Network analysis with the stochastic block model
(using variational approximations)

S. Robin

INRA / AgroParisTech

LIP6, February 2015, Paris

Outline

1. State-space models for networks (inc. SBM)
2. Variational inference (inc. SBM)
3. Extensions of SBM
4. (Variational) Bayesian model averaging
5. Towards W-graphs
6. Goodness-of-fit using network motif

Heterogeneity in interaction networks

Understanding network structure

Networks describe interactions between entities.

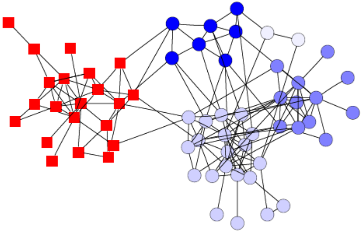
Observed networks display heterogeneous topologies, that one would like to decipher and better understand.

Understanding network structure

Networks describe interactions between entities.

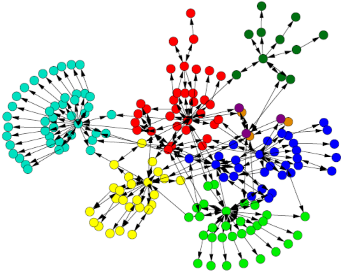
Observed networks display heterogeneous topologies, that one would like to decipher and better understand.

Dolphine social network.



[Newman and Girvan (2004)]

Hyperlink network.



Modeling network heterogeneity

Latent variable models allow to capture the underlying structure of a network (see review [Matias and R. (2014)]).

Modeling network heterogeneity

Latent variable models allow to capture the underlying structure of a network (see review [Matias and R. (2014)]).

General setting for binary graphs. [Bollobás *et al.* (2007)]:

- A latent (unobserved) variable Z_i is associated with each node:

$$\{Z_i\} \text{ iid } \sim \pi$$

- Edges $Y_{ij} = \mathbb{I}\{i \sim j\}$ are independent conditionally to the Z_i's:

$$\{Y_{ij}\} \text{ independent } |\{Z_i\} : \Pr\{Y_{ij} = 1\} = \gamma(Z_i, Z_j)$$

Modeling network heterogeneity

Latent variable models allow to capture the underlying structure of a network (see review [Matias and R. (2014)]).

General setting for binary graphs. [Bollobás *et al.* (2007)]:

- A latent (unobserved) variable Z_i is associated with each node:

$$\{Z_i\} \text{ iid } \sim \pi$$

- Edges $Y_{ij} = \mathbb{I}\{i \sim j\}$ are independent conditionally to the Z_i's:

$$\{Y_{ij}\} \text{ independent } |\{Z_i\} : \Pr\{Y_{ij} = 1\} = \gamma(Z_i, Z_j)$$

We focus here on model approaches, in contrast with, e.g.

- Graph clustering [Girvan and Newman (2002)], [Newman (2004)];
- Spectral clustering [von Luxburg *et al.* (2008)].

Latent space models

State-space model: principle.

Latent space models

State-space model: principle.

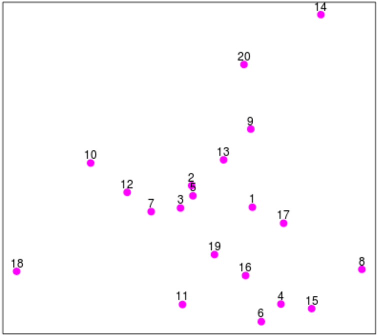
- Consider n nodes ($i = 1..n$);

Latent space models

State-space model: principle.

- Consider n nodes ($i = 1..n$);
- Z_i = unobserved position of no
  e.g.

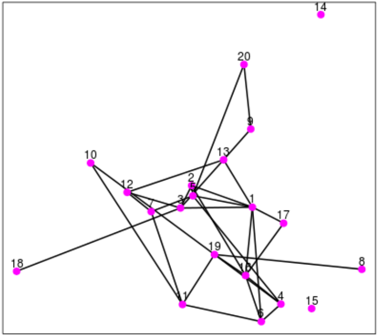
$$\{Z_i\} \text{ iid } \sim \mathcal{N}(0, I)$$



Latent space models

State-space model: principle.

- Consider n nodes ($i = 1..n$);
- Z_i = unobserved position of no
  e.g.
  $$\{Z_i\} \text{ iid } \sim \mathcal{N}(0, I)$$
- Edge $\{Y_{ij}\}$ independent given $\{$
  e.g.
  $$\Pr\{Y_{ij} = 1\} = \gamma(Z_i, Z_j)$$
  $$\gamma(z, z') = f(||z, z'||).$$



Latent space models

State-space model: principle.

- Consider n nodes ($i = 1..n$);
- Z_i = unobserved position of node i, e.g.
$$\{Z_i\} \text{ iid } \sim \mathcal{N}(0, I)$$
- Edge $\{Y_{ij}\}$ independent given $\{Z_i\}$, e.g.
$$\Pr\{Y_{ij} = 1\} = \gamma(Z_i, Z_j)$$
$$\gamma(z, z') = f(||z, z'||).$$

$$Y = \begin{pmatrix} 0 & 1 & 1 & 0 & 1 & \dots \\ 0 & 0 & 1 & 0 & 1 & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 1 & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

A variety of state-space models

Continuous. Latent position models.

- [Hoff *et al.* (2002)]:

$$Z_i \in \mathbb{R}^d, \qquad \text{logit}[\gamma(z, z')] = a - |z - z'|$$

- [Handcock *et al.* (2007)]:

$$Z_i \sim \sum_k p_k \mathcal{N}_d(\mu_k, \sigma_k^2 I)$$

- [Lovász and Szegedy (2006)]:

$$Z_i \sim \mathcal{U}_{[0,1]}, \qquad \gamma(z, z') : [0,1]^2 \to [0,1] =: \text{graphon function}$$

- [Daudin *et al.* (2010)]:

$$Z_i \in \mathcal{S}_K, \qquad \gamma(z, z') = \sum_{k,\ell} z_k z'_\ell \gamma_{k\ell}$$

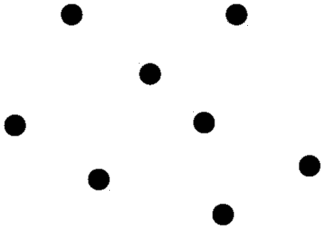
Stochastic Block Model (SBM)

A mixture model for random graphs.
[Nowicki and Snijders (2001)]

Stochastic Block Model (SBM)

A mixture model for random graphs.
[Nowicki and Snijders (2001)]

- Consider n nodes ($i = 1..n$);



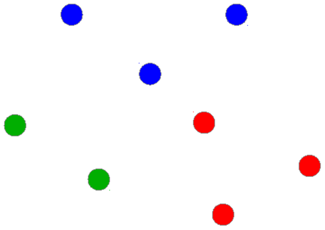
Stochastic Block Model (SBM)

A mixture model for random graphs.
[Nowicki and Snijders (2001)]

- Consider n nodes ($i = 1..n$);
- Z_i = unobserved label of node i:

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

$\pi = (\pi_1, ...\pi_K)$;



Stochastic Block Model (SBM)

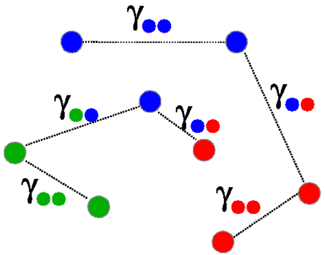
A mixture model for random graphs.
[Nowicki and Snijders (2001)]

- Consider n nodes ($i = 1..n$);
- Z_i = unobserved label of node i:

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

  $\pi = (\pi_1, ...\pi_K)$;
- Edge Y_{ij} depends on the labels:
  $\{Y_{ij}\}$ independent given $\{Z_i\}$,

  $$\Pr\{Y_{ij} = 1\} = \gamma(Z_i, Z_j)$$



Stochastic Block Model (SBM)

A mixture model for random graphs.
[Nowicki and Snijders (2001)]

- Consider n nodes ($i = 1..n$);

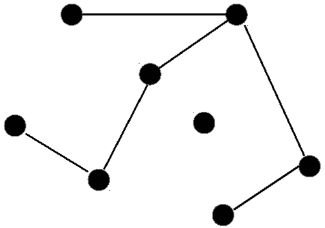
- Z_i = unobserved label of node i:

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

$\pi = (\pi_1, ...\pi_K)$;

- Edge Y_{ij} depends on the labels:
  $\{Y_{ij}\}$ independent given $\{Z_i\}$,

$$\Pr\{Y_{ij} = 1\} = \gamma(Z_i, Z_j)$$



Variational inference

Incomplete data models

Aim. Based on the observed network $Y = (Y_{ij})$, we want to infer

- the parameters
$$\theta = (\pi, \gamma)$$
- the hidden states
$$Z = (Z_i)$$

Incomplete data models

Aim. Based on the observed network $Y = (Y_{ij})$, we want to infer

- the parameters
$$\theta = (\pi, \gamma)$$
- the hidden states
$$Z = (Z_i)$$

State space models belong to the class of incomplete data models as

- the edges (Y_{ij}) are observed,
- the latent positions (or status) (Z_i) are not.

$\rightarrow$ usual issue in unsupervised classification.

Frequentist maximum likelihood inference

Likelihood. The (log-)likelihood

$$\log P(Y;\theta) = \log \sum_Z P(Y, Z; \theta)$$

can not be computed.

Frequentist maximum likelihood inference

Likelihood. The (log-)likelihood

$$\log P(Y;\theta) = \log \sum_Z P(Y, Z;\theta)$$

can not be computed.

EM trick. But it can be decomposed as

$$\log P(Y;\theta) = \log P(Y, Z;\theta) - \log P(Z|Y;\theta),$$

Frequentist maximum likelihood inference

Likelihood. The (log-)likelihood

$$\log P(Y;\theta) = \log \sum_Z P(Y, Z; \theta)$$

can not be computed.

EM trick. But it can be decomposed as

$$\log P(Y;\theta) = \log P(Y, Z; \theta) - \log P(Z|Y;\theta),$$

the conditional expectation of which gives

$$\mathbb{E}[\log P(Y;\theta)|Y] \quad = \quad \mathbb{E}[\log P(Y, Z;\theta)|Y] - \mathbb{E}[\log P(Z|Y;\theta)|Y]$$

Frequentist maximum likelihood inference

Likelihood. The (log-)likelihood

$$\log P(Y;\theta) = \log \sum_Z P(Y,Z;\theta)$$

can not be computed.

EM trick. But it can be decomposed as

$$\log P(Y;\theta) = \log P(Y,Z;\theta) - \log P(Z|Y;\theta),$$

the conditional expectation of which gives

$$\begin{aligned} \mathbb{E}[\log P(Y;\theta)|Y] &= \mathbb{E}[\log P(Y,Z;\theta)|Y] - \mathbb{E}[\log P(Z|Y;\theta)|Y] \\ \log P(Y;\theta) &= \mathbb{E}[\log P(Y,Z;\theta)|Y] + \mathcal{H}[P(Z|Y;\theta)] \end{aligned}$$

where $\mathcal{H}$ stands for the entropy.

EM algorithm

Aims at maximizing the log-likelihood

$$\log P(Y; \theta)$$

through the alternation of two steps [Dempster *et al.* (1977)]

EM algorithm

Aims at maximizing the log-likelihood

$$\log P(Y;\theta)$$

through the alternation of two steps [Dempster *et al.* (1977)]

M-step: maximize $\mathbb{E}[\log P(Y,Z;\theta)|Y]$ with respect to θ
$\rightarrow$ generally similar to standard MLE.

EM algorithm

Aims at maximizing the log-likelihood

$$\log P(Y; \theta)$$

through the alternation of two steps [Dempster *et al.* (1977)]

M-step: maximize $\mathbb{E}[\log P(Y, Z; \theta)|Y]$ with respect to θ
$\rightarrow$ generally similar to standard MLE.

E-step: calculate $P(Z|Y)$ (at least, some moments)

EM algorithm

Aims at maximizing the log-likelihood

$$\log P(Y;\theta)$$

through the alternation of two steps [Dempster *et al.* (1977)]

M-step: maximize $\mathbb{E}[\log P(Y,Z;\theta)|Y]$ with respect to θ
$\rightarrow$ generally similar to standard MLE.

E-step: calculate $P(Z|Y)$ (at least, some moments)
$\rightarrow$ sometimes straightforward (independent mixture models: Bayes formula)

EM algorithm

Aims at maximizing the log-likelihood

$$\log P(Y; \theta)$$

through the alternation of two steps [Dempster *et al.* (1977)]

M-step: maximize $\mathbb{E}[\log P(Y, Z; \theta) | Y]$ with respect to θ
$\rightarrow$ generally similar to standard MLE.

E-step: calculate $P(Z|Y)$ (at least, some moments)
$\rightarrow$ sometimes straightforward (independent mixture models: Bayes formula)
$\rightarrow$ sometimes tricky but doable (HMMs: forward-backward recursion)

EM algorithm

Aims at maximizing the log-likelihood

$$\log P(Y;\theta)$$

through the alternation of two steps [Dempster *et al.* (1977)]

M-step: maximize $\mathbb{E}[\log P(Y,Z;\theta)|Y]$ with respect to θ
$\rightarrow$ generally similar to standard MLE.

E-step: calculate $P(Z|Y)$ (at least, some moments)
$\rightarrow$ sometimes straightforward (independent mixture models: Bayes formula)
$\rightarrow$ sometimes tricky but doable (HMMs: forward-backward recursion)
$\rightarrow$ sometimes impossible (SBM: ...)

Conditional distribution of Z

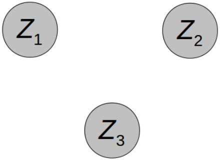
EM works provided that we can calculate

$$P(Z|Y;\theta)$$

Conditional distribution of Z

EM works provided that we can calculate

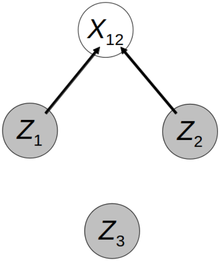
$$P(Z|Y;\theta)$$



Conditional distribution of Z

EM works provided that we can calculate

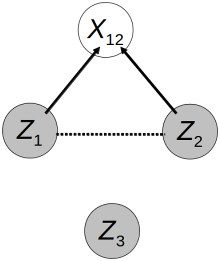
$$P(Z|Y;\theta)$$



Conditional distribution of Z

EM works provided that we can calculate

$$P(Z|Y;\theta)$$

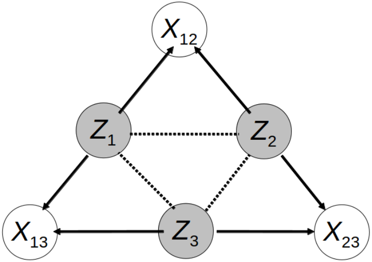


Conditional distribution of Z

EM works provided that we can calculate

$$P(Z|Y;\theta)$$

but we can not for state space models for graph.

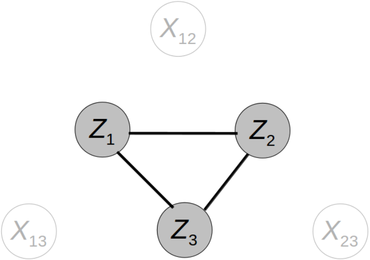


Conditional distribution of Z

EM works provided that we can calculate

$$P(Z|Y;\theta)$$

but we can not for state space models for graph.



Conditional distribution. The dependency graph of Z given Y is a clique.
$\rightarrow$ No factorization can be hoped (unlike for HMM).
$\rightarrow$ $P(Z|Y;\theta)$ can not be computed (efficiently).
$\rightarrow$ Variational techniques may help as they provide

$$Q(Z) \simeq P(Z|Y).$$

Variational inference

Lower bound of the log-likelihood. For any distribution $Q(Z)$ [Jaakkola (2000),Wainwright and Jordan (2008)],

$$\log P(Y) \geq \log P(Y) - KL[Q(Z), P(Z|Y)]$$

Variational inference

Lower bound of the log-likelihood. For any distribution $Q(Z)$ [Jaakkola (2000),Wainwright and Jordan (2008)],

$$\begin{aligned}\log P(Y) &\geq \log P(Y) - KL[Q(Z), P(Z|Y)] \\ &= \int Q(Z) \log P(Y,Z) \, \mathrm{d}Z - \int Q(Z) \log Q(Z) \, \mathrm{d}Z\end{aligned}$$

Variational inference

Lower bound of the log-likelihood. For any distribution $Q(Z)$ [Jaakkola (2000),Wainwright and Jordan (2008)],

$$\begin{aligned}\log P(Y) &\geq \log P(Y) - KL[Q(Z), P(Z|Y)] \\ &= \int Q(Z) \log P(Y, Z) \; \mathrm{d}Z - \int Q(Z) \log Q(Z) \; \mathrm{d}Z \\ &= \mathbb{E}_Q[\log P(Y, Z)] + \mathcal{H}[Q(Z)]\end{aligned}$$

Variational inference

Lower bound of the log-likelihood. For any distribution $Q(Z)$ [Jaakkola (2000),Wainwright and Jordan (2008)],

$$\begin{aligned}\log P(Y) &\geq \log P(Y) - KL[Q(Z), P(Z|Y)] \\ &= \int Q(Z) \log P(Y, Z) \, \mathrm{d}Z - \int Q(Z) \log Q(Z) \, \mathrm{d}Z \\ &= \mathbb{E}_Q[\log P(Y, Z)] + \mathcal{H}[Q(Z)]\end{aligned}$$

Link with EM. This is similar to

$$\log P(Y) = \mathbb{E}[\log P(Y, Z)|Y] + \mathcal{H}[P(Z|Y)]$$

replacing $P(Z|Y)$ with $Q(Z)$.

Variational EM algorithm

Variational EM.

- M-step: compute

$$\widehat{\theta} = \arg\max_{\theta} \mathbb{E}_{Q^*}[\log P(Y, Z; \theta)].$$

Variational EM algorithm

Variational EM.

- M-step: compute

$$\widehat{\theta} = \arg\max_{\theta} \mathbb{E}_{Q^*}[\log P(Y, Z; \theta)].$$

- E-step: replace the calculation of $P(Z|Y)$ with the search of

$$Q^*(Z) = \arg\min_{Q \in \mathcal{Q}} KL[Q(Z), P(Z|Y)].$$

Variational EM algorithm

Variational EM.

- M-step: compute
$$\widehat{\theta} = \arg\max_{\theta} \mathbb{E}_{Q^*}[\log P(Y, Z; \theta)].$$
- E-step: replace the calculation of $P(Z|Y)$ with the search of
$$Q^*(Z) = \arg\min_{Q \in \mathcal{Q}} KL[Q(Z), P(Z|Y)].$$

$\rightarrow$ Taking $\mathcal{Q} = \{\text{all possible distributions}\}$ gives $Q^*(Z) = P(Z|Y)$ ... like EM does.

$\rightarrow$ Variational approximations rely on the choice a set $\mathcal{Q}$ of 'good' and 'tractable' ditributions.

Variational EM for SBM [Daudin *et al.* (2008)]

Distribution class. $\mathcal{Q}$ = set of factorisable distributions:

$$\mathcal{Q} = \{Q : Q(Z) = \prod_i Q_i(Z_i)\}, \qquad Q_i(Z_i) = \prod_k \tau_{ik}^{Z_{ik}}.$$

→ The approximate joint distribution is $Q(Z_i, Z_j) = Q_i(Z_i)Q_j(Z_j)$.

Variational EM for SBM [Daudin *et al.* (2008)]

Distribution class. $\mathcal{Q}$ = set of factorisable distributions:

$$\mathcal{Q} = \{Q : Q(Z) = \prod_i Q_i(Z_i)\}, \qquad Q_i(Z_i) = \prod_k \tau_{ik}^{Z_{ik}}.$$

$\rightarrow$ The approximate joint distribution is $Q(Z_i, Z_j) = Q_i(Z_i)Q_j(Z_j)$.

The optimal approximation within this class satisfies a fix-point relation:

$$\tau_{ik} \propto \pi_k \prod_{j \neq i} \prod_{\ell} f_{k\ell}(Y_{ij})^{\tau_{j\ell}}$$

also known as mean-field approximation in physics [Parisi (1988)].

Variational EM for SBM [Daudin *et al.* (2008)]

Distribution class. $\mathcal{Q}$ = set of factorisable distributions:

$$\mathcal{Q} = \{Q : Q(Z) = \prod_i Q_i(Z_i)\}, \qquad Q_i(Z_i) = \prod_k \tau_{ik}^{Z_{ik}}.$$

$\rightarrow$ The approximate joint distribution is $Q(Z_i, Z_j) = Q_i(Z_i)Q_j(Z_j)$.

The optimal approximation within this class satisfies a fix-point relation:

$$\tau_{ik} \propto \pi_k \prod_{j \neq i} \prod_\ell f_{k\ell}(Y_{ij})^{\tau_{j\ell}}$$

also known as mean-field approximation in physics [Parisi (1988)].

Variational estimates.

- No general statistical guaranty for variational estimates.
- SBM is a very specific case for which the variational approximation is asymptotically exact [Celisse *et al.* (2012),Mariadassou and Matias (2015)].

Variational Bayes inference

Bayesian perspective.
$\theta = (\pi, \gamma)$ is random.

Model:

Variational Bayes inference

Bayesian perspective.
$\theta = (\pi, \gamma)$ is random.

Model:

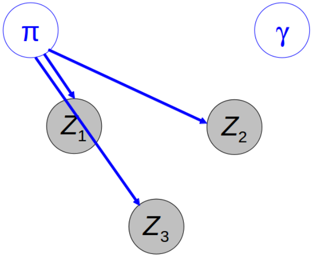
- $P(\theta)$

Variational Bayes inference

Bayesian perspective.
$\theta = (\pi, \gamma)$ is random.

Model:

- $P(\theta)$
- $P(Z|\pi)$

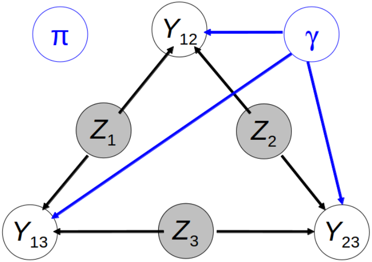


Variational Bayes inference

Bayesian perspective.
$\theta = (\pi, \gamma)$ is random.

Model:

- $P(\theta)$
- $P(Z|\pi)$
- $P(Y|\gamma, Z)$

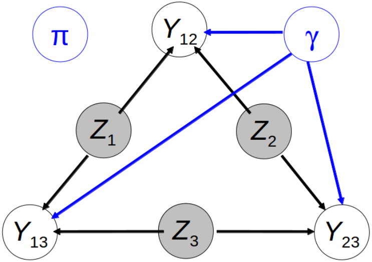


Variational Bayes inference

Bayesian perspective.
$\theta = (\pi, \gamma)$ is random.

Model:

- $P(\theta)$
- $P(Z|\pi)$
- $P(Y|\gamma, Z)$



Bayesian inference. The aim is then to get the joint conditional distribution of the parameters and of the hidden variables:

$$P(\theta, Z|Y).$$

Variational Bayes algorithm

Variational Bayes. As $P(\theta, Z|Y)$ is intractable, one look formula

$$Q^*(\theta, Z) = \arg\min_{Q \in \mathcal{Q}} KL\left[Q(\theta, Z) || P(\theta, Z|Y)\right]$$

taking, e.g., $\mathcal{Q} = \{Q(\theta, Z) = Q_\theta(\theta) Q_Z(Z)\}$

Variational Bayes algorithm

Variational Bayes. As $P(\theta, Z|Y)$ is intractable, one look formula

$$Q^*(\theta, Z) = \arg\min_{Q \in \mathcal{Q}} KL\left[Q(\theta, Z) || P(\theta, Z|Y)\right]$$

taking, e.g., $\mathcal{Q} = \{Q(\theta, Z) = Q_\theta(\theta) Q_Z(Z)\}$

Variational Bayes EM (VBEM). When

- $P(Z, Y|\theta)$ belongs to the exponential family,
- $P(\theta)$ is the corresponding conjugate prior,

Q^* can be obtained iteratively as [Beal and Ghahramani (2003)]

$$\log Q_\theta^h(\theta) \propto \mathbb{E}_{Q_Z^{h-1}}\left[\log P(Z, Y, \theta)\right], \quad \log Q_Z^h(Z) \propto \mathbb{E}_{Q_\theta^h}\left[\log P(Z, Y, \theta)\right].$$

Variational Bayes algorithm

Variational Bayes. As $P(\theta, Z|Y)$ is intractable, one look formula

$$Q^*(\theta, Z) = \arg\min_{Q \in \mathcal{Q}} KL\left[Q(\theta, Z) || P(\theta, Z|Y)\right]$$

taking, e.g., $\mathcal{Q} = \{Q(\theta, Z) = Q_\theta(\theta) Q_Z(Z)\}$

Variational Bayes EM (VBEM). When

- $P(Z, Y|\theta)$ belongs to the exponential family,
- $P(\theta)$ is the corresponding conjugate prior,

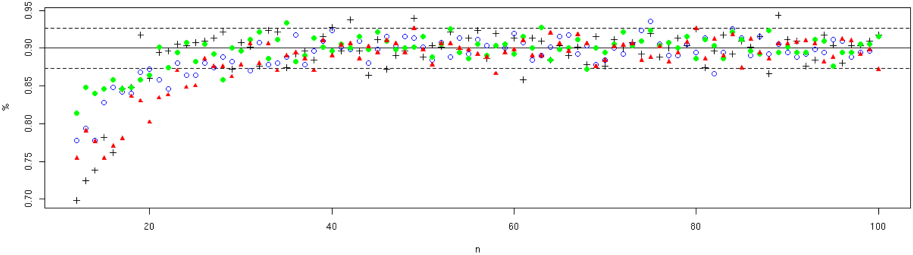
Q^* can be obtained iteratively as [Beal and Ghahramani (2003)]

$$\log Q_\theta^h(\theta) \propto \mathbb{E}_{Q_Z^{h-1}}\left[\log P(Z, Y, \theta)\right], \quad \log Q_Z^h(Z) \propto \mathbb{E}_{Q_\theta^h}\left[\log P(Z, Y, \theta)\right].$$

Application to SBM: [Latouche *et al.* (2012)]

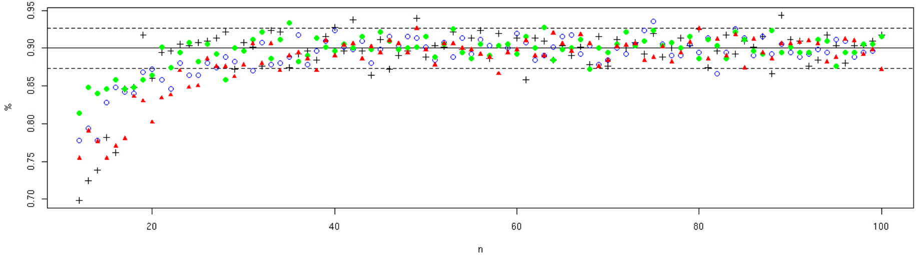
VBEM: Simulation study [Gazal *et al.* (2012)]

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

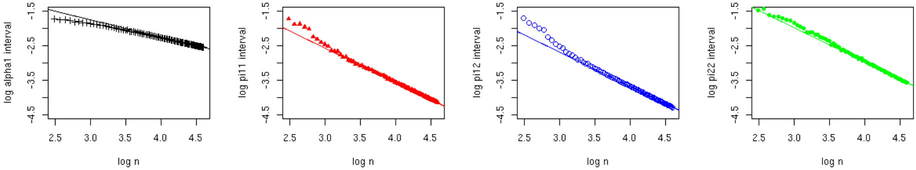


VBEM: Simulation study [Gazal *et al.* (2012)]

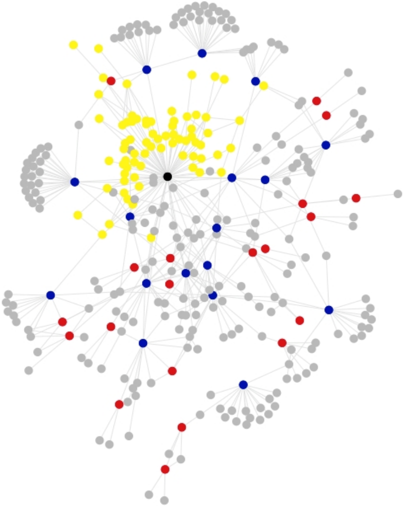
Credibility intervals: π_1: $+$, γ_{11}: $\triangle$, γ_{12}: $\circ$, γ_{22}: $\bullet$



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

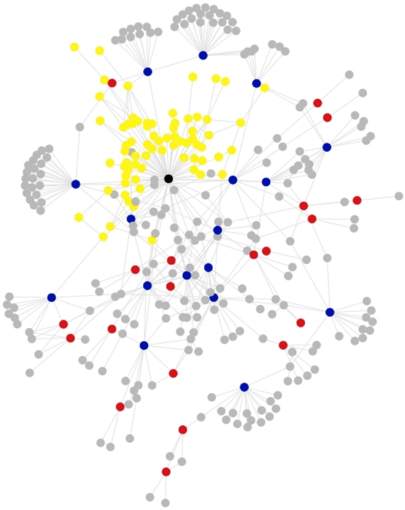


SBM analysis of *E. coli* operon networks



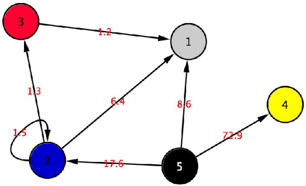
[Picard *et al.* (2009)]

SBM analysis of *E. coli* operon networks

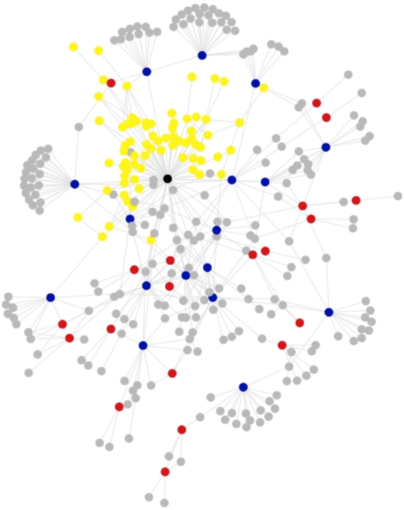


[Picard *et al.* (2009)]

Meta-graph representation.

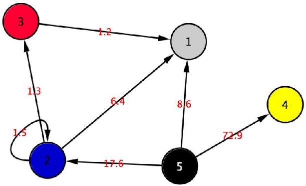


SBM analysis of *E. coli* operon networks

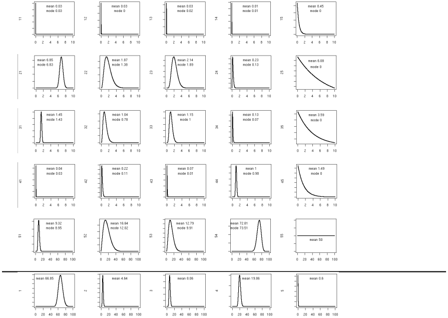


[Picard *et al.* (2009)]

Meta-graph representation.



Parameter estimates. $K = 5$



Some extensions of SBM

GLM framework [Mariadassou *et al.* (2010)]

SBM can be combined with generalized linear models (GLM) to deal with both valued graphs and covariates.

GLM framework [Mariadassou *et al.* (2010)]

SBM can be combined with generalized linear models (GLM) to deal with both valued graphs and covariates.

Valued graphs. Simply adapt the emission distribution:

$$(Y_{ij} \mid Z_i = k, Z_j = \ell) \sim F_{k\ell}(\cdot) := F(\cdot; \gamma_{k\ell})$$

where $F =$ some (parametric) distribution: Bernoulli (regular SBM), Poisson, Gaussian, etc.

GLM framework [Mariadassou *et al.* (2010)]

SBM can be combined with generalized linear models (GLM) to deal with both valued graphs and covariates.

Valued graphs. Simply adapt the emission distribution:

$$(Y_{ij} \mid Z_i = k, Z_j = \ell) \sim F_{k\ell}(\cdot) := F(\cdot; \gamma_{k\ell})$$

where $F =$ some (parametric) distribution: Bernoulli (regular SBM), Poisson, Gaussian, etc.

Covariates. In the context of exponential family, covariates x can be accounted for via a regression term

$$g(\mathbb{E}Y_{ij} \mid Z_i = k, Z_j = \ell) = \gamma_{k\ell} + x_{ij}\beta$$

where

- g stands for the link function (logit, log, identity, etc.);
- β may depend or not on the groups ($\beta \to \beta_{k\ell}$).

Tree interaction network

Data: $n = 51$ tree species,
X_{ij} = number of shared parasites
[Vacher *et al.* (2008)].

Tree interaction network

Data: $n = 51$ tree species,
X_{ij} = number of shared parasites
[Vacher *et al.* (2008)].

Model:

$$X_{ij} \sim \mathcal{P}(e^{\gamma_{k\ell}}),$$

$\gamma_{k\ell}$ = log-mean number of shared parasites.

Tree interaction network

Data: $n = 51$ tree species,
X_{ij} = number of shared parasites
[Vacher *et al.* (2008)].

Model:

$$X_{ij} \sim \mathcal{P}(e^{\gamma_{k\ell}}),$$

$\gamma_{k\ell}$ = log-mean number of shared parasites.

Results: ICL selects $K = 7$ groups that are partly related with phylums.

Tree interaction network

$e^{\widehat{\gamma}_{k\ell}}$	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

Data: $n = 51$ tree species,
X_{ij} = number of shared parasites
[Vacher *et al.* (2008)].

Model:

$$X_{ij} \sim \mathcal{P}(e^{\gamma_{k\ell}}),$$

$\gamma_{k\ell}$ = log-mean number of shared parasites.

Results: ICL selects $K = 7$ groups that are partly related with phylums.

Tree interaction network

Data: $n = 51$ tree species,
X_{ij} = number of shared parasites
[Vacher *et al.* (2008)].

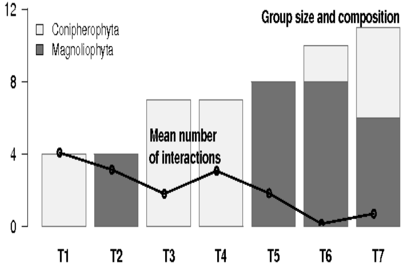
$e^{\widehat{\gamma}_{k\ell}}$	T1	T2	T3	T4	T5	T6	T7
T1	14.46	4.19	5.99	7.67	2.44	0.13	1.43
T2		14.13	0.68	2.79	4.84	0.53	1.54
T3			3.19	4.10	0.66	0.02	0.69
T4				7.42	2.57	0.04	1.05
T5					3.64	0.23	0.83
T6						0.04	0.06
T7							0.27
$\widehat{\pi}_k$	7.8	7.8	13.7	13.7	15.7	19.6	21.6

Model:

$$X_{ij} \sim \mathcal{P}(e^{\gamma_{k\ell}}),$$

$\gamma_{k\ell}$ = log-mean number of shared parasites.

Results: ICL selects $K = 7$ groups that are partly related with phylums.



Accounting for taxonomic distance

Model: $x_{ij} = \text{distance}(i, j)$

$$Y_{ij} \sim \mathcal{P}(e^{\gamma_{k\ell} + \beta x_{ij}}).$$

Accounting for taxonomic distance

Model: $x_{ij} = \text{distance}(i, j)$

$$Y_{ij} \sim \mathcal{P}(e^{\gamma_{k\ell} + \beta x_{ij}}).$$

Results: $\widehat{\beta} = -0.317$.
$\rightarrow$ for $\overline{x} = 3.82$,

$$e^{\widehat{\beta}\overline{x}} = .298$$

$\rightarrow$ The mean number of shared parasites decreases with taxonomic distance.

Accounting for taxonomic distance

Model: $x_{ij} = \text{distance}(i, j)$

$$Y_{ij} \sim \mathcal{P}(e^{\gamma_{k\ell} + \beta x_{ij}}).$$

Results: $\widehat{\beta} = -0.317$.
$\rightarrow$ for $\overline{x} = 3.82$,

$$e^{\widehat{\beta}\overline{x}} = .298$$

$\rightarrow$ The mean number of shared parasites decreases with taxonomic distance.

$e^{\widehat{\lambda}_{k\ell}}$	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			

Accounting for taxonomic distance

$e^{\widehat{\lambda}_{k\ell}}$	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			

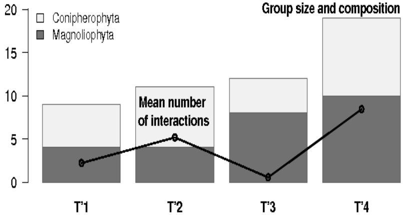
Model: $x_{ij} = \text{distance}(i,j)$

$$Y_{ij} \sim \mathcal{P}(e^{\gamma_{k\ell} + \beta x_{ij}}).$$

Results: $\widehat{\beta} = -0.317$.
$\rightarrow$ for $\overline{x} = 3.82$,

$$e^{\widehat{\beta}\overline{x}} = .298$$

$\rightarrow$ The mean number of shared parasites decreases with taxonomic distance.



Accounting for taxonomic distance

Model: $x_{ij} = \text{distance}(i,j)$

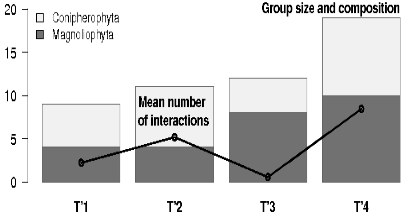
$$Y_{ij} \sim \mathcal{P}(e^{\gamma_{k\ell}+\beta x_{ij}}).$$

$e^{\widehat{\lambda}_{k\ell}}$	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			

Results: $\widehat{\beta} = -0.317$.
$\rightarrow$ for $\overline{x} = 3.82$,

$$e^{\widehat{\beta}\overline{x}} = .298$$

$\rightarrow$ The mean number of shared parasites decreases with taxonomic distance.



$\rightarrow$ Groups are no longer associated with the phylogenetic structure.
$\rightarrow$ Mixture = residual heterogeneity of the regression.

(Variational) Bayesian model averaging

Model choice

Model selection. The number of classes K generally needs to be estimated.

- In the frequentist setting, an approximate ICL criterion can be derived

$$ICL = \mathbb{E}[\log P(Y, Z)|Y] - \frac{1}{2}\left\{\frac{K(K+1)}{2}\log\frac{n(n-1)}{2} - (K-1)\log n\right\}.$$

- In the Bayesian setting, exact versions of BIC and ICL criteria can be calculated as

$$\log P(Y, K), \qquad \log P(Y, Z, K).$$

(up to the variational approximation) [Latouche *et al.* (2012)]

Model choice

Model selection. The number of classes K generally needs to be estimated.

- In the frequentist setting, an approximate ICL criterion can be derived

$$ICL = \mathbb{E}[\log P(Y, Z)|Y] - \frac{1}{2}\left\{\frac{K(K+1)}{2}\log\frac{n(n-1)}{2} - (K-1)\log n\right\}.$$

- In the Bayesian setting, exact versions of BIC and ICL criteria can be calculated as

$$\log P(Y, K), \qquad \log P(Y, Z, K).$$

(up to the variational approximation) [Latouche *et al.* (2012)]

But, in some applications, it may be useless or meaningless and model averaging may be preferred.

Bayesian model averaging (BMA)

General principle. [Hoeting *et al.* (1999)]

- Δ: a parameter that can be defined under a series of different models $\{\mathcal{M}_K\}_K$.
- Denote $P_K(\Delta|Y)$ its posterior distribution under model $\mathcal{M}_K$.

The posterior distribution of Δ can be averaged over all models as

$$P(\Delta|Y) = \sum_K P(\mathcal{M}_K|Y) P_K(\Delta|Y)$$

Bayesian model averaging (BMA)

General principle. [Hoeting *et al.* (1999)]

- Δ: a parameter that can be defined under a series of different models $\{\mathcal{M}_K\}_K$.
- Denote $P_K(\Delta|Y)$ its posterior distribution under model $\mathcal{M}_K$.

The posterior distribution of Δ can be averaged over all models as

$$P(\Delta|Y) = \sum_K P(\mathcal{M}_K|Y) P_K(\Delta|Y)$$

Remarks.

- $w_k = P(\mathcal{M}_K|Y)$: weight given to model $\mathcal{M}_K$ for the estimation of Δ.
- Calculating of w_K is not easy, but variational approximation may help.

Variational Bayesian model averaging [Volant *et al.* (2012)]

Variational Bayes formulation.

- $\mathcal{M}_K$ can be viewed as one more hidden layer
- Variational Bayes then aims at finding

$$Q^*(K, \theta, Z) = \arg\min_{Q \in \mathcal{Q}} KL\left[Q(K, \theta, Z) || P(K, \theta, Z|Y)\right]$$

with $\mathcal{Q} = \{Q(\theta, Z) = Q_\theta(\theta|K)Q_Z(Z|K)Q_K(K)\}$[1]

[1] No additional approximation w.r.t. the regular VBEM.

Variational Bayesian model averaging [Volant *et al.* (2012)]

Variational Bayes formulation.

- $\mathcal{M}_K$ can be viewed as one more hidden layer
- Variational Bayes then aims at finding

$$Q^*(K, \theta, Z) = \arg\min_{Q \in \mathcal{Q}} KL\left[Q(K, \theta, Z) || P(K, \theta, Z|Y)\right]$$

with $\mathcal{Q} = \{Q(\theta, Z) = Q_\theta(\theta|K) Q_Z(Z|K) Q_K(K)\}$[1]

Optimal variational weights:

$$\begin{aligned} Q_K^*(K) &\propto P(K) \exp\{\log P(Y|K) - KL[Q^*(Z, \theta|K); P(Z, \theta|Y, K)]\} \\ &= P(K|Y) e^{-KL[Q^*(Z,\theta|K); P(Z,\theta|Y,K)]}. \end{aligned}$$

[1] No additional approximation w.r.t. the regular VBEM.

VBMA: the recipe

For $K = 1 \dots K_{\max}$

- Use regular VBEM to compute the approximate conditional posterior

$$Q^*_{Z,\theta|K}(Z,\theta|K) = Q^*_{Z|K}(Z|K)Q^*_{\theta|K}(\theta|K);$$

- Compute the conditional lower bound of the log-likelihood

$$L_K = \log P(Y|K) - KL[Q^*(Z,\theta|K); P(Z,\theta|Y,K)].$$

VBMA: the recipe

For $K = 1 \dots K_{\max}$

- Use regular VBEM to compute the approximate conditional posterior

$$Q^*_{Z,\theta|K}(Z,\theta|K) = Q^*_{Z|K}(Z|K) Q^*_{\theta|K}(\theta|K);$$

- Compute the conditional lower bound of the log-likelihood

$$L_K = \log P(Y|K) - KL[Q^*(Z,\theta|K); P(Z,\theta|Y,K)].$$

Compute the approximate posterior of model $\mathcal{M}_K$:

$$w_K := Q_K(K) \propto P(K) \exp(L_K).$$

VBMA: the recipe

For $K = 1 \dots K_{\max}$

- Use regular VBEM to compute the approximate conditional posterior

$$Q^*_{Z,\theta|K}(Z,\theta|K) = Q^*_{Z|K}(Z|K) Q^*_{\theta|K}(\theta|K);$$

- Compute the conditional lower bound of the log-likelihood

$$L_K = \log P(Y|K) - KL[Q^*(Z,\theta|K); P(Z,\theta|Y,K)].$$

Compute the approximate posterior of model $\mathcal{M}_K$:

$$w_K := Q_K(K) \propto P(K) \exp(L_K).$$

Deduce the approximate posterior of Δ:

$$P(\Delta|Y) \approx \sum_K w_K Q^*_{\theta|K}(\Delta|K)$$

Towards W-graphs

W-graph model

W graph.
Latent variables:

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

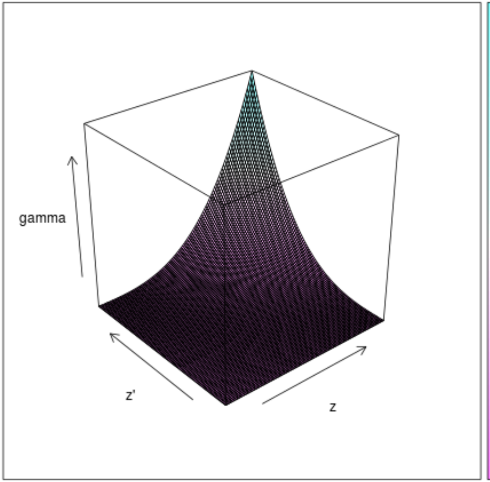
Graphon function γ:

$$\gamma(z, z') : [0,1]^2 \to [0,1]$$

Edges:

$$\Pr\{Y_{ij} = 1\} = \gamma(Z_i, Z_j)$$

Graphon function $\gamma(z, z')$



Inference of the graphon function

Probabilistic point of view.

- W-graph have been mostly studied in the probability literature: [Lovász and Szegedy (2006)], [Diaconis and Janson (2008)]
- Motif (sub-graph) frequencies are invariant characteristics of a W-graph.
- Intrinsic un-identifiability of the graphon function γ is often overcome by imposing that $u \mapsto \int \gamma(u, v) \, \mathrm{d}v$ is monotonous increasing.

Inference of the graphon function

Probabilistic point of view.

- W-graph have been mostly studied in the probability literature: [Lovász and Szegedy (2006)], [Diaconis and Janson (2008)]
- Motif (sub-graph) frequencies are invariant characteristics of a W-graph.
- Intrinsic un-identifiability of the graphon function γ is often overcome by imposing that $u \mapsto \int \gamma(u, v) \, \mathrm{d}v$ is monotonous increasing.

Statistical point of view.

- Not much attention has been paid to its inference until very recently: [Chatterjee (2012)], [Airoldi *et al.* (2013)], ...
- The latter also uses SBM as a proxy for W-graph.

SBM as a W-graph model

Latent variables:

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

$$Z_{ik} = \mathbb{I}\{\sigma_{k-1} \leq U_i < \sigma_k\}$$

where $\sigma_k = \sum_{\ell=1}^{k} \pi_\ell$.

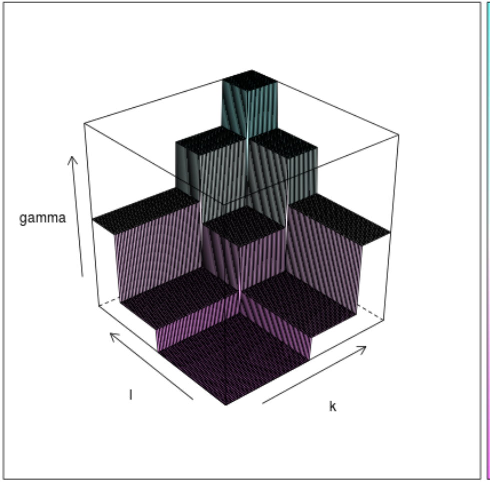
Blockwise constant graphon:

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

Edges:

$$\Pr\{Y_{ij} = 1\} = \gamma(Z_i, Z_j)$$

Graphon function $\gamma_K^{SBM}(z, z')$



Variational Bayes estimation of $\gamma(z, z')$ [Latouche and R. (2013)]

Posterior mean of $\gamma_K^{SBM}(z, z')$

VBEM inference provides the approximat posteriors:

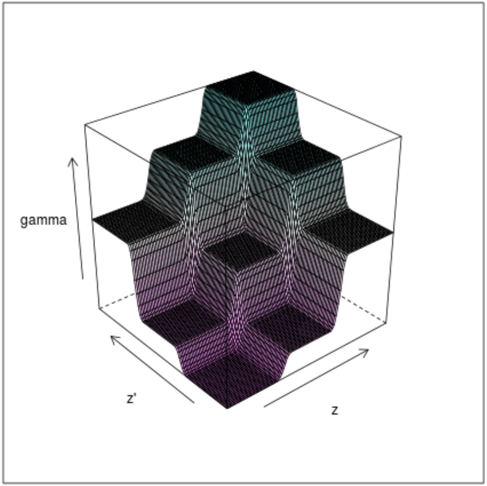
$$\begin{aligned}(\pi|Y) &\approx \text{Dir}(\pi^*) \\ (\gamma_{k\ell}|Y) &\approx \text{Beta}(\gamma_{k\ell}^{0*}, \gamma_{k\ell}^{1*})\end{aligned}$$

Estimate of $\gamma(u, v)$.

$$\widehat{\gamma}_K^{SBM}(u, v) = \widetilde{\mathbb{E}}\left(\gamma_{C(u),C(v)}|Y\right)$$

where $C(u) = 1 + \sum_k \mathbb{I}\{\sigma_k \leq u\}$.

[Gouda and Szántai (2010)]



Model averaging

Model averaging: There is no 'true K' in the W-graph model.

Apply VBMA recipe. For $K = 1..K_{\max}$, fit an SBM model via VBEM and compute

$$\widehat{\gamma}_K^{SBM}(z, z') = \mathbb{E}_Q[\gamma_{C(z), C(z')}].$$

Model averaging

Model averaging: There is no 'true K' in the W-graph model.

Apply VBMA recipe. For $K = 1..K_{\max}$, fit an SBM model via VBEM and compute

$$\widehat{\gamma}_K^{SBM}(z, z') = \mathbb{E}_Q[\gamma_{C(z),C(z')}].$$

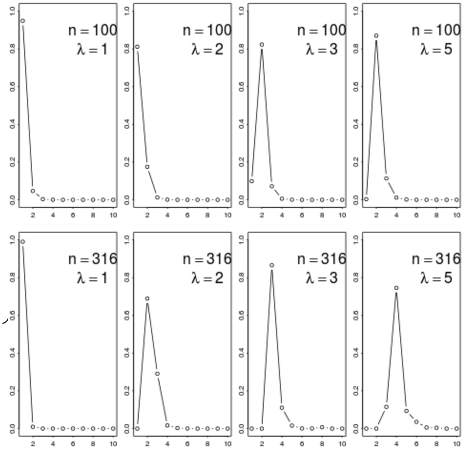
Perform model averaging as

$$\widehat{\gamma}(z, z') = \sum_K w_K \widehat{\gamma}_K^{SBM}(z, z')$$

where w_K is the variational weights arising from variational Bayes inference.

Some simulations

Variational posterior for K: $Q^*(K)$.



Design. Symetric graphon:

$$\gamma(u, v) = \rho\lambda^2(uv)^{\lambda-1}$$

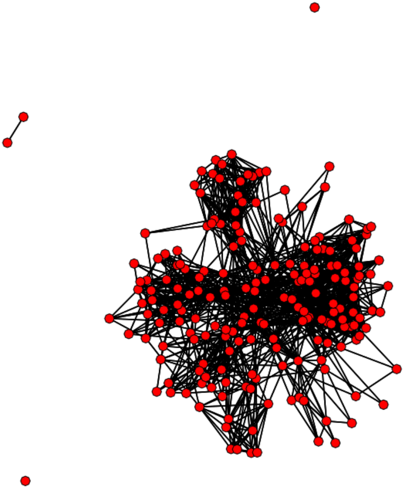
- $\lambda \uparrow$: imbalanced graph
- $\rho \uparrow$: dense graph

Results.

- More complex models as n and λ ↗
- Posterior fairly concentrated

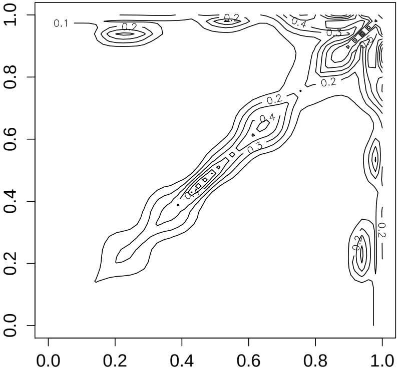
French political blogosphere

Website network. French political blogs: 196 nodes, 1432 edges.



French political blogosphere

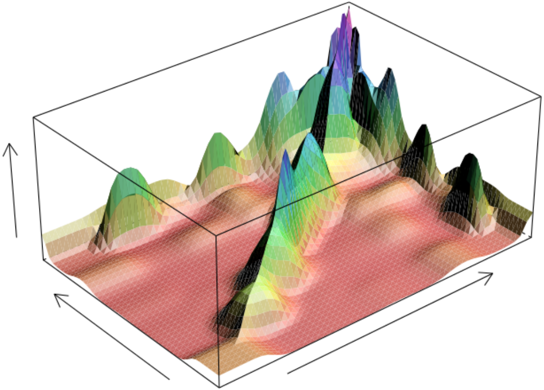
Infered graphon. $\widehat{W}(u,v) = \mathbb{E}(\gamma(u,v)|Y)$



Motif probability can be estimated as well as $\widehat{\mu}(m) = \mathbb{E}(\mu(m)|Y)$.

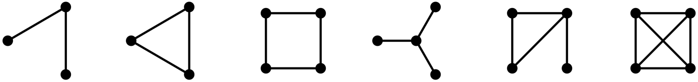
French political blogosphere

Infered graphon. $\widehat{W}(u,v) = \mathbb{E}(\gamma(u,v)|Y)$



Motif probability can be estimated as well as $\widehat{\mu}(m) = \mathbb{E}(\mu(m)|Y)$.

Goodness-of-fit using network motifs



- Network motifs have sociological interpretation (e.g. triangles).
- The first moments $\mathbb{E}N(m), \mathbb{V}N(m)$ of the count are known under SBM [Picard *et al.* (2008)]:

$$\mathbb{E}_{SBM}N(m) \propto \mu_{SBM}(m) = f(\theta_{SBM})$$

- Motif probability can be estimated as

$$\widehat{\mu}(m) = \sum_{k} Q^*_K(K)\widetilde{\mathbb{E}}(\mu_{SBM}(m)|X, K)$$

$\rightarrow$ Goodness of fit criterion

Network motifs in the blogosphere

Motif	Count ($\times 10^3$)	Mean ($\times 10^3$)	Std. dev. ($\times 10^3$)	approx p-value
	29.7	39.7	8.3	0.89
	3.8	4.6	1.3	0.69
	608.7	968.3	336.8	0.86
	279.8	428.9	154.0	0.83
	47.4	74.5	35.1	0.77
	270.5	397.0	177.0	0.75
	62.1	87.8	47.4	0.67
	6.5	8.8	5.4	0.61

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

Conclusion

- Real networks are heterogeneous.

Conclusion

- Real networks are heterogeneous.
- State space models (including SBM) allow to capture such an heterogeneity ... but raise inference problems.

Conclusion

- Real networks are heterogeneous.
- State space models (including SBM) allow to capture such an heterogeneity ... but raise inference problems.
- Variational approximations help to deal with complex (conditional) dependency structures.

Conclusion

- Real networks are heterogeneous.
- State space models (including SBM) allow to capture such an heterogeneity ... but raise inference problems.
- Variational approximations help to deal with complex (conditional) dependency structures.
- SBM is a special (rare?) case for which guaranties exist as for the performances of the variational approximation.

Conclusion

- Real networks are heterogeneous.
- State space models (including SBM) allow to capture such an heterogeneity ... but raise inference problems.
- Variational approximations help to deal with complex (conditional) dependency structures.
- SBM is a special (rare?) case for which guaranties exist as for the performances of the variational approximation.
- SBM can be easily generalized to handle weighted graphs and covariates.

Conclusion

- Real networks are heterogeneous.
- State space models (including SBM) allow to capture such an heterogeneity ... but raise inference problems.
- Variational approximations help to deal with complex (conditional) dependency structures.
- SBM is a special (rare?) case for which guaranties exist as for the performances of the variational approximation.
- SBM can be easily generalized to handle weighted graphs and covariates.
- SBM can be used as a proxy for smoother models, such as W-graphs.

Airoldi, E. M., Costa, T. B. and Chan, S. H. (2013). Stochastic blockmodel approximation of a graphon: Theory and consistent estimation. In *Advances in Neural Information Processing Systems*, 692–700.

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

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

Celisse, A., Daudin, J.-J. and Pierre, L. (2012). Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electron. J. Statis.* **6** 1847–99.

Chatterjee, S. (2012), Matrix estimation by Universal Singular Value Thresholding. Technical report, arXiv:1212.1247.

Daudin, J.-J., Picard, F. and Robin, S. (Jun, 2008). A mixture model for random graphs. *Stat. Comput.* **18 (2)** 173–83.

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

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. B.* **39** 1–38.

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

Gazal, S., Daudin, J.-J. and Robin, S. (2012). Accuracy of variational estimates for random graph mixture models. *Journal of Statistical Computation and Simulation.* **82 (6)** 849–862.

Girvan, M. and Newman, M. E. J. (2002). Community strucutre in social and biological networks. *Proc. Natl. Acad. Sci. USA.* **99 (12)** 7821–6.

Gouda, A. and Szántai, T. (2010). On numerical calculation of probabilities according to Dirichlet distribution. *Ann. Oper. Res.* **177** 185–200. DOI: 10.1007/s10479-009-0601-9.

Handcock, M., Raftery, A. and Tantrum, J. (2007). Model-based clustering for social networks. *JRSSA.* **170 (2)** 301–54. `doi: 10.1111/j.1467-985X.2007.00471.x`.

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

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

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

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

Latouche, P. and Robin, S. (2013), 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.

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

von Luxburg, U., Belkin, M. and Bousquet, O. (2008). Consistency of spectral clustering. *Ann. Stat.* **36 (2)** 555–586.

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

Mariadassou, M. and Matias, C. (2015). Convergence of the groups posterior distribution in latent or stochastic block models. *Bernoulli*. ??–?? to appear.

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

Newman, M. and Girvan, M. (2004). Finding and evaluating community structure in networks,. *Phys. Rev. E.* **69** 026113.

Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks. *Phys. Rev.* **E (69)** 066133.

Nowicki, K. and Snijders, T. (2001). Estimation and prediction for stochastic block-structures. *J. Amer. Statist. Assoc.* **96** 1077–87.

Parisi, G. (1988). *Statistical Field Theory*. Addison Wesley, New York),.

Picard, F., Daudin, J.-J., Koskas, M., Schbath, S. and Robin, S. (2008). Assessing the exceptionality of network motifs,. *J. Comp. Biol.* **15 (1)** 1–20.

Picard, F., Miele, V., Daudin, J.-J., Cottret, L. and Robin, S. (2009). Deciphering the connectivity structure of biological networks using mixnet. *BMC Bioinformatics*. **Suppl 6** S17. doi:10.1186/1471-2105-10-S6-S17.

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

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

WAINWRIGHT, M. J. and JORDAN, M. I. (2008). Graphical models, exponential families, and variational inference. *Found. Trends Mach. Learn.* **1 (1–2)** 1–305. http:/dx.doi.org/10.1561/2200000001.