Variational inference for the Stochastic Block-Model

S. Robin

AgroParisTech / INRA

Workshop on Random Graphs, April 2011, Lille

Stochastic block model (SBM)

Modelling network heterogeneity

Latent variable models allow to capture the underlying structure of a network.

Modelling network heterogeneity

Latent variable models allow to capture the underlying structure of a network.

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

Modelling network heterogeneity

Latent variable models allow to capture the underlying structure of a network.

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

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

$$\{Z_i\} \text{ i.i.d. } \sim \pi$$

- the edges X_{ij} are independent conditionally to the Z_i's:

$$\{X_{ij}\} \text{ independent } |\{Z_i\} : X_{ij} \sim \mathcal{B}[\gamma(Z_i, Z_j)]$$

Modelling network heterogeneity

Latent variable models allow to capture the underlying structure of a network.

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

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

$$\{Z_i\} \text{ i.i.d. } \sim \pi$$

- the edges X_{ij} are independent conditionally to the Z_i's:

$$\{X_{ij}\} \text{ independent } |\{Z_i\} : X_{ij} \sim \mathcal{B}[\gamma(Z_i, Z_j)]$$

Continuous (*Hoff* et al. *(2002)*): ($\simeq$ PCA)

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

Modelling network heterogeneity

Latent variable models allow to capture the underlying structure of a network.

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

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

$$\{Z_i\} \text{ i.i.d. } \sim \pi$$

- the edges X_{ij} are independent conditionally to the Z_i's:

$$\{X_{ij}\} \text{ independent } |\{Z_i\} : X_{ij} \sim \mathcal{B}[\gamma(Z_i, Z_j)]$$

Continuous (*Hoff* et al. *(2002)*): ($\simeq$ PCA)

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

Discrete (*Nowicki and Snijders (2001)*): ($\longrightarrow$ finite mixture = SBM)

$$Z_i \in \{1, \dots, K\}, \qquad \gamma(k, \ell) = \gamma_{k\ell}.$$

(Weighted) Stochastic Block-Model (SBM)

Discrete-valued latent labels: each node i belong to class k with probability π_k:

$$\{Z_i\}_i \text{ i.i.d.}, \qquad Z_i \sim \mathcal{M}(1; \boldsymbol{\pi})$$

where $\boldsymbol{\pi} = (\pi_1, \dots \pi_K)$.

(Weighted) Stochastic Block-Model (SBM)

Discrete-valued latent labels: each node i belong to class k with probability π_k:

$$\{Z_i\}_i \text{ i.i.d.}, \qquad Z_i \sim \mathcal{M}(1; \boldsymbol{\pi})$$

where $\boldsymbol{\pi} = (\pi_1, \dots \pi_K)$.

Observed edges: $\{X_{ij}\}_{i,j}$ are conditionally independent given the Z_i's:

$$(X_{ij} \mid Z_i = k, Z_j = \ell) \sim f_{k\ell}(\cdot)$$

where $f_{k\ell}(\cdot)$ is some parametric distribution $f_{k\ell}(x) = f(x; \gamma_{k\ell})$, e.g.

$$(X_{ij} \mid Z_i = k, Z_j = \ell) \sim \mathcal{B}(\gamma_{k\ell}) \qquad \text{(binary graph)}$$

We denote $\boldsymbol{\gamma} = \{\gamma_{k\ell}\}_{k,\ell}$.

(Weighted) Stochastic Block-Model (SBM)

Discrete-valued latent labels: each node i belong to class k with probability π_k:

$$\{Z_i\}_i \text{ i.i.d.}, \qquad Z_i \sim \mathcal{M}(1; \boldsymbol{\pi})$$

where $\boldsymbol{\pi} = (\pi_1, \dots \pi_K)$.

Observed edges: $\{X_{ij}\}_{i,j}$ are conditionally independent given the Z_i's:

$$(X_{ij} \mid Z_i = k, Z_j = \ell) \sim f_{k\ell}(\cdot)$$

where $f_{k\ell}(\cdot)$ is some parametric distribution $f_{k\ell}(x) = f(x; \gamma_{k\ell})$, e.g.

$$(X_{ij} \mid Z_i = k, Z_j = \ell) \sim \mathcal{B}(\gamma_{k\ell}) \qquad \text{(binary graph)}$$

We denote $\boldsymbol{\gamma} = \{\gamma_{k\ell}\}_{k,\ell}$.

Statistical inference: We want to estimate

$$\boldsymbol{\theta} = (\boldsymbol{\pi}, \boldsymbol{\gamma}) \qquad \text{and} \qquad P(\mathbf{Z}|\mathbf{X}).$$

Variational inference

Maximum likelihood inference

Maximum likelihood estimate: We are looking for

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \log P(\mathbf{X}; \boldsymbol{\theta})$$

but $P(\mathbf{X}; \boldsymbol{\theta}) = \sum_{\mathbf{Z}} P(\mathbf{X}, \mathbf{Z}; \boldsymbol{\theta})$ is not tractable.

Maximum likelihood inference

Maximum likelihood estimate: We are looking for

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \log P(\mathbf{X}; \boldsymbol{\theta})$$

but $P(\mathbf{X}; \boldsymbol{\theta}) = \sum_{\mathbf{Z}} P(\mathbf{X}, \mathbf{Z}; \boldsymbol{\theta})$ is not tractable.

Classical strategy: Based on the decomposition

$$\log P(\mathbf{X}) = \mathbb{E}[\log P(\mathbf{X}, \mathbf{Z})|\mathbf{X}] - \mathbb{E}[\log P(\mathbf{Z}|\mathbf{X})|\mathbf{X}],$$

Maximum likelihood inference

Maximum likelihood estimate: We are looking for

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \log P(\mathbf{X}; \boldsymbol{\theta})$$

but $P(\mathbf{X}; \boldsymbol{\theta}) = \sum_{\mathbf{Z}} P(\mathbf{X}, \mathbf{Z}; \boldsymbol{\theta})$ is not tractable.

Classical strategy: Based on the decomposition

$$\log P(\mathbf{X}) = \mathbb{E}[\log P(\mathbf{X}, \mathbf{Z}) | \mathbf{X}] - \mathbb{E}[\log P(\mathbf{Z}|\mathbf{X}) | \mathbf{X}],$$

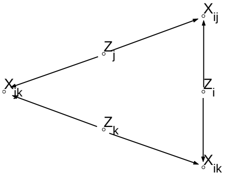
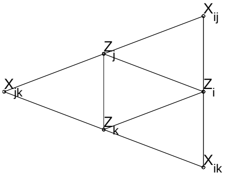
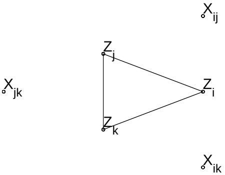
the EM algorithm aims at retrieving the maximum likelihood estimates via the alternation of 2 steps.

E-step: calculation of $P(\mathbf{Z}|\mathbf{X}; \widehat{\boldsymbol{\theta}})$.

M-step: maximisation of $\mathbb{E}[\log P(\mathbf{X}, \mathbf{Z}; \boldsymbol{\theta}) | \mathbf{X}]$ in $\boldsymbol{\theta}$.

Case of the Stochastic Block-Model

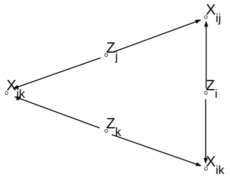
Dependency structure.

Dependecy graph: $P(\mathbf{Z})P(\mathbf{X}\|\mathbf{Z})$	Moral graph (*Lauritzen (1996)*)	Conditional dep.: $P(\mathbf{Z}\|\mathbf{X})$
		

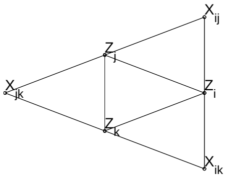
Case of the Stochastic Block-Model

Dependency structure.

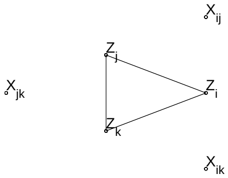
Dependecy graph: $P(\mathbf{Z})P(\mathbf{X}|\mathbf{Z})$



Moral graph (*Lauritzen (1996)*)



Conditional dep.: $P(\mathbf{Z}|\mathbf{X})$



The conditional dependency graph of $\mathbf{Z}$ is a clique
$\longrightarrow$ no factorisation can be hoped to calculate $P(\mathbf{Z}|\mathbf{X})$ (unlike hidden Markov random fields).

Variational approximation

As $P(\mathbf{Z}|\mathbf{X})$ can not be calculated, we need to find some approximate distribution $Q(\mathbf{Z})$.

Variational approximation

As $P(\mathbf{Z}|\mathbf{X})$ can not be calculated, we need to find some approximate distribution $Q(\mathbf{Z})$.

Lower bound of the log-likelihood: For any distribution $Q(\mathbf{Z})$, we have (*Jaakkola (2000)*,*Wainwright and Jordan (2008)*)

$$\log P(\mathbf{X}) \geq \log P(\mathbf{X}) - KL[Q(\mathbf{Z}); P(\mathbf{Z}|\mathbf{X})]$$

Variational approximation

As $P(\mathbf{Z}|\mathbf{X})$ can not be calculated, we need to find some approximate distribution $Q(\mathbf{Z})$.

Lower bound of the log-likelihood: For any distribution $Q(\mathbf{Z})$, we have (*Jaakkola (2000)*,*Wainwright and Jordan (2008)*)

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

where $\mathcal{H}(Q)$ is the entropy of Q: $\mathcal{H}(Q) = -\mathbb{E}_Q[\log Q(\mathbf{Z})]$.

Variational approximation

As $P(\mathbf{Z}|\mathbf{X})$ can not be calculated, we need to find some approximate distribution $Q(\mathbf{Z})$.

Lower bound of the log-likelihood: For any distribution $Q(\mathbf{Z})$, we have (*Jaakkola (2000)*,*Wainwright and Jordan (2008)*)

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

where $\mathcal{H}(Q)$ is the entropy of Q: $\mathcal{H}(Q) = -\mathbb{E}_Q[\log Q(\mathbf{Z})]$.

This amounts to replace $P(\cdot|\mathbf{X})$ with $Q(\cdot)$ in

$$\log P(\mathbf{X}) = \mathbb{E}_{P(\cdot|\mathbf{X})}[\log P(\mathbf{X}, \mathbf{Z})] + \mathcal{H}[P(\cdot|\mathbf{X})].$$

Variational EM

'Expectation' step (pseudo E-step): find the best lower bound of $\log P(\mathbf{X})$, i.e. the best approximation of $P(\cdot|\mathbf{X})$ as

$$Q^* = \arg\min_{Q \in \mathcal{Q}} KL[Q(\mathbf{Z}); P(\mathbf{Z}|\mathbf{X})]$$

where $\mathcal{Q}$ is a class of 'manageable' distributions.

Variational EM

'Expectation' step (pseudo E-step): find the best lower bound of $\log P(\mathbf{X})$, i.e. the best approximation of $P(\cdot|\mathbf{X})$ as

$$Q^* = \arg\min_{Q \in \mathcal{Q}} KL[Q(\mathbf{Z}); P(\mathbf{Z}|\mathbf{X})]$$

where $\mathcal{Q}$ is a class of 'manageable' distributions.

Maximisation step (M-step): estimate $\boldsymbol{\theta}$ as

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

which maximises the lower bound of $\log P(\mathbf{X})$.

Approximation of $P(\mathbf{Z}|\mathbf{X})$ for SBM

We are looking for

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

Approximation of $P(\mathbf{Z}|\mathbf{X})$ for SBM

We are looking for

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

- We restrict ourselves to the set of factorisable distributions:

$$\mathcal{Q} = \left\{ Q : Q(\mathbf{Z}) = \prod_i Q_i(Z_i) = \prod_i \prod_k \tau_{ik}^{Z_{ik}} \right\}, \quad \tau_{ik} \approx \Pr\{Z_i = k|\mathbf{X}\}.$$

Approximation of $P(\mathbf{Z}|\mathbf{X})$ for SBM

We are looking for

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

- We restrict ourselves to the set of factorisable distributions:

$$\mathcal{Q} = \left\{ Q : Q(\mathbf{Z}) = \prod_i Q_i(Z_i) = \prod_i \prod_k \tau_{ik}^{Z_{ik}} \right\}, \quad \tau_{ik} \approx \Pr\{Z_i = k|\mathbf{X}\}.$$

- The optimal τ_{ik}^*'s satisfy the fix-point relation:

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

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

Application to a regulatory network

Regulatory network = directed graph
where

- Nodes = genes (or groups of genes, e.g. operons)
- Edges = regulations:

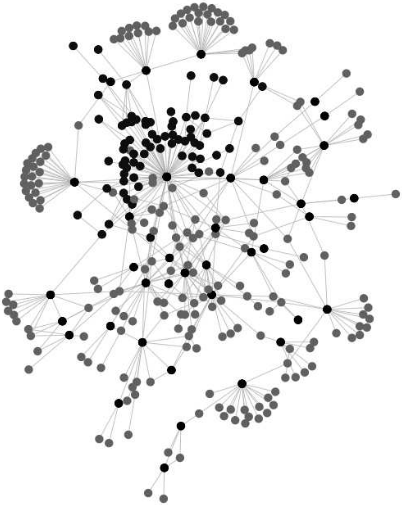
$$\{i \rightarrow j\} \quad \Leftrightarrow \quad i \text{ regulates } j$$

Application to a regulatory network

Regulatory network = directed graph where

- Nodes = genes (or groups of genes, e.g. operons)
- Edges = regulations:

$$\{i \to j\} \quad \Leftrightarrow \quad i \text{ regulates } j$$



Application to a regulatory network

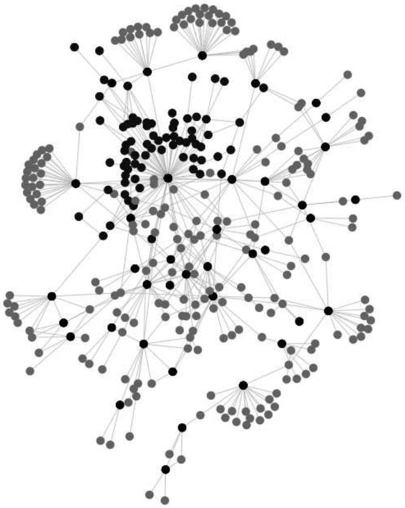
Regulatory network = directed graph where

- Nodes = genes (or groups of genes, e.g. operons)
- Edges = regulations:

$$\{i \rightarrow j\} \quad \Leftrightarrow \quad i \text{ regulates } j$$

Questions

- Do some nodes share similar connexion profiles?
- Is there a 'macroscopic' organisation of the network?



SBM analysis

Parameter estimates. $K = 5$

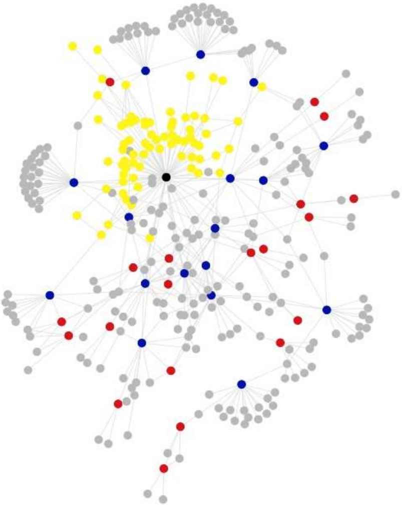
$\widehat{\gamma}_{k\ell}$ (%)	1	2	3	4	5
1	.	.	.	.	.
2	6.40	1.50	1.34	.	.
3	1.21	.	.	.	.
4	.	.	.	.	.
5	8.64	17.65	.	72.87	11.01
$\widehat{\pi}$ (%)	65.49	5.18	7.92	21.10	0.30

Picard et al. *(2009)*

SBM analysis

Parameter estimates. $K = 5$

$\widehat{\gamma}_{k\ell}$ (%)	1	2	3	4	5
1	.	.	.	.	.
2	6.40	1.50	1.34	.	.
3	1.21	.	.	.	.
4	.	.	.	.	.
5	8.64	17.65	.	72.87	11.01
$\widehat{\pi}$ (%)	65.49	5.18	7.92	21.10	0.30



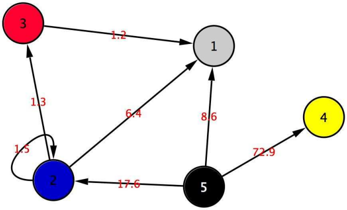
Picard et al. *(2009)*

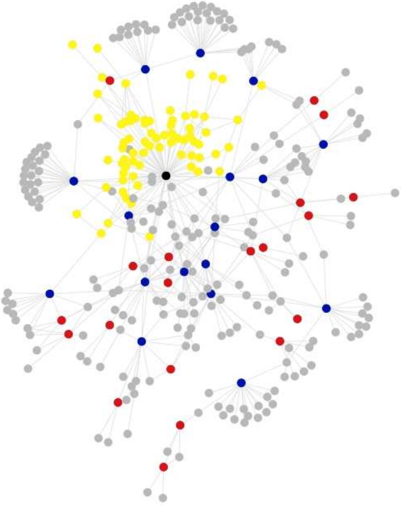
SBM analysis

Parameter estimates. $K = 5$

$\widehat{\gamma}_{k\ell}$ (%)	1	2	3	4	5
1	.	.	.	.	.
2	6.40	1.50	1.34	.	.
3	1.21	.	.	.	.
4	.	.	.	.	.
5	8.64	17.65	.	72.87	11.01
$\widehat{\pi}$ (%)	65.49	5.18	7.92	21.10	0.30

Meta-graph representation.





Picard et al. *(2009)*

Properties of variational inference

The quality of the inference based on the variational approximation is not very well known yet.

Properties of variational inference

The quality of the inference based on the variational approximation is not very well known yet.

Negative result: (*Gunawardana and Byrne (2005)*) The VEM algorithm converges to a different optimum than ML in the general case, except for 'degenerated' models.

Properties of variational inference

The quality of the inference based on the variational approximation is not very well known yet.

Negative result: (*Gunawardana and Byrne (2005)*) The VEM algorithm converges to a different optimum than ML in the general case, except for 'degenerated' models.

Specific case of graphs.

- Specific asymptotic framework: n^2 data, $'p' = n$ 'variables' per individual.
- Mean field approximation is asymptotically exact for some models with infinite range dependency (*Opper and Winther (2001)*: law of large number argument).

Concentration of $P(\mathbf{Z}|\mathbf{X})$ for binary graphs

Let us denote g, the conditional distribution

$$g(\mathbf{z}; \mathbf{X}) := \Pr\{\mathbf{Z} = \mathbf{z}|\mathbf{X}\} = \frac{1}{C} \prod_i \pi_{Z_i} \prod_{j \neq i} \gamma_{Z_i Z_j}^{X_{ij}} [1 - \gamma_{Z_i Z_j}]^{1 - X_{ij}}$$

Concentration of $P(\mathbf{Z}|\mathbf{X})$ for binary graphs

Let us denote g, the conditional distribution

$$g(\mathbf{z};\mathbf{X}) := \Pr\{\mathbf{Z}=\mathbf{z}|\mathbf{X}\} = \frac{1}{C}\prod_i \pi_{Z_i} \prod_{j\neq i} \gamma_{Z_i Z_j}^{X_{ij}} [1-\gamma_{Z_i Z_j}]^{1-X_{ij}}$$

Theorem (*Célisse & al. (2011)*). Under identifiability conditions and if $\forall k, \ell : 0 < a < \gamma_{k\ell} < 1-a, 0 < b < \pi_k$, then we have

$$\forall t > 0, \qquad \Pr\left\{ \frac{\sum_{\mathbf{z}\neq\mathbf{z}^*} g(\mathbf{z};\mathbf{X})}{g(\mathbf{z}^*;\mathbf{X})} > t \middle| \mathbf{Z}=\mathbf{z}^* \right\} = \mathcal{O}(ne^{-\kappa(t)n}).$$

Concentration of $P(\mathbf{Z}|\mathbf{X})$ for binary graphs

Let us denote g, the conditional distribution

$$g(\mathbf{z}; \mathbf{X}) := \Pr\{\mathbf{Z} = \mathbf{z}|\mathbf{X}\} = \frac{1}{C} \prod_i \pi_{Z_i} \prod_{j \neq i} \gamma_{Z_i Z_j}^{X_{ij}} [1 - \gamma_{Z_i Z_j}]^{1 - X_{ij}}$$

Theorem (*Célisse & al. (2011)*). Under identifiability conditions and if $\forall k, \ell : 0 < a < \gamma_{k\ell} < 1 - a, 0 < b < \pi_k$, then we have

$$\forall t > 0, \qquad \Pr\left\{ \frac{\sum_{\mathbf{z} \neq \mathbf{z}^*} g(\mathbf{z}; \mathbf{X})}{g(\mathbf{z}^*; \mathbf{X})} > t \middle| \mathbf{Z} = \mathbf{z}^* \right\} = \mathcal{O}(ne^{-\kappa(t)n}).$$

$\longrightarrow$ If the true labels are $\mathbf{z}^*$, then $P(\mathbf{Z}|\mathbf{X})$ concentrates around $\mathbf{z}^*$.

Concentration of $P(\mathbf{Z}|\mathbf{X})$ for binary graphs

Let us denote g, the conditional distribution

$$g(\mathbf{z}; \mathbf{X}) := \Pr\{\mathbf{Z} = \mathbf{z}|\mathbf{X}\} = \frac{1}{C} \prod_i \pi_{Z_i} \prod_{j \neq i} \gamma_{Z_i Z_j}^{X_{ij}} [1 - \gamma_{Z_i Z_j}]^{1 - X_{ij}}$$

Theorem (*Célisse & al. (2011)*). Under identifiability conditions and if $\forall k, \ell : 0 < a < \gamma_{k\ell} < 1 - a, 0 < b < \pi_k$, then we have

$$\forall t > 0, \qquad \Pr\left\{ \frac{\sum_{\mathbf{z} \neq \mathbf{z}^*} g(\mathbf{z}; \mathbf{X})}{g(\mathbf{z}^*; \mathbf{X})} > t \middle| \mathbf{Z} = \mathbf{z}^* \right\} = \mathcal{O}(ne^{-\kappa(t)n}).$$

$\longrightarrow$ If the true labels are $\mathbf{z}^*$, then $P(\mathbf{Z}|\mathbf{X})$ concentrates around $\mathbf{z}^*$.
$\longrightarrow$ SBM is a 'degenerated' model.

Concentration of $P(\mathbf{Z}|\mathbf{X})$ for binary graphs

Let us denote g, the conditional distribution

$$g(\mathbf{z};\mathbf{X}) := \Pr\{\mathbf{Z}=\mathbf{z}|\mathbf{X}\} = \frac{1}{C}\prod_i \pi_{Z_i} \prod_{j\neq i} \gamma_{Z_iZ_j}^{X_{ij}}[1-\gamma_{Z_iZ_j}]^{1-X_{ij}}$$

Theorem (*Célisse & al. (2011)*). Under identifiability conditions and if $\forall k,\ell: 0 < a < \gamma_{k\ell} < 1-a, 0 < b < \pi_k$, then we have

$$\forall t > 0, \qquad \Pr\left\{ \frac{\sum_{\mathbf{z}\neq\mathbf{z}^*} g(\mathbf{z};\mathbf{X})}{g(\mathbf{z}^*;\mathbf{X})} > t \middle| \mathbf{Z}=\mathbf{z}^* \right\} = \mathcal{O}(ne^{-\kappa(t)n}).$$

$\longrightarrow$ If the true labels are $\mathbf{z}^*$, then $P(\mathbf{Z}|\mathbf{X})$ concentrates around $\mathbf{z}^*$.
$\longrightarrow$ SBM is a 'degenerated' model.

Ongoing work about the convergence $P(\cdot|\mathbf{X}) \rightarrow \delta\{\mathbf{z}^0\}$ (*Matias (2011)*).

Concentration of the degree distribution

Binary graph. Binomial distribution of the degrees

$$K_i | (i \in q) \sim \mathcal{B}(n-1, \overline{\gamma}_k)$$

where $\overline{\gamma}_k = \sum_\ell \pi_\ell \gamma_{k,\ell}$.

Concentration of the degree distribution

Binary graph. Binomial distribution of the degrees

$$K_i | (i \in q) \sim \mathcal{B}(n-1, \overline{\gamma}_k)$$

where $\overline{\gamma}_k = \sum_\ell \pi_\ell \gamma_{k,\ell}$.

Normalised degree: $D_i = K_i/(n-1)$ concentrates around $\overline{\gamma}_k$.

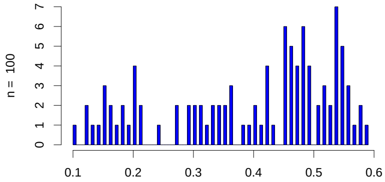
Concentration of the degree distribution

Binary graph. Binomial distribution of the degrees

$$K_i|(i \in q) \sim \mathcal{B}(n-1, \overline{\gamma}_k)$$

where $\overline{\gamma}_k = \sum_\ell \pi_\ell \gamma_{k,\ell}$.

Normalised degree: $D_i = K_i/(n-1)$ concentrates around $\overline{\gamma}_k$.



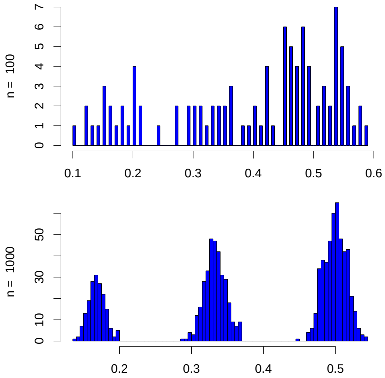
Concentration of the degree distribution

Binary graph. Binomial distribution of the degrees

$$K_i | (i \in q) \sim \mathcal{B}(n-1, \overline{\gamma}_k)$$

where $\overline{\gamma}_k = \sum_\ell \pi_\ell \gamma_{k,\ell}$.

Normalised degree: $D_i = K_i/(n-1)$ concentrates around $\overline{\gamma}_k$.



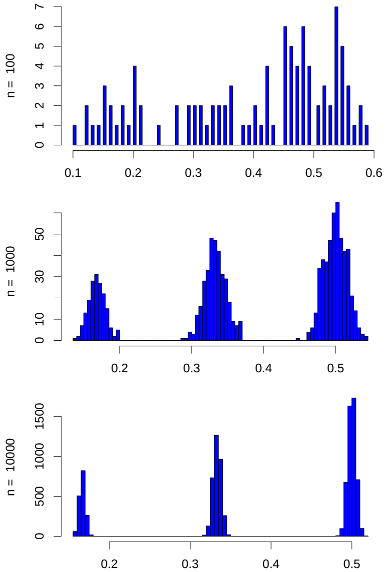
Concentration of the degree distribution

Binary graph. Binomial distribution of the degrees

$$K_i | (i \in q) \sim \mathcal{B}(n-1, \overline{\gamma}_k)$$

where $\overline{\gamma}_k = \sum_\ell \pi_\ell \gamma_{k,\ell}$.

Normalised degree: $D_i = K_i/(n-1)$ concentrates around $\overline{\gamma}_k$.



Concentration of the degree distribution

Binary graph. Binomial distribution of the degrees

$$K_i | (i \in q) \sim \mathcal{B}(n-1, \overline{\gamma}_k)$$

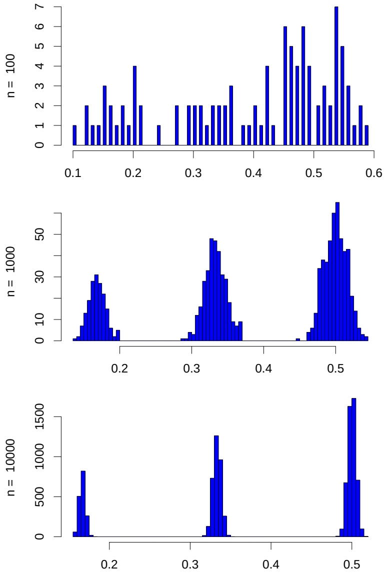
where $\overline{\gamma}_k = \sum_\ell \pi_\ell \gamma_{k,\ell}$.

Normalised degree: $D_i = K_i/(n-1)$ concentrates around $\overline{\gamma}_k$.

Linear algorithm

- based on the gaps between the ordered $D_{(i)}$,
- provides consistent estimates of π and γ can be derived.

(*Channarond (2011)*).



Variational Bayes inference

Variational Bayes inference

Bayesian setting: Both θ and $\mathbf{Z}$ are random and unobserved and we want to retrieve $P(\mathbf{Z}, \theta | \mathbf{X})$

Variational Bayes inference

Bayesian setting: Both $\boldsymbol{\theta}$ and $\mathbf{Z}$ are random and unobserved and we want to retrieve $P(\mathbf{Z}, \boldsymbol{\theta}|\mathbf{X})$ so we look for

$$Q^* = \arg\min_{Q \in \mathcal{Q}} KL[Q(\mathbf{Z}, \boldsymbol{\theta}); P(\mathbf{Z}, \boldsymbol{\theta}|\mathbf{X})]$$

within $\mathcal{Q} = \{Q : Q(\mathbf{Z}, \boldsymbol{\theta}) = Q_Z(\mathbf{Z})Q_\theta(\boldsymbol{\theta})\}$.

Variational Bayes inference

Bayesian setting: Both $\boldsymbol{\theta}$ and $\mathbf{Z}$ are random and unobserved and we want to retrieve $P(\mathbf{Z}, \boldsymbol{\theta}|\mathbf{X})$ so we look for

$$Q^* = \arg\min_{Q \in \mathcal{Q}} KL[Q(\mathbf{Z}, \boldsymbol{\theta}); P(\mathbf{Z}, \boldsymbol{\theta}|\mathbf{X})]$$

within $\mathcal{Q} = \{Q : Q(\mathbf{Z}, \boldsymbol{\theta}) = Q_Z(\mathbf{Z})Q_\theta(\boldsymbol{\theta})\}$.

VB-EM algorithm: In the exponential family / conjugate prior context

$$P(\mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}) \propto \exp\{\phi(\boldsymbol{\theta})'[u(\mathbf{X}, \mathbf{Z}) + \boldsymbol{\nu}]\}$$

Variational Bayes inference

Bayesian setting: Both $\boldsymbol{\theta}$ and $\mathbf{Z}$ are random and unobserved and we want to retrieve $P(\mathbf{Z}, \boldsymbol{\theta}|\mathbf{X})$ so we look for

$$Q^* = \arg\min_{Q \in \mathcal{Q}} KL[Q(\mathbf{Z}, \boldsymbol{\theta}); P(\mathbf{Z}, \boldsymbol{\theta}|\mathbf{X})]$$

within $\mathcal{Q} = \{Q : Q(\mathbf{Z}, \boldsymbol{\theta}) = Q_Z(\mathbf{Z})Q_\theta(\boldsymbol{\theta})\}$.

VB-EM algorithm: In the exponential family / conjugate prior context

$$P(\mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}) \propto \exp\{\phi(\boldsymbol{\theta})'[u(\mathbf{X}, \mathbf{Z}) + \boldsymbol{\nu}]\}$$

the optimal $Q^*(\mathbf{Z}, \boldsymbol{\theta})$ is recovered (*Beal and Ghahramani (2003)*) via

$$\begin{array}{lrcl} \text{pseudo-M:} & Q_{\boldsymbol{\theta}}(\boldsymbol{\theta}) & \propto & \exp\left(\phi(\boldsymbol{\theta})' \left\{\mathbb{E}_{Q_Z}[u(\mathbf{X}, \mathbf{Z})] + \boldsymbol{\nu}\right\}\right) \\ \text{pseudo-E:} & Q_Z(\mathbf{Z}) & \propto & \exp\{\mathbb{E}_{Q_\theta}[\phi(\boldsymbol{\theta})]' u(\mathbf{X}, \mathbf{Z})\} \end{array}$$

Variational Bayes inference

Bayesian setting: Both $\boldsymbol{\theta}$ and $\mathbf{Z}$ are random and unobserved and we want to retrieve $P(\mathbf{Z}, \boldsymbol{\theta}|\mathbf{X})$ so we look for

$$Q^* = \arg\min_{Q \in \mathcal{Q}} KL[Q(\mathbf{Z}, \boldsymbol{\theta}); P(\mathbf{Z}, \boldsymbol{\theta}|\mathbf{X})]$$

within $\mathcal{Q} = \{Q : Q(\mathbf{Z}, \boldsymbol{\theta}) = Q_Z(\mathbf{Z}) Q_\theta(\boldsymbol{\theta})\}$.

VB-EM algorithm: In the exponential family / conjugate prior context

$$P(\mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}) \propto \exp\{\phi(\boldsymbol{\theta})'[u(\mathbf{X}, \mathbf{Z}) + \boldsymbol{\nu}]\}$$

the optimal $Q^*(\mathbf{Z}, \boldsymbol{\theta})$ is recovered (*Beal and Ghahramani (2003)*) via

$$\begin{array}{rrcl} \text{pseudo-M:} & Q_{\boldsymbol{\theta}}(\boldsymbol{\theta}) & \propto & \exp\left(\phi(\boldsymbol{\theta})'\left\{\mathbb{E}_{Q_Z}[u(\mathbf{X}, \mathbf{Z})] + \boldsymbol{\nu}\right\}\right) \\ \text{pseudo-E:} & Q_Z(\mathbf{Z}) & \propto & \exp\{\mathbb{E}_{Q_\theta}[\phi(\boldsymbol{\theta})]' u(\mathbf{X}, \mathbf{Z})\} \end{array}$$

See *Latouche* et al. *(2010)* for binary SBM inference.

Operon network: Comparison of VEM and VB

VEM estimates for the $K = 5$ group model lie within the VB approximate 90% credibility intervals (*Gazal* et al. *(2011)*).

$\gamma_{k\ell}$	1	2	3	4	5
1	0.03	0.00	0.03	0.00	0.00
2	6.40	1.50	1.34	0.44	0.00
3	1.21	0.89	0.58	0.00	0.00
4	0.00	0.09	0.00	0.95	0.00
5	8.64	17.65	0.05	72.87	11.01
π	65.49	5.18	7.92	21.10	0.30

Operon network: Comparison of VEM and VB

VEM estimates for the $K = 5$ group model lie within the VB approximate 90% credibility intervals (*Gazal* et al. *(2011)*).

$\gamma_{k\ell}$	1	2	3	4	5
1	0.03	0.00	0.03	0.00	0.00
2	6.40	1.50	1.34	0.44	0.00
3	1.21	0.89	0.58	0.00	0.00
4	0.00	0.09	0.00	0.95	0.00
5	8.64	17.65	0.05	72.87	11.01
π	65.49	5.18	7.92	21.10	0.30
1	[0.02;0.04]	[0.00;0.10]	[0.01;0.08]	[0.00;0.03]	[0.02;1.34]
2	[6.14;7.60]	[0.61;3.68]	[1.07;3.50]	[0.05;0.54]	[0.33;17.62]
3	[1.20;1.72]	[0.35;2.02]	[0.56;1.92]	[0.03;0.30]	[0.19;10.57]
4	[0.01;0.07]	[0.04;0.51]	[0.01;0.20]	[0.76;1.27]	[0.08;4.43]
5	[6.35;12.70]	[4.60;33.36]	[4.28;24.37]	[63.56;81.28]	[5.00;95.00]
π	[59.65;74.38]	[2.88;6.74]	[5.68;10.77]	[16.02;24.04]	[0.11;1.42]

VEM and VB estimates for the $K = 5$ group model (approximate 90% credibility intervals).

Variational Bayes approximation: Simulation Study

Few is known about the properties of variational-Bayes inference:

- Consistency is proved for some incomplete data models (*McGrory and Titterington (2009)*).
- In practice, VB-EM often under-estimates the posterior variances.

Variational Bayes approximation: Simulation Study

Few is known about the properties of variational-Bayes inference:

- Consistency is proved for some incomplete data models (*McGrory and Titterington (2009)*).
- In practice, VB-EM often under-estimates the posterior variances.

Simulation design:

Variational Bayes approximation: Simulation Study

Few is known about the properties of variational-Bayes inference:

- Consistency is proved for some incomplete data models (*McGrory and Titterington (2009)*).
- In practice, VB-EM often under-estimates the posterior variances.

Simulation design:

- 2-group binary SBM with parameters with 2 scenarios

$$\pi = \left(\begin{array}{cc} 0.6 & 0.4 \end{array} \right), \qquad \gamma = \left(\begin{array}{cc} 0.8 & 0.2 \\ 0.2 & 0.5/0.3 \end{array} \right)$$

Variational Bayes approximation: Simulation Study

Few is known about the properties of variational-Bayes inference:

- Consistency is proved for some incomplete data models (*McGrory and Titterington (2009)*).
- In practice, VB-EM often under-estimates the posterior variances.

Simulation design:

- 2-group binary SBM with parameters with 2 scenarios

$$\pi = \left(\begin{array}{cc} 0.6 & 0.4 \end{array} \right), \qquad \gamma = \left(\begin{array}{cc} 0.8 & 0.2 \\ 0.2 & 0.5/0.3 \end{array} \right)$$

- Comparison of 4 methods: EM (when possible), VEM, BP and VB

Variational Bayes approximation: Simulation Study

Few is known about the properties of variational-Bayes inference:

- Consistency is proved for some incomplete data models (*McGrory and Titterington (2009)*).
- In practice, VB-EM often under-estimates the posterior variances.

Simulation design:

- 2-group binary SBM with parameters with 2 scenarios

$$\pi = \left(\begin{array}{cc} 0.6 & 0.4 \end{array} \right), \qquad \gamma = \left(\begin{array}{cc} 0.8 & 0.2 \\ 0.2 & 0.5/0.3 \end{array} \right)$$

- Comparison of 4 methods: EM (when possible), VEM, BP and VB
- Belief Propagation (BP) algorithm:

$$\mathbb{E}_Q[\log P(\mathbf{X}, \mathbf{Z})] = \sum_{i,k} \underbrace{\mathbb{E}_Q[Z_{ik}]}_{\tau_{ik}} \log \pi_k + \sum_{i,j} \sum_{k,\ell} \underbrace{\mathbb{E}_Q[Z_{ik} Z_{j\ell}]}_{\Delta_{ijk\ell} \neq \tau_{ik}\tau_{j\ell}} \log f(X_{ij}; \gamma_{k\ell}).$$

Variational Bayes approximation: Simulation Study

Few is known about the properties of variational-Bayes inference:

- Consistency is proved for some incomplete data models (*McGrory and Titterington (2009)*).
- In practice, VB-EM often under-estimates the posterior variances.

Simulation design:

- 2-group binary SBM with parameters with 2 scenarios

$$\pi = \left(\begin{array}{cc} 0.6 & 0.4 \end{array} \right), \qquad \gamma = \left(\begin{array}{cc} 0.8 & 0.2 \\ 0.2 & 0.5/0.3 \end{array} \right)$$

- Comparison of 4 methods: EM (when possible), VEM, BP and VB
- Belief Propagation (BP) algorithm:

$$\mathbb{E}_Q[\log P(\mathbf{X}, \mathbf{Z})] = \sum_{i,k} \underbrace{\mathbb{E}_Q[Z_{ik}]}_{\tau_{ik}} \log \pi_k + \sum_{i,j} \sum_{k,\ell} \underbrace{\mathbb{E}_Q[Z_{ik} Z_{j\ell}]}_{\Delta_{ijk\ell} \neq \tau_{ik}\tau_{j\ell}} \log f(X_{ij}; \gamma_{k\ell}).$$

- 500 graphs are simulated for each scenario and each graph size.

Estimates, standard deviation and likelihood

Comparison on small graphs ($n = 18$):

	π_1	γ_{11}	γ_{12}	γ_{22}	$\log P(X)$
Scenario 1	60%	80%	20%	50%	
EM	59.1 (13.1)	78.5 (13.5)	20.9 (8.4)	50.9 (15.4)	-90.68
VEM	57.7 (16.6)	78.8 (12.4)	22.4 (10.7)	50.3 (14.6)	-90.87
BP	57.9 (16.2)	78.9 (12.3)	22.2 (10.5)	50.3 (14.5)	-90.85
VB	58.1 (13.3)	78.2 (9.7)	21.6 (7.7)	50.8 (13.3)	-90.71

Scenario 2	60%	80%	20%	30%	
EM	59.5 (14.1)	78.7 (15.6)	21.2 (8.7)	30.3 (14.3)	-88.18
VEM	55.6 (19.0)	80.1 (14.0)	24.0 (11.8)	30.8 (13.8)	-88.54
BP	56.6 (17.8)	80.0 (13.6)	23.2 (11.0)	30.8 (13.8)	-88.40
VB	58.4 (14.6)	77.9 (12.0)	22.3 (9.3)	32.1 (12.3)	-88.26

Estimates, standard deviation and likelihood

Comparison on small graphs ($n = 18$):

	π_1	γ_{11}	γ_{12}	γ_{22}	$\log P(X)$
Scenario 1	60%	80%	20%	50%	
EM	59.1 (13.1)	78.5 (13.5)	20.9 (8.4)	50.9 (15.4)	-90.68
VEM	57.7 (16.6)	78.8 (12.4)	22.4 (10.7)	50.3 (14.6)	-90.87
BP	57.9 (16.2)	78.9 (12.3)	22.2 (10.5)	50.3 (14.5)	-90.85
VB	58.1 (13.3)	78.2 (9.7)	21.6 (7.7)	50.8 (13.3)	-90.71

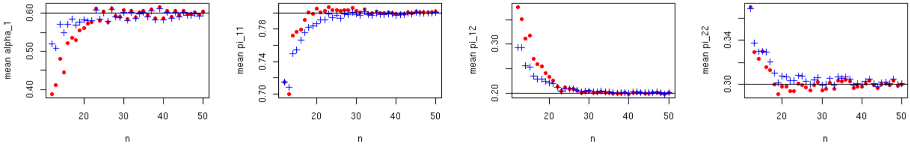
Scenario 2	60%	80%	20%	30%	
EM	59.5 (14.1)	78.7 (15.6)	21.2 (8.7)	30.3 (14.3)	-88.18
VEM	55.6 (19.0)	80.1 (14.0)	24.0 (11.8)	30.8 (13.8)	-88.54
BP	56.6 (17.8)	80.0 (13.6)	23.2 (11.0)	30.8 (13.8)	-88.40
VB	58.4 (14.6)	77.9 (12.0)	22.3 (9.3)	32.1 (12.3)	-88.26

- All methods provide similar results.
- EM achieves the best ones.
- Belief propagation (BP) does not significantly improve VEM.

Influence of the graph size

Comparison of VEM: • and VB: + in scenario 2 (most difficult).
Left to right: π_1, γ_{11}, γ_{12}, γ_{22}.

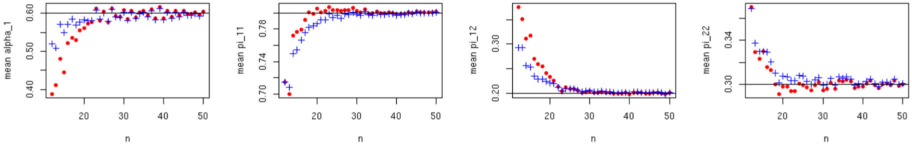
Means.



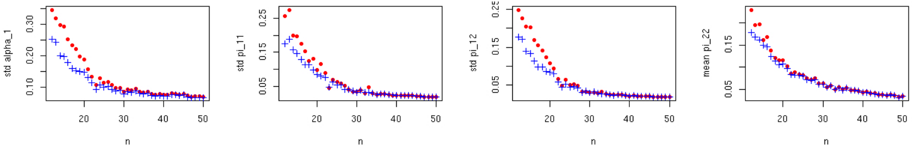
Influence of the graph size

Comparison of VEM: • and VB: + in scenario 2 (most difficult).
Left to right: π_1, γ_{11}, γ_{12}, γ_{22}.

Means.



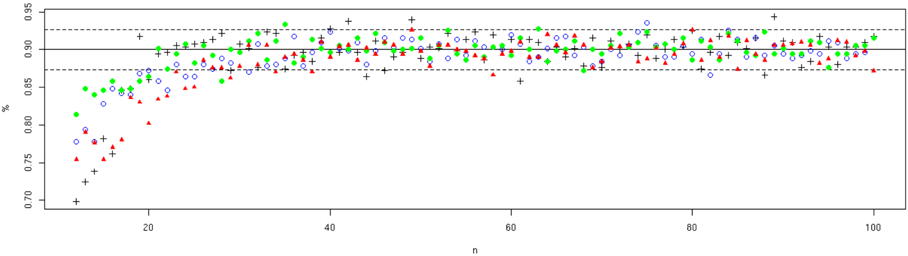
Standard deviations.



- VB estimates converge more rapidly than VEM estimates.
- Their precision is also better.

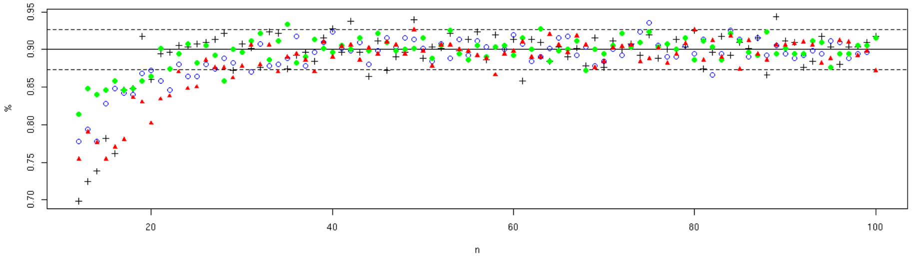
VB Credibility intervals

Actual level as a function of n: π_1: $+$, γ_{11}: $\triangle$, γ_{12}: $\circ$, γ_{22}: $\bullet$



VB Credibility intervals

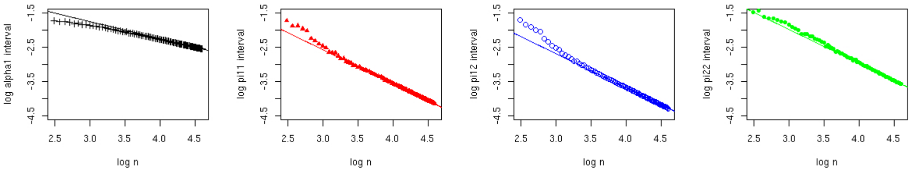
Actual level as a function of n: π_1: $+$, γ_{11}: $\triangle$, γ_{12}: $\circ$, γ_{22}: $\bullet$



- For all parameters, VB posterior credibility intervals achieve the nominal level (90%), as soon as $n \geq 30$.
- $\rightarrow$ The VB approximation seems to work well.

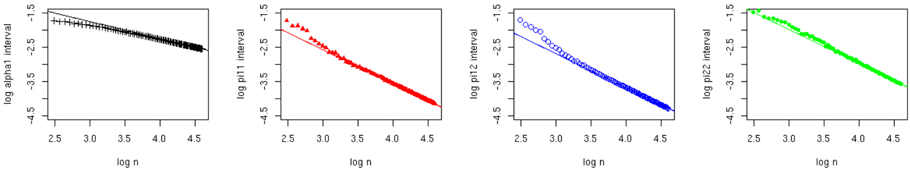
Convergence rate of the VB estimates

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



Convergence rate of the VB estimates

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



- The width decreases as $1/\sqrt{n}$ for π_1.
- It decreases as $1/n = 1/sqrt n^2$ for γ_{11}, γ_{12} and γ_{22}.
- Consistent with the penalty of the ICL criterion proposed by *Daudin* et al. *(2008)* (see next slide).

Few more about inference

Identifiability. Even for binary edges, MixNet (SBM) is identifiable (*Allman et al. (2009)*) ... although mixtures of Bernoulli are not.

Few more about inference

Identifiability. Even for binary edges, MixNet (SBM) is identifiable (*Allman* et al. *(2009)*) ... although mixtures of Bernoulli are not.

Model selection.

- *Daudin* et al. *(2008)* propose the penalised criterion

$$ICL(K) = \mathbb{E}_{Q^*}[\log P(\mathbf{Z}, \mathbf{X})] - \frac{1}{2}\left\{(K-1)\log n + K^2 \log[n(n-1)/2]\right\}.$$

Few more about inference

Identifiability. Even for binary edges, MixNet (SBM) is identifiable (*Allman* et al. *(2009)*) ... although mixtures of Bernoulli are not.

Model selection.

- *Daudin* et al. *(2008)* propose the penalised criterion

$$ICL(K) = \mathbb{E}_{Q^*}[\log P(\mathbf{Z}, \mathbf{X})] - \frac{1}{2} \left\{(K-1) \log n + K^2 \log[n(n-1)/2]\right\}.$$

- The difference between ICL and BIC is the entropy term $\mathcal{H}(Q^*)$ ... which is almost zero (due to the concentration of $P(\mathbf{Z}|\mathbf{X})$).

Few more about inference

Identifiability. Even for binary edges, MixNet (SBM) is identifiable (*Allman* et al. *(2009)*) ... although mixtures of Bernoulli are not.

Model selection.

- *Daudin* et al. *(2008)* propose the penalised criterion

$$ICL(K) = \mathbb{E}_{Q^*}[\log P(\mathbf{Z}, \mathbf{X})] - \frac{1}{2} \left\{ (K-1) \log n + K^2 \log[n(n-1)/2] \right\}.$$

- The difference between ICL and BIC is the entropy term $\mathcal{H}(Q^*)$ ... which is almost zero (due to the concentration of $P(\mathbf{Z}|\mathbf{X})$).
- BIC and ICL-like criteria are also considered in *Latouche* et al. *(2011b)* for SBM in the context of variational Bayes inference.

Covariates in weighted networks

Weighted network

Understanding the mixture components: Observed clusters may be related to exogenous covariates.

Model-based clustering (such as SBM) provides a comfortable set-up to account for covariates.

Weighted network

Understanding the mixture components: Observed clusters may be related to exogenous covariates.

Model-based clustering (such as SBM) provides a comfortable set-up to account for covariates.

Generalised linear model. In the context of exponential family, covariates **y** can be accounted for via a regression term

$$g(\mathbb{E}X_{ij}) = \mu_{k\ell} + \mathbf{y}_{ij}\boldsymbol{\beta}, \qquad \text{if } Z_{ik}Z_{j\ell} = 1$$

where $\boldsymbol{\beta}$ does not depend on the group (*Mariadassou* et al. *(2010)*).

Weighted network

Understanding the mixture components: Observed clusters may be related to exogenous covariates.

Model-based clustering (such as SBM) provides a comfortable set-up to account for covariates.

Generalised linear model. In the context of exponential family, covariates **y** can be accounted for via a regression term

$$g(\mathbb{E}X_{ij}) = \mu_{k\ell} + \mathbf{y}_{ij}\boldsymbol{\beta}, \qquad \text{if } Z_{ik}Z_{j\ell} = 1$$

where $\boldsymbol{\beta}$ does not depend on the group (*Mariadassou* et al. *(2010)*).

Both VEM or VBEM inference can be performed.

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:

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

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

Results: ICL selects $K = 7$ 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

Tree interaction network

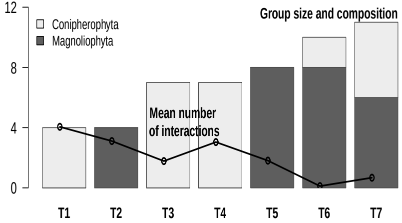
Data: $n = 51$ tree species,
X_{ij} = number of shared parasites (*Vacher* et al. *(2008)*).

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

Model:

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

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



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

Accounting for taxonomic distance

Model: $d_{ij} = d_{taxo}(i,j)$,

$$X_{ij} \sim \mathcal{P}(\lambda_{k\ell} e^{\beta d_{ij}}).$$

Accounting for taxonomic distance

Model: $d_{ij} = d_{taxo}(i,j)$,

$$X_{ij} \sim \mathcal{P}(\lambda_{k\ell} e^{\beta d_{ij}}).$$

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

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

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

Accounting for taxonomic distance

Model: $d_{ij} = d_{taxo}(i, j)$,

$$X_{ij} \sim \mathcal{P}(\lambda_{k\ell} e^{\beta d_{ij}}).$$

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

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

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

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

Accounting for taxonomic distance

Model: $d_{ij} = d_{taxo}(i, j)$,

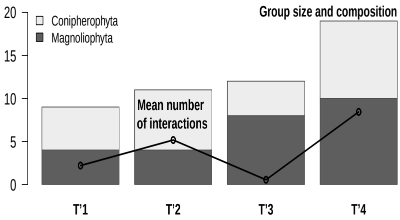
$$X_{ij} \sim \mathcal{P}(\lambda_{k\ell} e^{\beta d_{ij}}).$$

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

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

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

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



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

Conclusion

Conclusion

Stochastic block-model: flexible and already widely used mixture model to uncover some underlying heterogeneity in networks.

Conclusion

Stochastic block-model: flexible and already widely used mixture model to uncover some underlying heterogeneity in networks.

Variational inference

- Efficient and scalable (*Daudin (2011)*: $n > 2000$) in terms of computation times (as opposed to MCMC).
- Seems to work well, because the conditional distribution $P(\mathbf{Z}|\mathbf{X})$ (and therefore $P(\mathbf{Z}, \theta|\mathbf{X})$) asymptotically belongs to the class $\mathcal{Q}$ within which the optimisation if achieved.
- Due to the specific asymptotic framework of networks.

Conclusion

Stochastic block-model: flexible and already widely used mixture model to uncover some underlying heterogeneity in networks.

Variational inference

- Efficient and scalable (*Daudin (2011)*: $n > 2000$) in terms of computation times (as opposed to MCMC).
- Seems to work well, because the conditional distribution $P(\mathbf{Z}|\mathbf{X})$ (and therefore $P(\mathbf{Z}, \theta|\mathbf{X})$) asymptotically belongs to the class $\mathcal{Q}$ within which the optimisation if achieved.
- Due to the specific asymptotic framework of networks.

Alternative methods. Faster algorithms do exist for large graphs:

- Based on the degree distribution (*Channarond (2011)*)
- Based on spectral clustering (*Rohe* et al. *(2010)*).

Future work

Theoretical properties of variational estimates. Although the graph context seems favourable, we still need more understanding about variational and variational Bayes inference properties.

Future work

Theoretical properties of variational estimates. Although the graph context seems favourable, we still need more understanding about variational and variational Bayes inference properties.

SBM = discrete version of W-graph. Let

$$\begin{array}{rcl} \phi : [0,1]^2 & \rightarrow & [0,1] \\ \{Z_i\} \text{ i.i.d.} & \sim & \mathcal{U}[0,1] \\ \{X_{ij}\} \text{ indep.} | \{Z_i\} & \sim & \mathcal{B}[\phi(Z_i, Z_j)] \end{array}$$

- Approximation of the $\phi(u,v)$ function by a step function $\gamma_{k\ell}$ (SBM)
- Model averaging based on optimal variational weights (*Volant (2011)*)

Acknowledgements

People:

A. Célisse, A. Channarond, J.-J. Daudin, S. Gazall, M. Mariadassou, V. Miele, F. Picard, C. Vacher

Grant:

Supported by the French Agence Nationale de la Recherche NeMo project ANR-08-BLAN-0304-01

Acknowledgements

People:

A. Célisse, A. Channarond, J.-J. Daudin, S. Gazall, M. Mariadassou, V. Miele, F. Picard, C. Vacher

Grant:

Supported by the French Agence Nationale de la Recherche
NeMo project ANR-08-BLAN-0304-01

Softwares:

- Stand-alone MixNet:
  `stat.genopole.cnrs.fr/software/mixnet/`
- R-package Mixer:
  `cran.r-project.org/web/packages/mixer/index.html`
- R-package NeMo: Network motif detection
  in preparation

AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E. and XING, E. P. (2008). Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.* **9** 1981–2014.

ALLMAN, E., MATIAS, C. and RHODES, J. (2009). Identifiability of parameters in latent structure models with many observed variables. *Ann. Statist.* **37 (6A)** 3099–132.

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.

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

DAUDIN, J., PIERRE, L. and VACHER, C. (Jan, 2010). Model for heterogeneous random networks using continuous latent variables and an application to a tree-fungus network. DOI:10.1111/j.1541-0420.2009.01378.x.

GAZAL, S., DAUDIN, J.-J. and ROBIN, S. (2011). Accuracy of variational estimates for random graph mixture models. *to appear.*

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

GUNAWARDANA, A. *and* BYRNE, W. *(2005). Convergence theorems for generalized alternating minimization procedures.* J. Mach. Learn. Res. **6** *2049–73.*

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.*

HOFMAN, J. M. *and* WIGGINS, C. H. *(2008). A bayesian approach to network modularity.* Physical Review Letters. **100** *258701.* `doi:10.1103/PhysRevLett.100.258701`.

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. *(2010). A non-asymptotic bic criterion for stochastic blockmodels. submitted, Tech. Report,* genome.jouy.inra.fr/ssb/preprint/, *# 17.*

LATOUCHE, P., BIRMELE, E. *and* AMBROISE, C. *(2011a). Overlapping stochastic block models with application to the french political blogosphere. to appear, ArXiv:*arxiv.org/pdf/0910.2098*.*

LATOUCHE, P., BIRMELE, E. *and* AMBROISE, C. *(2011b). Variational bayesian inference and complexity control for stochastic block models. to appear.*

LAURITZEN, S. *(1996).* Graphical Models*. Oxford Statistical Science Series. Clarendon Press.*

LOVÁSZ, L. *and* SZEGEDY, B. *(2006). Limits of dense graph sequences.* J. Combin. Theory, Series B. **96 (6)** *933–57.*

VON LUXBURG, U., BELKIN, M. *and* BOUSQUET, O. *(2007). Consistency of spectral clustering.* Ann. Statist. **36 (2)** *555–86.*

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

MCGRORY, C. A. *and* TITTERINGTON, D. M. *(2009). Variational Bayesian analysis for hidden Markov models.* Austr. & New Zeal. J. Statist. **51 (2)** *227–44.*

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.*

OPPER, M. *and* WINTHER, O. *(2001).* Advanced mean field methods: Theory and practice*. chapter From Naive Mean Field Theory to the TAP Equations. The MIT Press.*

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

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.

Rohe, K., Chatterjee, S. *and* Yu, B. *(July, 2010). Spectral clustering and the high-dimensional Stochastic Block Model.*

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.

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.*

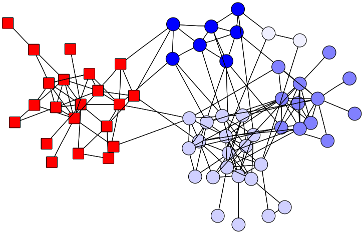
Understanding network structure

- Network constitute a natural way to depict interactions between entities.
- They are now present in many scientific fields (biology, sociology, communication, economics, ...).
- Most observed networks display an heterogeneous topology, that one would like to decipher and better understand.

Understanding network structure

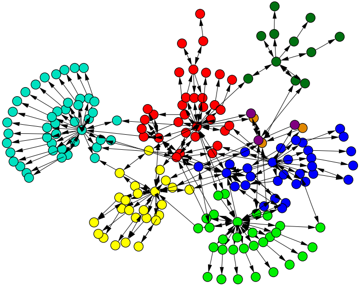
- Network constitute a natural way to depict interactions between entities.
- They are now present in many scientific fields (biology, sociology, communication, economics, ...).
- Most observed networks display an heterogeneous topology, that one would like to decipher and better understand.

Dolphine social network.

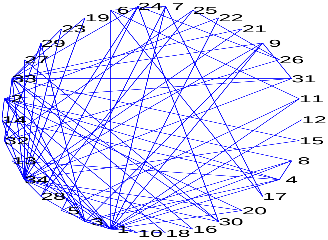


Newman and Girvan (2004)

Hyperlink network.

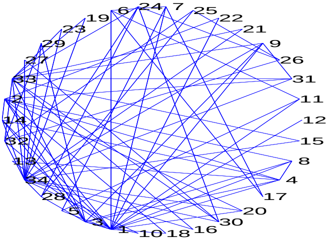


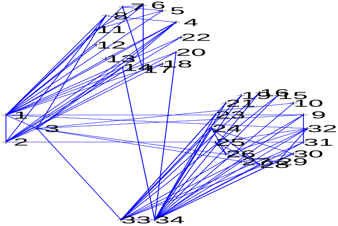
SBM for a binary social network



Zachary data. Social binary network of friendship within a sport club.

SBM for a binary social network

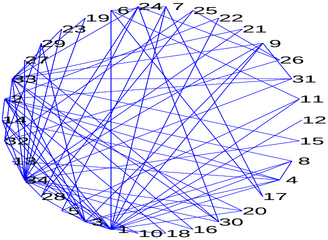


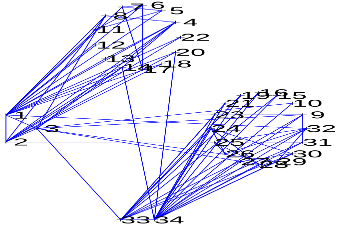


Zachary data. Social binary network of friendship within a sport club.

Results. The split is recovered and the role of few leaders is underlined.

SBM for a binary social network





Zachary data. Social binary network of friendship within a sport club.

Results. The split is recovered and the role of few leaders is underlined.

$$X_{ij}|Z_i = q, Z_j = \ell \sim \mathcal{B}(\gamma_{q\ell})$$

(%)	$\widehat{\gamma}_{k\ell}$			
k/ℓ	1	2	3	4
1	100	53	16	16
2	-	12	0	7
3	-	-	8	73
4	-	-	-	100
$\widehat{\pi}_\ell$	9	38	47	6

Extensions and variations

Algorithmic approaches: Looking for communities

- Graph clustering (*Girvan and Newman (2002)*, *Newman (2004)*);
- Spectral clustering (*von Luxburg* et al. *(2007)*).

Extensions and variations

Algorithmic approaches: Looking for communities

- Graph clustering (*Girvan and Newman (2002)*, *Newman (2004)*);
- Spectral clustering (*von Luxburg* et al. *(2007)*).

Variations around SBM:

- Community structure (*Hofman and Wiggins (2008)*),
- Mixed-membership (*Airoldi* et al. *(2008)*), overlapping groups (*Latouche* et al. *(2011a)*)
- Continuous version (*Daudin* et al. *(2010)*),
- SBM = Step-function version of W-random graphs (*Lovász and Szegedy (2006)*)

Extensions and variations

Algorithmic approaches: Looking for communities

- Graph clustering (*Girvan and Newman (2002)*, *Newman (2004)*);
- Spectral clustering (*von Luxburg* et al. *(2007)*).

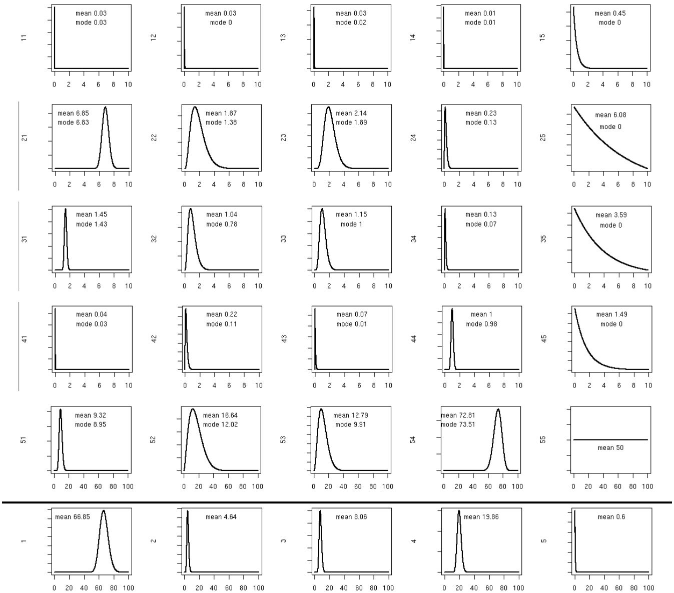
Variations around SBM:

- Community structure (*Hofman and Wiggins (2008)*),
- Mixed-membership (*Airoldi* et al. *(2008)*), overlapping groups (*Latouche* et al. *(2011a)*)
- Continuous version (*Daudin* et al. *(2010)*),
- SBM = Step-function version of W-random graphs (*Lovász and Szegedy (2006)*)

In this talk:

- Variational inference for SBM;
- Variational Bayes inference for SBM;
- Including covariates.

Approximate posterior distribution Q_θ^*



Comparison of classifications and G-O-F

Accounting for taxonomy deeply modifies the group structure:

	T'1	T'2	T'3	T'4
T1	-	-	-	4
T2	-	-	-	4
T3	2	5	-	-
T4	-	2	-	5
T5	-	2	-	6
T6	-	-	10	-
T7	7	2	2	-

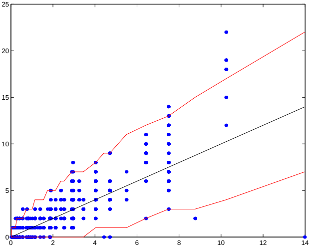
Comparison of classifications and G-O-F

Accounting for taxonomy deeply modifies the group structure:

	T'1	T'2	T'3	T'4
T1	-	-	-	4
T2	-	-	-	4
T3	2	5	-	-
T4	-	2	-	5
T5	-	2	-	6
T6	-	-	10	-
T7	7	2	2	-

Goodness of fit can be assessed via the predicted intensities $\widehat{X}_{ij}$ or degrees $\widehat{K}_i$.

Edges X_{ij}:



Degrees K_i: