Goodness of fit of logistic models for random graphs: a variational Bayes approach

S. Robin

Joint work with P. Latouche and S. Ouadah

INRA / AgroParisTech

MIAT Seminar, Toulouse, November 2016

Motivating example

Data: $n = 51$ tree species,
Y_{ij} = number of common parasites [23].

Motivating example

Data: $n = 51$ tree species, Y_{ij} = number of common parasites [23].

Valued stochastic block model.
If $i \in \mathcal{C}_k$, $j \in \mathcal{C}_\ell$:

$$Y_{ij} \sim \mathcal{P}(\lambda_{k\ell}),$$

$\lambda_{k\ell}$ = mean number of shared parasites.

Pseudo-BIC: $K = 7$ groups

$\widehat{\lambda}_{k\ell}$	T1	T2	T3	T4	T5	T6	T7
T1	14.46	4.19	5.99	7.67	2.44	0.13	1.43
T2		14.13	0.68	2.79	4.84	0.53	1.54
T3			3.19	4.10	0.66	0.02	0.69
T4				7.42	2.57	0.04	1.05
T5					3.64	0.23	0.83
T6						0.04	0.06
T7							0.27
$\widehat{\pi}_k$	7.8	7.8	13.7	13.7	15.7	19.6	21.6

Motivating example

Data: $n = 51$ tree species,
Y_{ij} = number of common parasites [23].

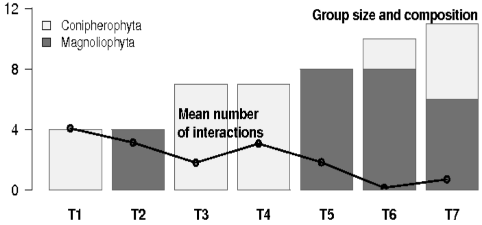
Valued stochastic block model.
If $i \in \mathcal{C}_k$, $j \in \mathcal{C}_\ell$:

$$Y_{ij} \sim \mathcal{P}(\lambda_{k\ell}),$$

$\lambda_{k\ell}$ = mean number of shared parasites.

Pseudo-BIC: $K = 7$ groups

$\widehat{\lambda}_{k\ell}$	T1	T2	T3	T4	T5	T6	T7
T1	14.46	4.19	5.99	7.67	2.44	0.13	1.43
T2		14.13	0.68	2.79	4.84	0.53	1.54
T3			3.19	4.10	0.66	0.02	0.69
T4				7.42	2.57	0.04	1.05
T5					3.64	0.23	0.83
T6						0.04	0.06
T7							0.27
$\widehat{\pi}_k$	7.8	7.8	13.7	13.7	15.7	19.6	21.6



Accounting for the taxonomic distance d_{ij}

Model: [19]

$$Y_{ij} \sim \mathcal{P}[\lambda_{k\ell} \; e^{\beta d_{ij}}].$$

Accounting for the taxonomic distance d_{ij}

$\widehat{\lambda}_{k\ell}$	T'1	T'2	T'3	T'4
T'1	0.75	2.46	0.40	3.77
T'2		4.30	0.52	8.77
T'3			0.080	1.05
T'4				14.22
$\widehat{\pi}_k$	17.7	21.5	23.5	37.3
$\widehat{\beta}$	-0.317			

Model: [19]

$$Y_{ij} \sim \mathcal{P}[\lambda_{k\ell}\ e^{\beta d_{ij}}].$$

Results: $\widehat{\beta} = -0.317$.

Pseudo-BIC: $K = 4$ groups.

Accounting for the taxonomic distance d_{ij}

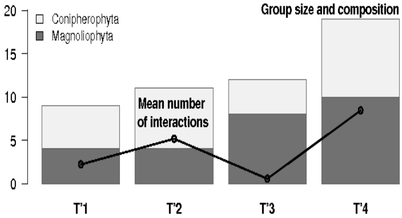
$\widehat{\lambda}_{k\ell}$	T'1	T'2	T'3	T'4
T'1	0.75	2.46	0.40	3.77
T'2		4.30	0.52	8.77
T'3			0.080	1.05
T'4				14.22
$\widehat{\pi}_k$	17.7	21.5	23.5	37.3
$\widehat{\beta}$	-0.317			

Model: [19]

$$Y_{ij} \sim \mathcal{P}[\lambda_{k\ell} \; e^{\beta d_{ij}}].$$

Results: $\widehat{\beta} = -0.317$.

Pseudo-BIC: $K = 4$ groups.



Problem

Ecological network $Y = (Y_{ij})_{1 \leq i,j \leq n}$: interactions between n tree species:

$$Y_{ij} = \mathbb{I}\{i \sim j\} = \mathbb{I}\{i \text{ and } j \text{ share some fungal parasite}\}$$

→ in this talk, binary networks only.

Problem

Ecological network $Y = (Y_{ij})_{1 \leq i,j \leq n}$: interactions between n tree species:

$$Y_{ij} = \mathbb{I}\{i \sim j\} = \mathbb{I}\{i \text{ and } j \text{ share some fungal parasite}\}$$

→ in this talk, binary networks only.

$+$ Edge covariates $x = (x_{ij})_{1 \leq i,j \leq n}$:

- $x_{ij}^1 =$ phylogenetic distance between i and j,
- $x_{ij}^2 =$ geographic distance between the typical habitat of i and j,
- ...

Problem

Ecological network $Y = (Y_{ij})_{1 \leq i,j \leq n}$: interactions between n tree species:

$$Y_{ij} = \mathbb{I}\{i \sim j\} = \mathbb{I}\{i \text{ and } j \text{ share some fungal parasite}\}$$

→ in this talk, binary networks only.

$+$ Edge covariates $x = (x_{ij})_{1 \leq i,j \leq n}$:

- $x_{ij}^1 =$ phylogenetic distance between i and j,
- $x_{ij}^2 =$ geographic distance between the typical habitat of i and j,
- ...

Questions:

- Can we explain the topology of Y based on x?
- Is there a residual heterogeneity in the network Y?

Problem

Ecological network $Y = (Y_{ij})_{1 \leq i,j \leq n}$: interactions between n tree species:

$$Y_{ij} = \mathbb{I}\{i \sim j\} = \mathbb{I}\{i \text{ and } j \text{ share some fungal parasite}\}$$

$\rightarrow$ in this talk, binary networks only.

$+$ Edge covariates $x = (x_{ij})_{1 \leq i,j \leq n}$:

- $x_{ij}^1 =$ phylogenetic distance between i and j,
- $x_{ij}^2 =$ geographic distance between the typical habitat of i and j,
- ...

Questions:

- Can we explain the topology of Y based on x? $\rightarrow$ logistic regression
- Is there a residual heterogeneity in the network Y? $\rightarrow$ W-graph

Outline

Reminder of variational Bayes inference

Logistic regression

W-graphs and graphon

Goodness-of-fit

Illustrations

Outline

Reminder of variational Bayes inference

Logistic regression

W-graphs and graphon

Goodness-of-fit

Illustrations

Variational Bayes inference: General principle

$Y =$ observed data, $Z =$ latent variable, $\theta =$ parameter.

Frequentist or Bayesian inference requires

$$p(Z|Y), \qquad p(\theta|Y), \qquad p(\theta, Z|Y).$$

Variational Bayes inference: General principle

Y = observed data, Z = latent variable, θ = parameter.

Frequentist or Bayesian inference requires

$$p(Z|Y), \qquad p(\theta|Y), \qquad p(\theta, Z|Y).$$

Variational inference:

$$\boxed{\text{find} \quad \widetilde{p}(\cdot) \approx p(\cdot|Y)}$$

Variational Bayes inference: General principle

Y = observed data, Z = latent variable, θ = parameter.

Frequentist or Bayesian inference requires

$$p(Z|Y), \qquad p(\theta|Y), \qquad p(\theta, Z|Y).$$

Variational inference:

$$\boxed{\text{find} \quad \widetilde{p}(\cdot) \approx p(\cdot|Y)}$$

Typically [11,21,25]

$$\widetilde{p} = \arg\min_{q \in \mathcal{Q}} D[q(\cdot) \parallel p(\cdot|Y)]$$

- $D[q \parallel p] = KL[q \parallel p]$;
- $\widetilde{p}(\theta) = \mathcal{N}$;
- $\widetilde{p}(Z) = \prod_i q_i(Z_i)$;
- $\widetilde{p}(\theta, Z) = \widetilde{p}(\theta)\widetilde{p}(Z)$.

Variational Bayes inference: Some tools

Latent variable models: mean field approximation

$$\widetilde{p}(\theta, Z) := \widetilde{p}(\theta) \times \widetilde{p}(Z),$$

The VBEM algorithm [1] achieves

$$\min_{\widetilde{p}} \; KL[\widetilde{p}(\theta, Z) \parallel p(\theta, Z|Y)] \quad \Leftrightarrow \quad \max_{\widetilde{p}} \; \log p(Y) - KL[\widetilde{p}(\theta, Z) \parallel p(\theta, Z|Y)]$$

Variational Bayes inference: Some tools

Latent variable models: mean field approximation

$$\widetilde{p}(\theta, Z) := \widetilde{p}(\theta) \times \widetilde{p}(Z),$$

The VBEM algorithm [1] achieves

$$\min_{\widetilde{p}} KL[\widetilde{p}(\theta, Z) \parallel p(\theta, Z|Y)] \quad \Leftrightarrow \quad \max_{\widetilde{p}} \log p(Y) - KL[\widetilde{p}(\theta, Z) \parallel p(\theta, Z|Y)]$$

Bayesian model averaging (BMA) [8]: Consider models $M_1, \dots, M_K, \dots$

$$\mathbb{E}[h(\theta)|Y] = \sum_K p(K|Y) \, \mathbb{E}[h(\theta)|Y, K]$$

Variational Bayes inference: Some tools

Latent variable models: mean field approximation

$$\widetilde{p}(\theta, Z) := \widetilde{p}(\theta) \times \widetilde{p}(Z),$$

The VBEM algorithm [1] achieves

$$\min_{\widetilde{p}} KL[\widetilde{p}(\theta, Z) \parallel p(\theta, Z|Y)] \quad \Leftrightarrow \quad \max_{\widetilde{p}} \log p(Y) - KL[\widetilde{p}(\theta, Z) \parallel p(\theta, Z|Y)]$$

Bayesian model averaging (BMA) [8]: Consider models $M_1, \dots, M_K, \dots$

$$\mathbb{E}[h(\theta)|Y] = \sum_K p(K|Y)\ \mathbb{E}[h(\theta)|Y, K]$$

Variational approximation of $p(K|Y)$ [24]:

$$\widetilde{p}(K) \quad \propto \quad p(K|Y)\ e^{-KL_K^*} \quad \propto \quad p(K)\ e^{\log p(Y|K) - KL_K^*}$$

where KL_K^* = minimized KL divergence for model M_K.

Outline

Reminder of variational Bayes inference

Logistic regression

W-graphs and graphon

Goodness-of-fit

Illustrations

Logistic regression

Forget about the network structure: (Y_{ij}) are independent

$$Y_{ij} \sim \mathcal{B}[g(x_{ij}^\intercal \beta)], \qquad g(u) = 1/(1+e^{-u})$$

Logistic regression

Forget about the network structure: (Y_{ij}) are independent

$$Y_{ij} \sim \mathcal{B}[g(x_{ij}^\intercal \beta)], \qquad g(u) = 1/(1 + e^{-u})$$

Bayesian inference. Prior on β:

$$\beta \sim \mathcal{N}(m, G^{-1})$$

Posterior distribution:

$$p(\beta|Y) \propto p(\beta)p(Y|\beta)$$

- No nice conjugacy holds for the normal prior / logistic regression model.
- No close form for the posterior $p(\beta|Y)$.

Variational Bayes inference

Nasty term in $\log p(Y|\beta)$:

$$-\log(1+e^{-x_{ij}\beta}).$$

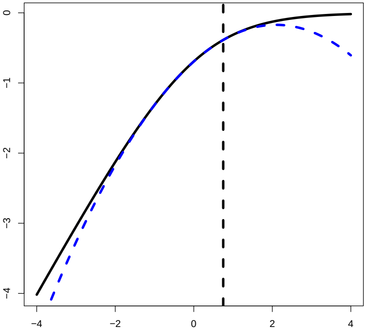
Variational Bayes inference

Nasty term in $\log p(Y|\beta)$:

$$-\log(1 + e^{-x_{ij}\beta}).$$

[12]: quadratic lower bound of $\log p(\beta, Y)$ to get

$$p(\beta|Y) \approx \widetilde{p}(\beta) = \mathcal{N}(\beta; \widetilde{m}, \widetilde{G}^{-1}).$$



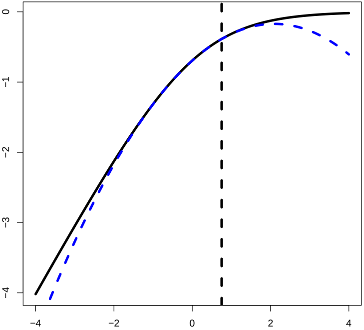
Variational Bayes inference

Nasty term in $\log p(Y|\beta)$:

$$-\log(1 + e^{-x_{ij}\beta}).$$

[12]: quadratic lower bound of $\log p(\beta, Y)$ to get

$$p(\beta|Y) \approx \widetilde{p}(\beta) = \mathcal{N}(\beta; \widetilde{m}, \widetilde{G}^{-1}).$$



Example: Tree network. $n = 51$ tree species, ($N = 1275$ pairs), 3 distances:

	intercept	genetic	geographic	taxonomic
μ_β	0.221	0.060	−0.337	−0.811
σ_β	0.058	0.062	0.059	0.060

Outline

Reminder of variational Bayes inference

Logistic regression

W-graphs and graphon

Goodness-of-fit

Illustrations

Latent space models

Generic framework [2]:

- Latent variable for node i: $Z_i \sim \pi$.
- Edges Y_{ij} conditionally independent:

$$(Z_i) \text{ iid } \sim \pi, \qquad Y_{ij} \mid (Z_i) : Y_{ij} \sim \mathcal{B}[\gamma(Z_i, Z_j)]$$

Latent space models

Generic framework [2]:

- Latent variable for node *i*: $Z_i \sim \pi$.
- Edges Y_{ij} conditionally independent:

$$(Z_i) \text{ iid } \sim \pi, \qquad Y_{ij} \mid (Z_i) : Y_{ij} \sim \mathcal{B}[\gamma(Z_i, Z_j)]$$

Includes [20]:

- stochastic block-model [5],
- latent position models [9,7,3],
- *W*-graph [18,4]

→ 'limit' model of most of these models.

Latent space models

Generic framework [2]:

- Latent variable for node i: $Z_i \sim \pi$.
- Edges Y_{ij} conditionally independent:

$$(Z_i) \text{ iid } \sim \pi, \qquad Y_{ij} \mid (Z_i) : Y_{ij} \sim \mathcal{B}[\gamma(Z_i, Z_j)]$$

Includes [20]:

- stochastic block-model [5],
- latent position models [9,7,3],
- W-graph [18,4]

→ 'limit' model of most of these models.

Intricate dependency structure: sampling and/or approximation is required for the inference of $p(Z, \theta | Y)$

W-graph (graphon) model

Latent variables:

$$(U_i) \text{ iid } \sim \mathcal{U}_{[0,1]},$$

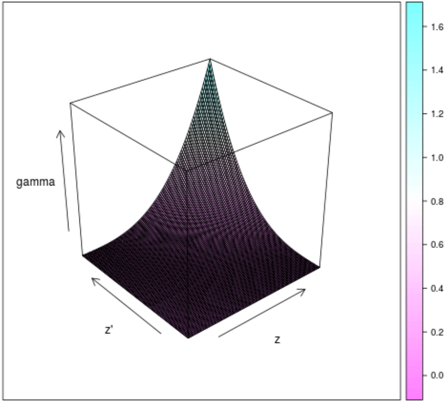
Graphon function γ:

$$\gamma(u, v) : [0, 1]^2 \mapsto [0, 1]$$

Edges:

$$P(Y_{ij} = 1 | U_i, U_j) = \gamma(U_i, U_j)$$

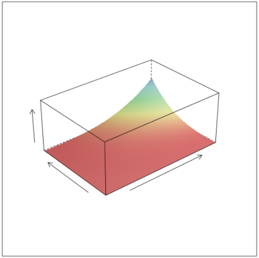
Graphon function $\gamma(u, v)$



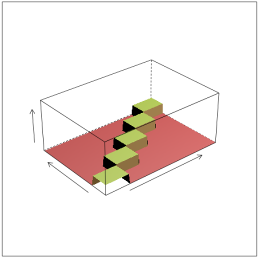
Some typical graphon functions

Graphon function γ: a global picture of the network's topology.

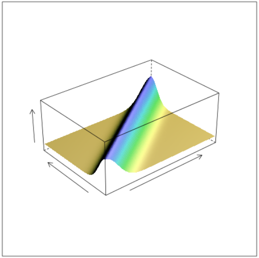
'Scale free'



Community

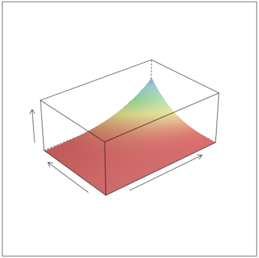


Small world

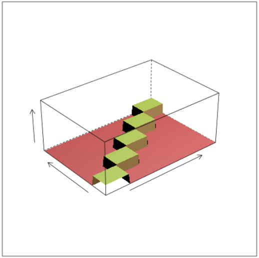


Some typical graphon functions

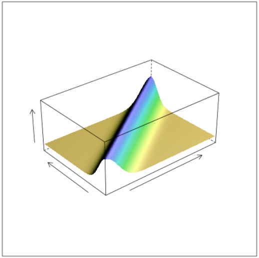
Graphon function γ: a global picture of the network's topology.



'Scale free'



Community



Small world

Remark: $\gamma(\cdot,\cdot) =$cst $\longrightarrow$ ER model

Variational Bayes inference of the graphon (1/3)

Stochastic Block-model (SBM) is a W-graph model.

Latent variables:

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

Blockwise constant graphon:

$$\gamma(z, z') = \gamma_{k\ell}$$

Edges:

$$P(Y_{ij} = 1 | Z_i, Z_j) \quad = \quad \gamma(Z_i, Z_j)$$

Variational Bayes inference of the graphon (1/3)

Stochastic Block-model (SBM) is a *W*-graph model.

Latent variables:

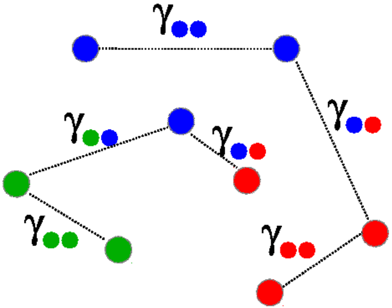
$$(Z_i) \text{ iid } \sim \mathcal{M}(1, \pi)$$

Blockwise constant graphon:

$$\gamma(z, z') = \gamma_{k\ell}$$

Edges:

$$P(Y_{ij} = 1 | Z_i, Z_j) \quad = \quad \gamma(Z_i, Z_j)$$



Variational Bayes inference of the graphon (1/3)

Stochastic Block-model (SBM) is a *W*-graph model.

Latent variables:

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

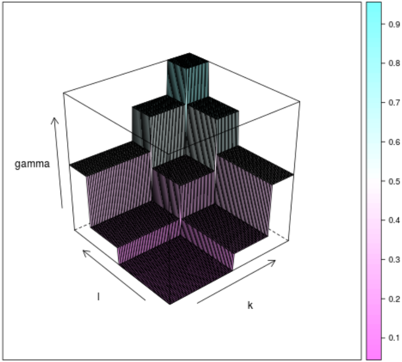
Blockwise constant graphon:

$$\gamma(z, z') = \gamma_{k\ell}$$

Edges:

$$P(Y_{ij} = 1 | Z_i, Z_j) \quad = \quad \gamma(Z_i, Z_j)$$

Graphon function of SBM_K



→ block widths $= \pi_k$, block heights $\gamma_{k\ell}$

Variational Bayes inference of the graphon (2/3)

VBEM for SBM: conjugate prior-exponential family setting [13].

Variational Bayes inference of the graphon (2/3)

VBEM for SBM: conjugate prior-exponential family setting [13].

Approximate posteriors

$$\begin{array}{rcl}\widetilde{p}(\pi) & = & \mathcal{D}(\widetilde{\pi}) \\ \widetilde{p}(\gamma_{k\ell}) & = & \mathrm{B}(\widetilde{\gamma}^0_{k\ell}, \widetilde{\gamma}^1_{k\ell}) \\ \widetilde{p}(Z_i) & = & \mathcal{M}(1; \widetilde{\tau}_i)\end{array}$$

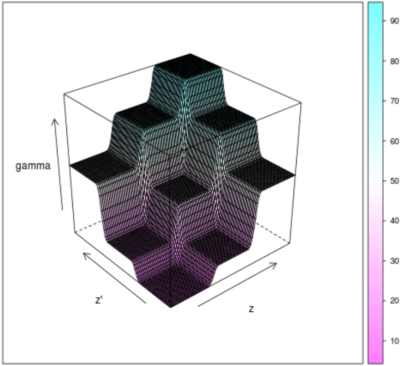
Variational Bayes inference of the graphon (2/3)

VBEM for SBM: conjugate prior-exponential family setting [13].

Approximate posteriors

$$\begin{array}{rcl}\widetilde{p}(\pi) & = & \mathcal{D}(\widetilde{\pi}) \\ \widetilde{p}(\gamma_{k\ell}) & = & \mathrm{B}(\widetilde{\gamma}^0_{k\ell}, \widetilde{\gamma}^1_{k\ell}) \\ \widetilde{p}(Z_i) & = & \mathcal{M}(1; \widetilde{\tau}_i)\end{array}$$

Posterior mean $\widetilde{\mathbb{E}}^{SBM}_K[\gamma(u, v)]$



Variational Bayes inference of the graphon (3/3)

Variational Bayes model averaging.

- Each auxiliary model SBM_K provides an estimates of γ as:

$$\widehat{\gamma}_K(u, v) = \widetilde{\mathbb{E}}_K^{SBM}[\gamma(u, v)].$$

- VB inference also provides an approximation of the posterior probability of each auxiliary model SBM_K:

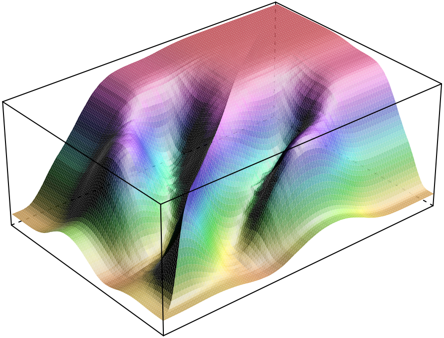
$$\widetilde{p}(K) \approx p(K|Y)$$

- A averaged estimate of γ is given by [15]:

$$\widetilde{\mathbb{E}}[\gamma(u, v)] = \sum_K \widetilde{p}(K) \; \widetilde{\mathbb{E}}_K^{SBM}[\gamma(u, v)].$$

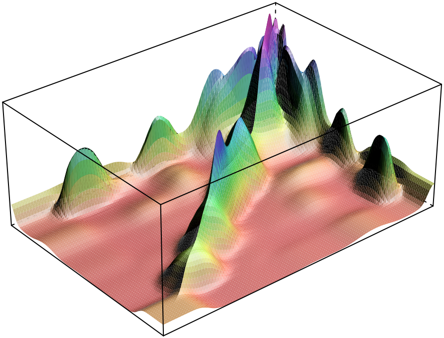
Example: Tree network

$n = 51$ tree species. Inferred graphon function:



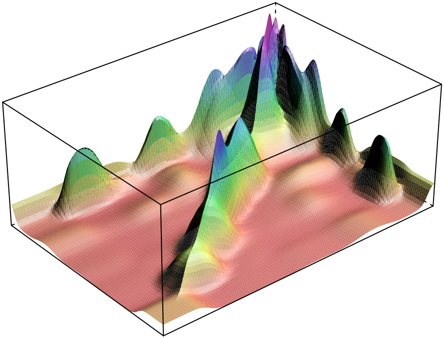
Example: Political blog network

$n = 196$ blogs. Inferred graphon function:



Example: Political blog network

$n = 196$ blogs. Inferred graphon function:



- Graphical representation of the network.
- Interpretability...

Outline

Reminder of variational Bayes inference

Logistic regression

W-graphs and graphon

Goodness-of-fit

Illustrations

Back to the original problem

Data:

- Y = observed (binary) network
- x = covariates

Questions:

- Does x explain the topology of Y?
- Residual (wrt x) heterogeneity in Y?

Combined model

Logistic regression:

$$\text{logit } P(Y_{ij} = 1) = \beta_0 + \sum_k \beta_k x_{ij}^k$$

Combined model

Logistic regression:

$$\text{logit } P(Y_{ij} = 1) = \beta_0 + \sum_k \beta_k x_{ij}^k$$

Logistic regression + graphon residual term: (U_i) iid $\sim \mathcal{U}[0, 1]$,

$$\text{logit } P(Y_{ij} = 1 | U_i, U_j) = \phi(U_i, U_j) + \sum_k \beta_k x_{ij}^k$$

Combined model

Logistic regression:

$$\text{logit } P(Y_{ij} = 1) = \beta_0 + \sum_k \beta_k x_{ij}^k$$

Logistic regression + graphon residual term: (U_i) iid $\sim \mathcal{U}[0, 1]$,

$$\text{logit } P(Y_{ij} = 1 | U_i, U_j) = \phi(U_i, U_j) + \sum_k \beta_k x_{ij}^k$$

Goodness of fit [16]: Check if

$$\phi(u, v) = \text{cst} \qquad (= \beta_0)$$

Goodness-of-fit as model comparison

Auxiliary model M_K:

$$\text{logit } P(Y_{ij} = 1 | U_i, U_j, K) = \phi_K^{SBM}(U_i, U_j) + \sum_k \beta_k x_{ij}^k.$$

Goodness-of-fit as model comparison

Auxiliary model M_K:

$$\text{logit } P(Y_{ij} = 1 | U_i, U_j, K) = \phi_K^{SBM}(U_i, U_j) + \sum_k \beta_k x_{ij}^k.$$

Goodness of fit:

$$\begin{array}{lcll} H_0 & = & \{\text{logistic regression is sufficient}\} & = & M_1 \\ H_1 & = & \{\text{logistic regression is not sufficient}\} & = & \bigcup_{K>1} M_K \end{array}$$

GOF is a assessed if

$$p(H_0|Y) = p(M_1|Y) \text{ is large.}$$

Variational Bayes inference

VBEM: A global variational Bayes EM algorithm can be designed to obtain all needed (approximate) posteriors:

$$\begin{aligned} p(\theta, Z, K|Y) &\approx \widetilde{p}(\theta, Z, K) = \widetilde{p}(K)\, \widetilde{p}(\theta|K) \times \widetilde{p}(Z|K) \\ &= \widetilde{p}(K)\, \left(\widetilde{p}(\alpha|K)\, \widetilde{p}(\beta|K)\, \widetilde{p}(\pi|K)\right) \times \left(\prod_i \widetilde{p}(Z_i|K)\right) \end{aligned}$$

Variational Bayes inference

VBEM: A global variational Bayes EM algorithm can be designed to obtain all needed (approximate) posteriors:

$$\begin{aligned} p(\theta, Z, K|Y) \;\approx\; \widetilde{p}(\theta, Z, K) \;&=\; \widetilde{p}(K)\; \widetilde{p}(\theta|K) \times \widetilde{p}(Z|K) \\ &=\; \widetilde{p}(K)\; \left(\widetilde{p}(\alpha|K)\; \widetilde{p}(\beta|K)\; \widetilde{p}(\pi|K)\right) \times \left(\prod_i \widetilde{p}(Z_i|K)\right) \end{aligned}$$

Posterior quantities of interest:

- Goodness of fit criterion:

$$p(H_0|Y) \approx \widetilde{P}(K = 1)$$

- Residual graphon:

$$\widehat{\phi}(u, v) = \sum_K \widetilde{p}(K)\; \widetilde{\mathbb{E}}[\phi_K^{SBM}(u, v)]$$

Outline

Reminder of variational Bayes inference

Logistic regression

W-graphs and graphon

Goodness-of-fit

Illustrations

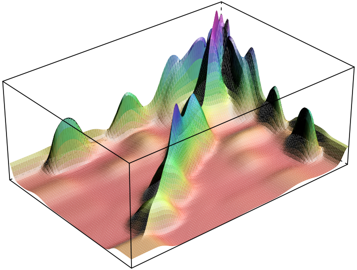
Some examples

network	size (n)	nb. covariates (d)	density	$\hat{p}(H_0 \mid Y)$
Blog	196	3	0.075	3e-172
Tree	51	3	0.54	2e-115
Karate	34	8	0.14	3e-2
Florentine (marriage)	16	3	0.17	0.995
Florentine (business)	16	3	0.125	0.991
Faux Dixon High	248	17	0.02	1
CKM	219	39	0.015	1
AddHealth 67	530	21	0.007	2e-25

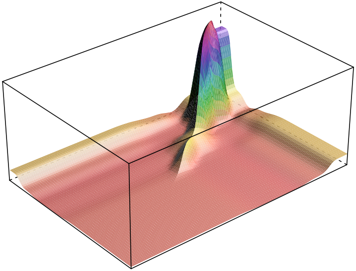
Political blog network

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

Inferred graphon (no covariate)



Residual graphon (3 covariates)

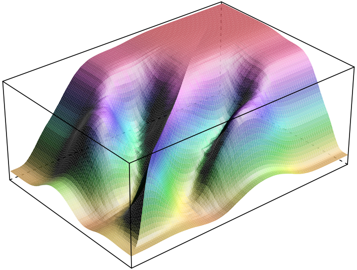


$\widetilde{P}(H_0) \simeq 10^{-172}$

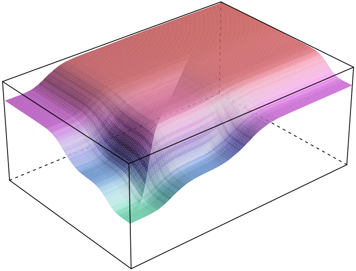
Tree network

$n = 51$ species ($N = 1275$ pairs), 3 covariates, density $= .54$

Inferred graphon (no covariate)



Residual graphon (3 covariates)

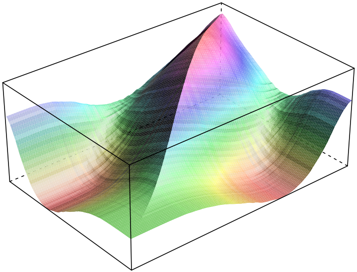


$\widetilde{P}(H_0) \simeq 10^{-115}$

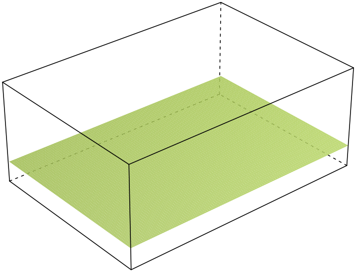
Florentine business

$n = 16$ families ($N = 120$ pairs), 3 covariates, density = .12

Inferred graphon (no covariate)



Residual graphon (3 covariates)

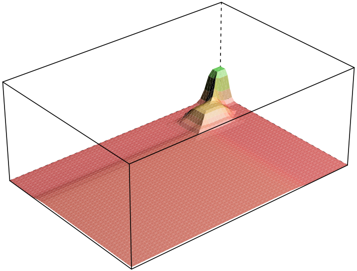


$\widetilde{P}(H_0) = .991$

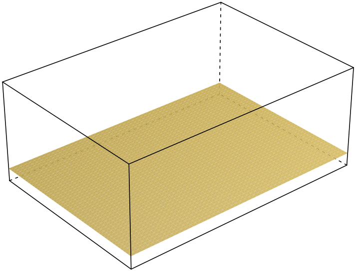
Physicians' friendship network

$n = 219$ individuals, 24 covariates (39 df), density $= .015$

Inferred graphon (no covariate)



Residual graphon (3 covariates)



$\widetilde{P}(H_0) \simeq 1$

Discussion

Contribution:

- Generic logistic regression model with a network-oriented residual term.
- Detour through SBM provides a natural goodness-of-fit criterion.
- R package on `github.com/platouche/gofNetwork` (soon on CRAN)

Discussion

Contribution:

- Generic logistic regression model with a network-oriented residual term.
- Detour through SBM provides a natural goodness-of-fit criterion.
- R package on `github.com/platouche/gofNetwork` (soon on CRAN)

Comments:

- Strongly relies on variational Bayes approximation of the posteriors.
- VBEM asymptotically accurate for logistic regression and SBM separately.
- No clue about accuracy in the combined model.

Discussion

Contribution:

- Generic logistic regression model with a network-oriented residual term.
- Detour through SBM provides a natural goodness-of-fit criterion.
- R package on `github.com/platouche/gofNetwork` (soon on CRAN)

Comments:

- Strongly relies on variational Bayes approximation of the posteriors.
- VBEM asymptotically accurate for logistic regression and SBM separately.
- No clue about accuracy in the combined model.

On-going work:

- Relevant *edge* covariates made of *node* covariates [10].
- Efficient sampling in the true posterior using sequential importance sampling:
  $\longrightarrow$ Move from $\widetilde{p}(\cdot)$ to $p(\cdot|Y)$ using a tempering scheme.

References I

J. Beal, M. and Z. Ghahramani.
The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures.
Bayes. Statist., 7:543–52, 2003.

B. Bollobás, S. Janson, and O. Riordan.
The phase transition in inhomogeneous random graphs.
Rand. Struct. Algo., 31(1):3–122, 2007.

J.-J. Daudin, L. Pierre, and C. Vacher.
Model for heterogeneous random networks using continuous latent variables and an application to a tree–fungus network.
Biometrics, 66(4):1043–1051, 2010.

P. Diaconis and S. Janson.
Graph limits and exchangeable random graphs.
Rend. Mat. Appl., 7(28):33–61, 2008.

O. Frank and F. Harary.
Cluster inference by using transitivity indices in empirical graphs.
J. Amer. Statist. Assoc., 77(380):835–840, 1982.

Steven Gazal, Jean-Jacques Daudin, and Stéphane Robin.
Accuracy of variational estimates for random graph mixture models.
Journal of Statistical Computation and Simulation, 82(6):849–862, 2012.

M. S. Handcock, A. E. Raftery, and J. M. Tantrum.
Model-based clustering for social networks.
Journal of the Royal Statistical Society: Series A (Statistics in Society), 170(2):301–354, 2007.
`doi: 10.1111/j.1467-985X.2007.00471.x.`

References II

J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky.
Bayesian model averaging: A tutorial.
Statistical Science, 14(4):382–417, 1999.

P. D. Hoff, A. E. Raftery, and M. S. Handcock.
Latent space approaches to social network analysis.
J. Amer. Statist. Assoc., 97(460):1090–98, 2002.

D. R. Hunter, S. M. Goodreau, and M. S. Handcock.
Goodness of fit of social network models.
Journal of the American Statistical Association, 103(481):248–258, 2008.

T. Jaakkola.
Advanced mean field methods: theory and practice, chapter Tutorial on variational approximation methods.
MIT Press, 2000.

T. S. Jaakkola and M. I. Jordan.
Bayesian parameter estimation via variational methods.
Statistics and Computing, 10(1):25–37, 2000.

P. Latouche, E. Birmelé, and C. Ambroise.
Variational bayesian inference and complexity control for stochastic block models.
Statis. Model., 12(1):93–115, 2012.

P. Latouche and S Robin.
Bayesian model averaging of stochastic block models to estimate the graphon function and motif frequencies in a W-graph model.
Technical report, arXiv:1310.6150, 2013.
to appear in Stat. Comput.

P. Latouche and S. Robin.
Variational bayes model averaging for graphon functions and motif frequencies inference in W-graph models.
Statistics and Computing*, pages 1–13, 2015.*

References III

P. Latouche, S. Robin, and S. Ouadah.
Goodness of fit of logistic models for random graphs.
Technical report, arXiv:1508.00286, 2015.

S.L. Lauritzen.
Graphical Models.
Oxford Statistical Science Series. Clarendon Press, 1996.

L. Lovász and B. Szegedy.
Limits of dense graph sequences.
Journal of Combinatorial Theory, Series B, *96(6):933 – 957, 2006.*

M. Mariadassou, S. Robin, and C. Vacher.
Uncovering structure in valued graphs: a variational approach.
Ann. Appl. Statist., *4(2):715–42, 2010.*

C. Matias and S. Robin.
Modeling heterogeneity in random graphs through latent space models: a selective review.
ESAIM: Proc., *47:55–74, 2014.*

Tom Minka.
Divergence measures and message passing.
Technical Report MSR-TR-2005-173, Microsoft Research Ltd, 2005.

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.*

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.`

References IV

S. Volant, M.-L. Martin Magniette, and S. Robin.
Variational bayes approach for model aggregation in unsupervised classification with markovian dependency.
Comput. Statis. & Data Analysis, *56(8):2375 – 2387, 2012.*

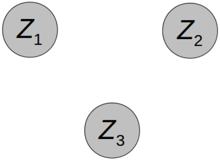
M. J. Wainwright and M. I. Jordan.
Graphical models, exponential families, and variational inference.
Found. Trends Mach. Learn., *1(1–2):1–305, 2008.*

Variational Bayes for heterogeneous network models

Graphical model formalism [17].

Variational Bayes for heterogeneous network models

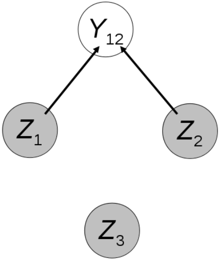
Graphical model formalism [17].



'Frequentist' setting ($p(\cdot|\theta)$):

- iid Z_i's,

Variational Bayes for heterogeneous network models

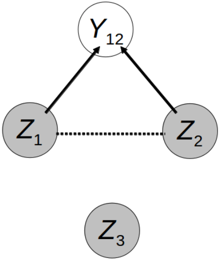


Graphical model formalism [17].

'Frequentist' setting ($p(\cdot|\theta)$):

- iid Z_i's,
- $p(Y_{ij}|Z_i, Z_j)$,

Variational Bayes for heterogeneous network models

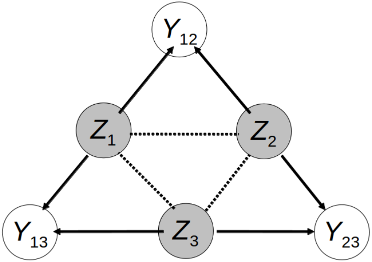


Graphical model formalism [17].

'Frequentist' setting ($p(\cdot|\theta)$):

- iid Z_i's,
- $p(Y_{ij}|Z_i, Z_j)$,
- $p(Z_i, Z_j|Y)$: graph moralization,

Variational Bayes for heterogeneous network models

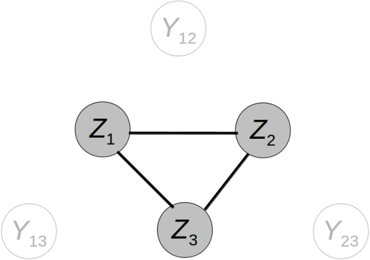


Graphical model formalism [17].

'Frequentist' setting ($p(\cdot|\theta)$):

- iid Z_i's,
- $p(Y_{ij}|Z_i, Z_j)$,
- $p(Z_i, Z_j|Y)$: graph moralization,
- this holds for each pair (i, j),

Variational Bayes for heterogeneous network models



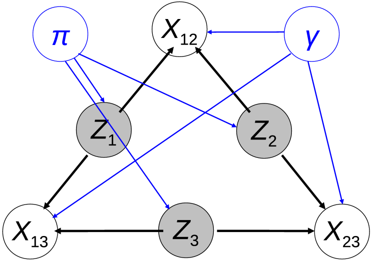
Graphical model formalism [17].

'Frequentist' setting ($p(\cdot|\theta)$):

- iid Z_i's,
- $p(Y_{ij}|Z_i, Z_j)$,
- $p(Z_i, Z_j|Y)$: graph moralization,
- this holds for each pair (i, j),

Dependency graph of $p(Z|Y, \theta) =$ clique.
→ No factorization can be hoped (unlike for HMM).
→ $p(Z|Y, \theta)$ can not be computed (efficiently).

Variational Bayes for heterogeneous network models



Graphical model formalism [17].

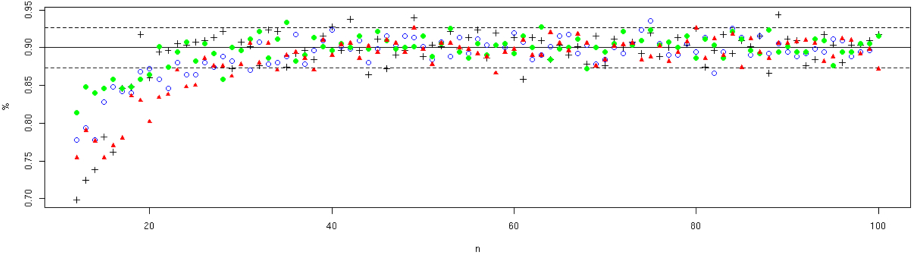
'Frequentist' setting ($p(\cdot|\theta)$):

- iid Z_i's,
- $p(Y_{ij}|Z_i, Z_j)$,
- $p(Z_i, Z_j|Y)$: graph moralization,
- this holds for each pair (i, j),

Dependency graph of $p(Z|Y, \theta)$ = clique.
$\rightarrow$ No factorization can be hoped (unlike for HMM).
$\rightarrow$ $p(Z|Y, \theta)$ can not be computed (efficiently).
Bayesian setting. Things get worst.

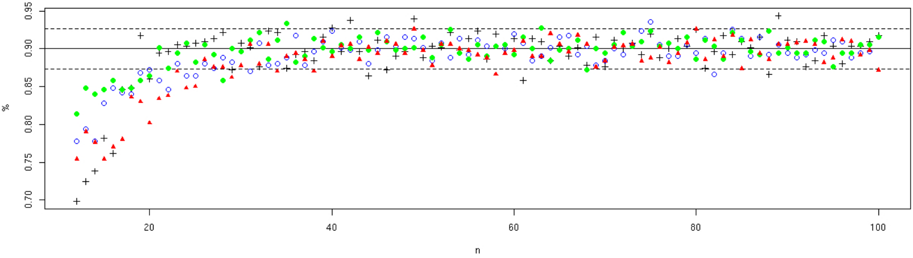
Accuracy of VBEM estimates for SBM: Simulation study

Credibility intervals: π_1: $+$, γ_{11}: $\triangle$, γ_{12}: $\circ$, γ_{22}: $\bullet$

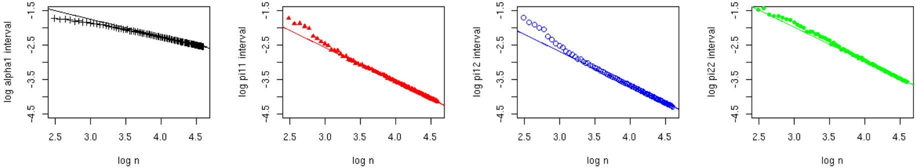


Accuracy of VBEM estimates for SBM: Simulation study

Credibility intervals: π_1: +, γ_{11}: △, γ_{12}: ○, γ_{22}: ●

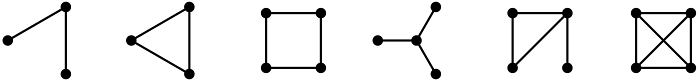


Width of the posterior credibility intervals. π_1, γ_{11}, γ_{12}, γ_{22}



[6]

A first criterion for goodness-of-fit: Motifs frequency



- Network motifs have a biological or sociological interpretation in terms of building blocks of the global network

  → Triangles = 'friends of my friends are my friends'.

- Latent space graph models only describe binary interactions, conditional on the latent positions

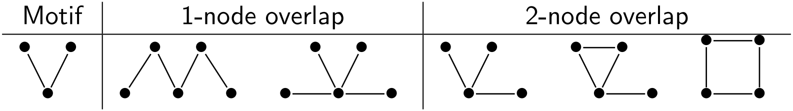
→ Goodness of fit criterion based on motif frequencies? [15].

Moments of motif counts

First moments $\mathbb{E}N(m)$, $\mathbb{V}N(m)$ known for exchangeable graphs (incl. SBM) [22]:

$$\mathbb{E}_{SBM}N(m) \propto \mu_{SBM}(m), \qquad \mu_{SBM}(m) = \text{motif occurrence probability.}$$

$\mathbb{V}N(m)$ depends on the frequency of super-motifs:



Moments under W-graph: Motif probability under the W -graph can be estimated as

$$\widehat{\mu}(m) = \sum_{k} \widetilde{P}(K)\widetilde{\mathbb{E}}(\mu^{K}_{SBM}(m))$$

Estimates of $\mathbb{E}_W N(m)$ and $\mathbb{V}_W N(m)$: see [14].

Network frequencies in the blog network

Motif	Count $(\times 10^3)$	Mean $(\times 10^3)$	Std. dev. $(\times 10^3)$	Rel. diff. $(N - \mathbb{E})/\sqrt{\mathbb{V}}$
	29.7	39.7	8.3	-1.20
	3.8	4.6	1.3	-0.62
	608.7	968.3	336.8	-1.07
	279.8	428.9	154.0	-0.97
	47.4	74.5	35.1	-0.77
	270.5	397.0	177.0	-0.71
	62.1	87.8	47.4	-0.54
	6.5	8.8	5.4	-0.43

No specific structure seems to be exceptional wrt the model's expectations.