pi_j, \sum_{j=l+1}^{K} \pi_j\right) \sim \text{Dir}\left(\sum_{j=1}^{q} a_j, \sum_{j=q+1}^{l} a_j, \sum_{j=l+1}^{K} a_j\right)$$

- $\tilde{\mathbb{P}}(C(u) = k, C(v) = l | X, K) = \tilde{\mathbb{P}}(\sigma_{k-1} < u < \sigma_k, \sigma_{l-1} < v < \sigma_l)$

Posterior distribution of W

Proposition

For given $(u, v) \in [0, 1]^2$, $u \leq v$, using a SBM with K groups, the variational Bayes approximate expectation of $W(u, v)$ is
$\tilde{\mathbb{E}}[W(u, v)|Y, M_K] =$

$$\sum_{k \leq \ell} \frac{\eta_{kl}}{\eta_{kl} + \zeta_{kl}} \Big[F_{k-1,l-1}(u, v; \mathbf{a}) - F_{k,l-1}(u, v; \mathbf{a}) - F_{k-1,l}(u, v; \mathbf{a}) + F_{k,l}(u, v; \mathbf{a}) \Big]$$

where

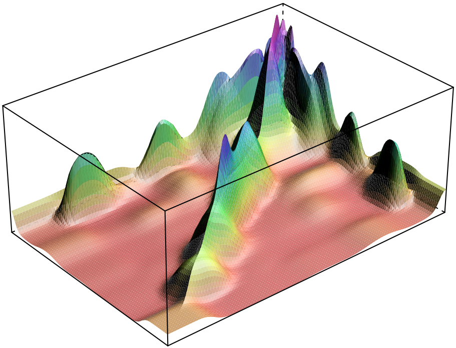
- $F_{k,l}(u, v; \mathbf{a})$ is the joint cdf of (σ_k, σ_l) when π has Dirichlet distribution $\mathrm{Dir}(\pi)$

Latouche and Robin, 2015

Model averaging

- $\log p(Y|M_K) \approx \mathcal{L}_K(q)$
- $\tilde{p}(M_K|Y) \propto \exp \mathcal{L}_K(q)$ with $\sum_K \tilde{p}(M_K|Y) = 1$
- $\tilde{p}(W(u,v)|Y) = \sum_K \tilde{p}(W(u,v)|Y,K)\tilde{p}(M_K|Y)$

Example



Estimation of the blogosphere Graphon function

Outline

Models

- H_0 : $\quad Y_{ij} \sim \mathcal{B}\left[g(x_{ij}^{\intercal}\beta + \alpha)\right]$
- H_1 : $\quad Y_{ij} \sim \mathcal{B}\left[g(x_{ij}^{\intercal}\beta + \phi(U_i, U_j))\right]$, where $U_i \sim \mathcal{U}(0,1)$ iid.
  Alternative model
- M_K : $\quad Y_{ij} \sim \mathcal{B}\left[g(x_{ij}^{\intercal}\beta + Z_i^{\intercal}\alpha Z_j)\right]$ where $Z_i \sim \mathcal{M}(1,\pi)$ iid.
  Blockwise constant alternative

Models

Note

- $H_0 = M_1$
- $H_1' = \bigcup_{K \geq 2} M_K$
- Set of model parameters : $\theta = (\beta, \pi, \alpha)$

- The comparison of H_0 and $H_1^{'}$ becomes a model selection problem

Bayesian mixture model comparison

- $p(H_0) = p(M_1)$ and $p(H_1') = \sum_{K \geq 2} p(M_K)$
- Posterior probability of Model M_K

$$p(M_K|Y) = \frac{p(Y|M_K)p(M_K)}{p(Y)} = \frac{p(Y|M_K)p(M_K)}{\sum_{K' \geq 1} p(Y|M_{K'})p(M_{K'})}$$

- Posterior probability of Model H_0 : $P(H_0|Y) = p(M_1|Y)$
- Bayes factor $B_{01} = \frac{p(Y|H_0)}{p(Y|H_1')}$ where $p(Y|H_1') = \sum_{K \geq 2} p(M_K)p(Y|M_K) \ / \ p(H_1')$

Inference

Problem

All goodness-of-fit criteria depend on marginal likelihood terms $p(Y|M_K)$ which have to be estimated

Outline

Bayesian framework

- Recall : $\theta = (\beta, \pi, \alpha)$
- We consider $p(\theta|M_K) = p(\beta|M_K)p(\pi|M_K)p(\alpha|M_K)$
- We give equal weights to H_0 and $H_1^{'}$: $p(H_0) = p(H_1') = 1/2$
- $p(M_1) = 1/2$
- We give equal prior probabilities to the models M_K $(K \geq 2)$ such that $p(H_1^{'}) = 1/2$

Bayesian framework

Prior distributions

- $p(\pi|M_K) = \mathrm{Dir}(\pi; e)$ where e is a vector with K components such that $e_k = e_0 > 0, \forall k \in \{1, \dots, K\}$
- $p(\beta|\eta, M_K) = \mathcal{N}(\beta;\ 0, \frac{I_d}{\eta}) = \prod_{j=1}^{d} \mathcal{N}(\beta_j;\ 0, \frac{1}{\eta})$, with I_d the $d \times d$ identity matrix and $\eta > 0$ a parameter controlling the inverse variance
- $p(\alpha|\gamma, M_K) = \prod_{k \leq l}^{K} \mathcal{N}(\alpha_{kl};\ 0, \frac{1}{\gamma})$
- $p(\gamma|M_K) = \mathrm{Gam}(\gamma;\ a_0, b_0), \quad a_0, b_0 > 0$
- $p(\eta|M_K) = \mathrm{Gam}(\eta;\ c_0, d_0), \quad c_0, d_0 > 0$

Note

The prior distributions $p(\beta|M_K)$ and $p(\alpha|M_K)$ are obtained by marginalizing over $p(\eta|M_K)$ and $p(\gamma|M_K)$. This results in prior distributions from the class of generalized hyperbolic distributions (Caron and Doucet, 2008)

Marginal likelihood

- $Z = (Z_i)_i$

$$\log p(Y|M_K) =$$

$$\log\left\{\sum_Z \int p(Y|Z,\alpha,\beta)p(Z|\pi)p(\alpha|\gamma)p(\beta|\eta)p(\pi)p(\gamma)p(\eta)\mathrm{d}\pi\mathrm{d}\alpha\mathrm{d}\beta\mathrm{d}\gamma\mathrm{d}\eta\right\}$$

- The marginal likelihood is not tractable
- Use variational approximations

VBEM

$$\log p(Y|M_K) = \mathcal{L}_K(q) + \text{KL}\left(q(\cdot)||p(\cdot|Y, M_K)\right),$$

where

$$\mathcal{L}_K(q) = \sum_Z \int q(Z,\pi,\alpha,\beta,\gamma,\eta) \log \frac{p(Y,Z,\pi,\alpha,\beta,\gamma,\eta)}{q(Z,\pi,\alpha,\beta,\gamma,\eta)} \mathrm{d}\pi \mathrm{d}\alpha \mathrm{d}\beta \mathrm{d}\gamma \mathrm{d}\eta.$$

- Assume $q(Z,\pi,\alpha,\beta,\gamma,\eta) = q(\pi)q(\alpha)q(\beta)q(\gamma)q(\eta)\prod_{i=1}^{n} q(Z_i)$
- Still not tractable !

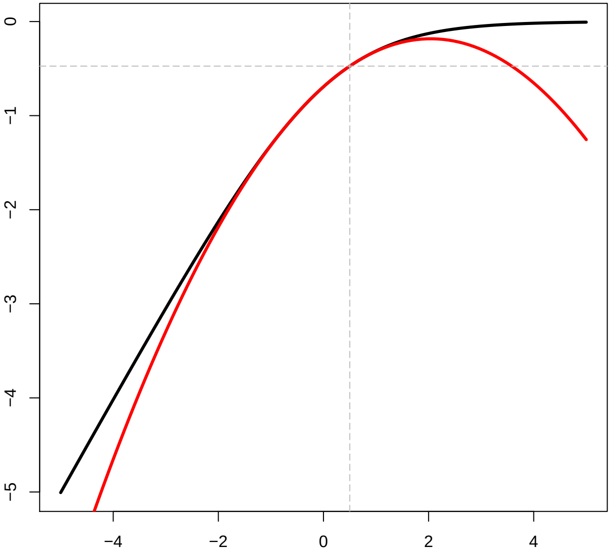
A bound

$$\log g(x) = -\log(1+e^{-x}) = \frac{x}{2} \underbrace{-\log(e^{\frac{x}{2}} + e^{-\frac{x}{2}})}_{\text{concave function in } x^2}$$

$$\geq \log g(\xi) + \frac{x-\xi}{2} - \lambda(\xi)(x^2 - \xi^2), \forall x, \xi \in \mathbb{R},$$

where $\lambda(\xi) = \frac{1}{4\xi}\tanh(\xi/2)$

A bound



$$\xi = 0.5$$

Local variational approximations

Proposition

Given any $n \times n$ positive real matrix $\xi = (\xi_{ij})_{1 \leq i,j \leq n}$, a lower bound of the first lower bound is given by

$$\log p(Y|M_K) \geq \mathcal{L}_K(q) \geq \mathcal{L}_K(q;\, \xi),$$

where

$$\mathcal{L}_K(q;\, \xi) =$$
$$\sum_Z \int q(Z,\pi,\alpha,\beta,\gamma,\eta) \log \frac{\sqrt{h(Z,\alpha,\beta,\xi)}\,p(Z,\pi,\alpha,\beta,\gamma,\eta)}{q(Z,\pi,\alpha,\beta,\gamma,\eta)} \mathrm{d}\pi \mathrm{d}\alpha \mathrm{d}\beta \mathrm{d}\gamma \mathrm{d}\eta,$$

and

Local variational approximations

Proposition (suite)

$$\log h(Z, \alpha, \beta, \xi) = \sum_{i \neq j}^{n} \Big\{ (Y_{ij} - \frac{1}{2})(Z_i^{\intercal} \alpha Z_j + x_{ij}^{\intercal} \beta) + \log g(\xi_{ij}) - \frac{\xi_{ij}}{2} - \lambda(\xi_{ij}) \left((Z_i^{\intercal} \alpha Z_j + x_{ij}^{\intercal} \beta)^2 - \xi_{ij}^2 \right) \Big\},$$

with $\xi_{ij} \in \mathbb{R}^+$, $\xi_{ij} = \xi_{ji}$. Moreover, $\lambda(\xi_{ij}) = (g(\xi_{ij}) - 1/2) / (2\xi_{ij})$

- Jaakkola and Jordan, 2000

Optimization

- Use VBEM to maximize $\mathcal{L}_K(q;\ \xi)$ with respect to the functional q, ξ being fixed
- Maximize $\mathcal{L}_K(q;\ \xi)$ with respect to ξ, q being fixed to obtain better local approximations

Optimization

Proposition

$$q(Z_i) = \mathcal{M}(Z_i;\ 1, \tau_i),$$

where $\sum_{k=1}^{K} \tau_{ik} = 1$ and

$$\begin{aligned}\tau_{ik} \propto \\ &\exp\Bigg\{ \sum_{l=1}^{K} (m_\alpha)_{kl} \sum_{j \neq i}^{n} \Big((Y_{ij} - \frac{1}{2}) - 2\lambda(\xi_{ij}) x_{ij}^\intercal m_\beta \Big) \tau_{jl} \\ &\quad - \sum_{l=1}^{K} \mathbb{E}_{\alpha_{kl}}[\alpha_{kl}^2] \sum_{j \neq i}^{n} \lambda(\xi_{ij}) \tau_{jl} + \psi(e_k^n) - \psi\Big(\sum_{l=1}^{K} e_l^n \Big) \Bigg\}.\end{aligned}$$

Proposition

$$q(\pi) = \text{Dir}(\pi;\ e^n),$$

where, $\forall k \in \{1, \ldots, K\}$, $e^n_k = e_0 + \sum_{i=1}^{n} \tau_{ik}$

Proposition

$$q(\beta) = \mathcal{N}(\beta;\ m_\beta, S_\beta),$$

where

$$S_\beta^{-1} = \frac{c_n}{d_n} I_d + \sum_{i \neq j}^{n} \lambda(\xi_{ij}) x_{ij} x_{ij}^\intercal,$$

and

$$m_\beta = S_\beta \frac{1}{2} \sum_{i \neq j}^{n} \left(Y_{ij} - \frac{1}{2} - 2\lambda(\xi_{ij}) \tau_i^\intercal m_\alpha \tau_j \right) x_{ij}.$$

Proposition

$$q(\gamma) = \mathrm{Gam}(\gamma;\ a_n, b_n),$$

where $a_n = a_0 + \frac{K(K+1)}{4}$ and $b_n = b_0 + \frac{1}{2}\sum_{k\leq l}^{K} \mathbb{E}_{\alpha_{kl}}[\alpha_{kl}^2]$.

Proposition

$$q(\eta) = \mathrm{Gam}(\eta;\ c_n, d_n),$$

where $c_n = c_0 + \frac{d}{2}$ and $d_n = d_0 + \frac{1}{2}\mathrm{Tr}(S_\beta) + \frac{1}{2}m_\beta^\intercal m_\beta$, S_β

Optimization

Proposition

$$q(\alpha) = \prod_{k \neq l}^{K} \mathcal{N}\left(\alpha_{kl}; (m_\alpha)_{kl}, (\sigma_\alpha^2)_{kl}\right),$$

where

$$(\sigma_\alpha^2)_{kk}^{-1} = \frac{a_n}{b_n} + \sum_{i \neq j}^{n} \lambda(\xi_{ij}) \tau_{ik} \tau_{jk}, \forall k,$$

$$(\sigma_\alpha^2)_{kl}^{-1} = \frac{a_n}{b_n} + 2 \sum_{i \neq j}^{n} \lambda(\xi_{ij}) \tau_{ik} \tau_{jl}, \forall k \neq l,$$

$$(m_\alpha)_{kk} = (\sigma_\alpha^2)_{kk} \sum_{i \neq j}^{n} \left(\frac{1}{2}(Y_{ij} - \frac{1}{2}) - \lambda(\xi_{ij}) x_{ij}^\intercal m_\beta \right) \tau_{ik} \tau_{jk}, \forall k,$$

$$(m_\alpha)_{kl} = (\sigma_\alpha^2)_{kl} \sum_{i \neq j}^{n} \left((Y_{ij} - \frac{1}{2}) - 2\lambda(\xi_{ij}) x_{ij}^\intercal m_\beta \right) \tau_{ik} \tau_{jl}, \forall k \neq l.$$

Proposition

The estimate $\widehat{\xi}_{ij}$ of ξ_{ij} maximizing $\mathcal{L}_K(q;\, \xi)$ is given by

$$\xi_{ij} = \sqrt{\sum_{k,l}^{K} \tau_{ik}\tau_{jl}\mathbb{E}_{\alpha_{kl}}[\alpha_{kl}^2] + 2\sum_{k,l}^{K} \tau_{ik}\tau_{jl}(m_\alpha)_{kl} x_{ij}^\intercal m_\beta + \mathrm{Tr}(x_{ij}x_{ij}^\intercal(S_\beta + m_\beta m_\beta^\intercal))}.$$

Note that $\widehat{\xi}_{ij} = \widehat{\xi}_{ji}, \forall i \neq j$ since the networks considered are undirected.

Trick

- Approximate $\hat{p}(Y|M_K) = \exp\{\mathcal{L}(q;\ \xi)\}$
- Approximate posterior probability : $\hat{p}(M_K|Y)$
- Approximate posterior probability of Model H_0 : $\hat{P}(H_0|Y) = \hat{P}(M_1|Y)$
- Approximate Bayes factor : $\hat{B}_{01}$

Residual structures

Proposition

$$\widehat{\mathbb{E}}\big[\phi(u,v)|Y\big] = \sum_{K\geq 1} \widehat{p}(M_K|Y)\widehat{\mathbb{E}}\big[\phi(u,v)|Y, M_K\big],$$

where

$$\widehat{\mathbb{E}}\big[\phi(u,v)|Y, M_K\big] =$$
$$\sum_{k\leq l}(m_\alpha)_{kl}\left[F_{k-1,l-1}(u,v;e) - F_{k,l-1}(u,v;e) - F_{k-1,l}(u,v;e) + F_{k,l}(u,v;e)\right]$$

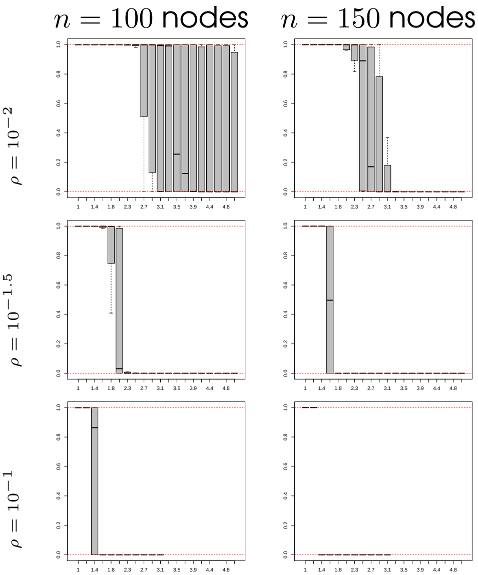
$F_{k,l}(u,v;e)$ denotes the joint cdf of the Dirichlet variables (σ_k, σ_l) such that $\sigma_k = \sum_{l=1}^{k} \pi_l$ and π has a Dirichlet distribution Dir(e).

Outline

Simulation study

- Simulate networks using Model H_1
- $\phi(u, v) = \rho\lambda^2(uv)^{\lambda-1}$, $\lambda \in [1, 9]$
- $x_i \in \mathbb{R}^d$ is drawn for each node, using a standardized Gaussian distribution with $d = 2$
- $x_{ij} = x_i - x_j$
- $\beta = (1, 1)^\intercal$

Results

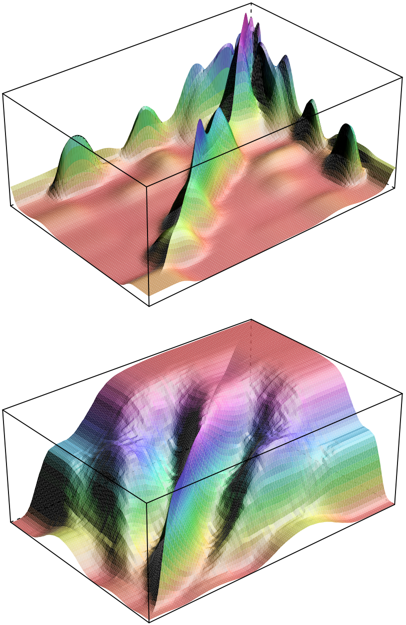

$n = 100$ nodes
$n = 150$ nodes
$\rho = 10^{-2}$
$\rho = 10^{-1.5}$
$\rho = 10^{-1}$
0.0
0.2
0.4
0.6
0.8
1.0
1
1.4
1.8
2.3
2.7
3.1
3.5
3.9
4.4
4.8

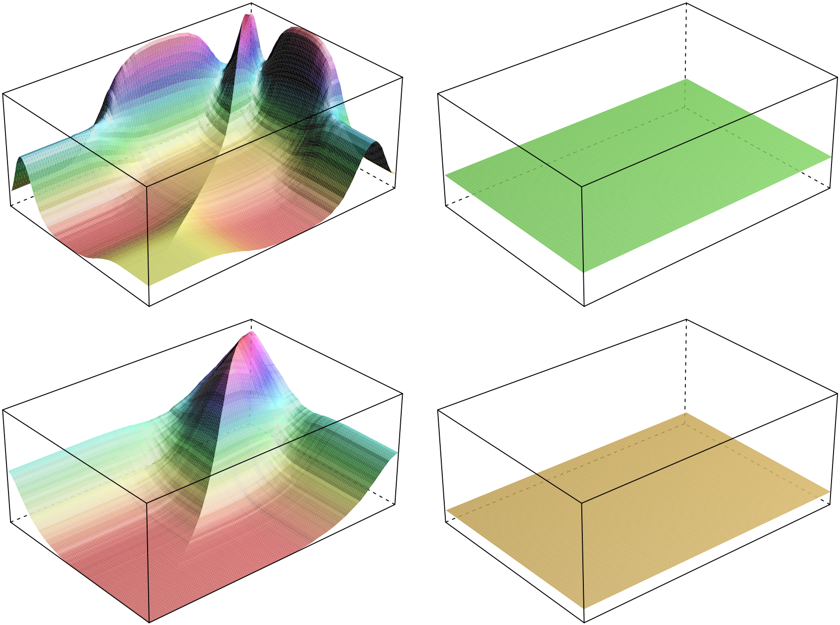
Real data

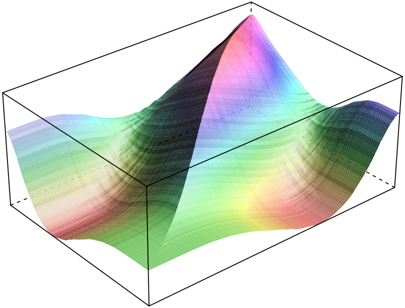
- Blog network
- Tree network
- Karate network
- Florentine marriage network
- Florentine business network
- Faux Dixon high network

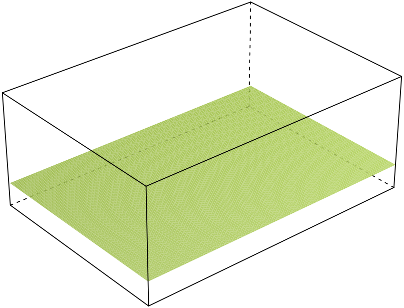
network	size (n)	nb. covariates (d)	density	$\hat{p}(H_0\|Y)$
Blog	196	3	0.07	3.6e-172
Tree	51	3	0.54	7.5e-153
Karate	34	10	0.14	0.998
Florentine (marriage)	16	3	0.17	0.995
Florentine (business)	16	3	0.12	0.991
Faux dixon high	248	3	0.02	1

Table: Estimation of $p(H_0|Y)$, for the six networks considered.









Conclusion

- How to assess the fit of a logistic model for network ?
- Consider a blockwise constant alternative for testing
- Bayesian model comparison / averaging context
- Estimate the residual structure

References

- K. Nowicki and T.A.B. Snijders (2001), Estimation and prediction for stochastic blockstructures. Journal of the American Statistical Association, 2001
- P. Latouche and S. Robin, Variational Bayes model averaging for graphon functions and motif frequencies inference in W-graph model, 2015
- T.S. Jaakkola and M.I. Jordan, Bayesian parameter estimation via variational methods. Statistics and Computing, 2000.
- L. Lovasz and B. Szegedy, Limits of dense graph sequences. Journal of Combinatorial Theory, Series B, 2006