

Bayesian non parametric inference of discrete valued networks

L. Nouedoui, P. Latouche

Université Paris 1 Panthéon-Sorbonne
Laboratoire SAMM

ESANN 13



Contents

Introduction

Real networks

Graph clustering

Stochastic block models

The model

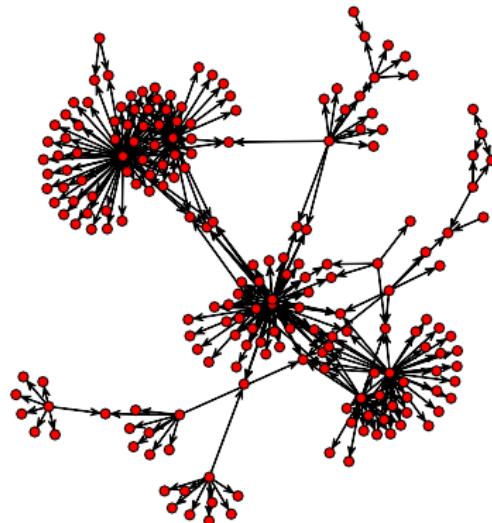
Poisson mixture model

Infinite Poisson mixture model

Chinese restaurant process

Inference

Experiments



Subset of the yeast transcriptional regulatory network (Milo et al., 2002).

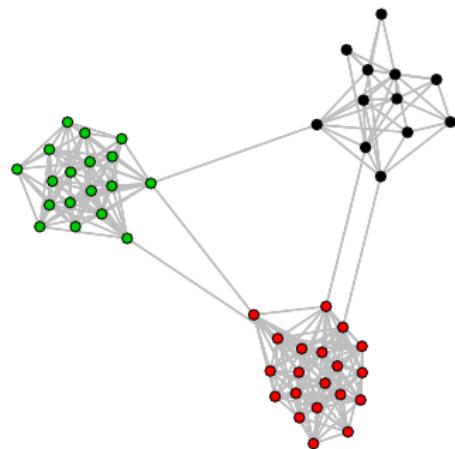
Graph clustering

- ▶ **Existing methods look for :**

- ▶ Community structure
- ▶ Disassortative mixing
- ▶ Heterogeneous structure

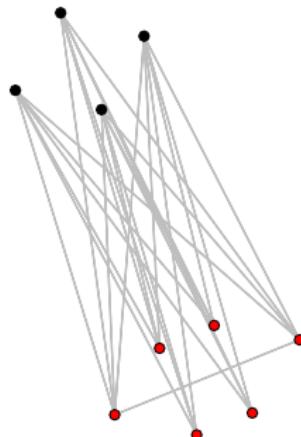
Graph clustering

- ▶ Existing methods look for :
 - ▶ Community structure
 - ▶ Disassortative mixing
 - ▶ Heterogeneous structure



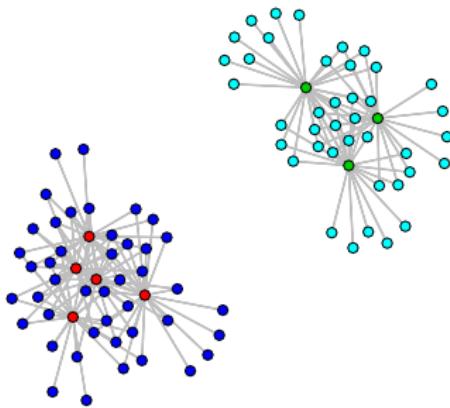
Graph clustering

- ▶ Existing methods look for :
 - ▶ Community structure
 - ▶ Disassortative mixing
 - ▶ Heterogeneous structure



Graph clustering

- ▶ Existing methods look for :
 - ▶ Community structure
 - ▶ Disassortative mixing
 - ▶ Heterogeneous structure



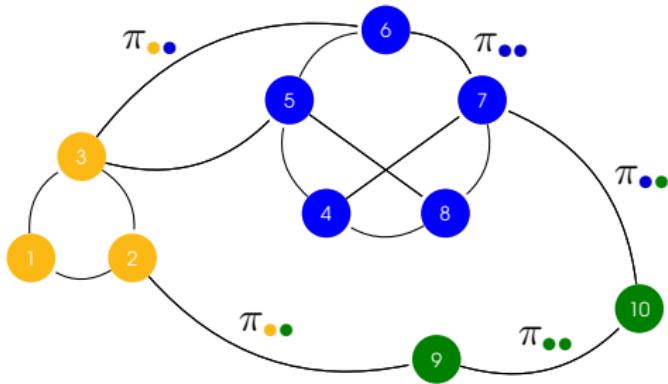
Stochastic Block Model (SBM)

- ▶ Nowicki and Snijders (2001)
 - ▶ Earlier work : Govaert et al. (1977)
- ▶ \mathbf{Z}_i independent hidden variables :
 - ▶ $\mathbf{Z}_i \sim \mathcal{M}\left(1, \boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_K)\right)$
 - ▶ $Z_{ik} = 1$: vertex i belongs to class k
- ▶ $\mathbf{X} | \mathbf{Z}$ edges drawn independently :

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

- ▶ A mixture model for graphs :

$$X_{ij} \sim \sum_{k=1}^K \sum_{l=1}^K \alpha_k \alpha_l \mathcal{B}(\pi_{kl})$$

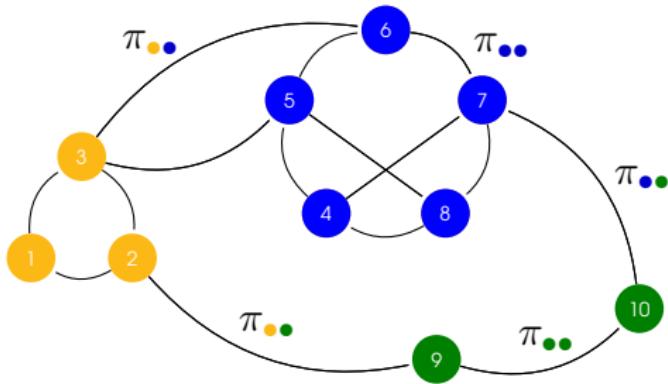


Approximations

Gibbs : Nowicki and Snijders (2001)

VEM : Daudin et al. (2008)

VBEM : Latouche et al. (2012)



Approximations

Gibbs : Nowicki and Snijders (2001)

VEM : Daudin et al. (2008)

VBEM : Latouche et al. (2012)

Poisson mixture model for networks

- ▶ Many networks have discrete edges
- ▶ Extension of SBM to discrete edges : $X_{ij} \in \mathbb{N}$
- ▶ $X_{ij} | \{Z_{ik} Z_{jl} = 1\} \sim \mathcal{P}(\lambda_{kl})$
- ▶ Poisson mixture model (PM) (Mariadassou et al., 2010)
- ▶ Inference : VEM + ICL

ICL

- ▶ Based on a Laplace (asymptotic) approximation
- ▶ Problem for small sample sizes

- ▶ Many networks have discrete edges
- ▶ Extension of SBM to discrete edges : $X_{ij} \in \mathbb{N}$
- ▶ $X_{ij} | \{Z_{ik} Z_{jl} = 1\} \sim \mathcal{P}(\lambda_{kl})$
- ▶ Poisson mixture model (PM) (Mariadassou et al., 2010)
- ▶ Inference : VEM + ICL

ICL

- ▶ Based on a Laplace (asymptotic) approximation
- ▶ Problem for small sample sizes

Chinese restaurant process

- ▶ Non parametric prior for PM
- ▶ Each class attracts new data points depending on its current size
- ▶ Assume they are the $m - 1$ observations classified
- ▶ A new data point is assigned to
 - ▶ class k with probability $\propto n_k$
 - ▶ a new class with probability $\propto \eta_0$
- ▶ Exchangeable distribution

Chinese restaurant process

- ▶ Stick-Breaking prior
 - ▶ $\beta_k \sim \text{Beta}(1; \eta_0), \forall k$
 - ▶ $\alpha_1 = \beta_1$
 - ▶ $\alpha_k = \beta_k \prod_{l=1}^{k-1} (1 - \beta_l)$
- ▶ $Z_i | \alpha \sim \mathcal{M}(1, \alpha)$
- ▶ Conjugate prior
 - ▶ $\lambda_{kl} | a, b \sim \text{Gamma}(a, b)$
 - ▶ Choice for the hyperparameters a and b

Gibbs sampling

- ▶ $p(\mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\lambda} | \mathbf{X})$ not tractable
- ▶ Gibbs sampling procedure :
 - ▶ $\boldsymbol{\beta} \sim p(\boldsymbol{\beta} | \mathbf{X}, \mathbf{Z}, \boldsymbol{\lambda})$ then compute $\boldsymbol{\alpha}$
 - ▶ $\mathbf{Z}_i \sim p(\mathbf{Z}_i | \mathbf{X}, \mathbf{Z}_{\setminus i}, \boldsymbol{\alpha}, \boldsymbol{\lambda})$
 - ▶ $\boldsymbol{\lambda} \sim p(\boldsymbol{\lambda} | \mathbf{X}, \mathbf{Z}, \boldsymbol{\alpha})$
- ▶ Start with $K = K_{up}$ classes
- ▶ Some classes get empty during the algorithm
- ▶ Number of non empty classes : estimate of K

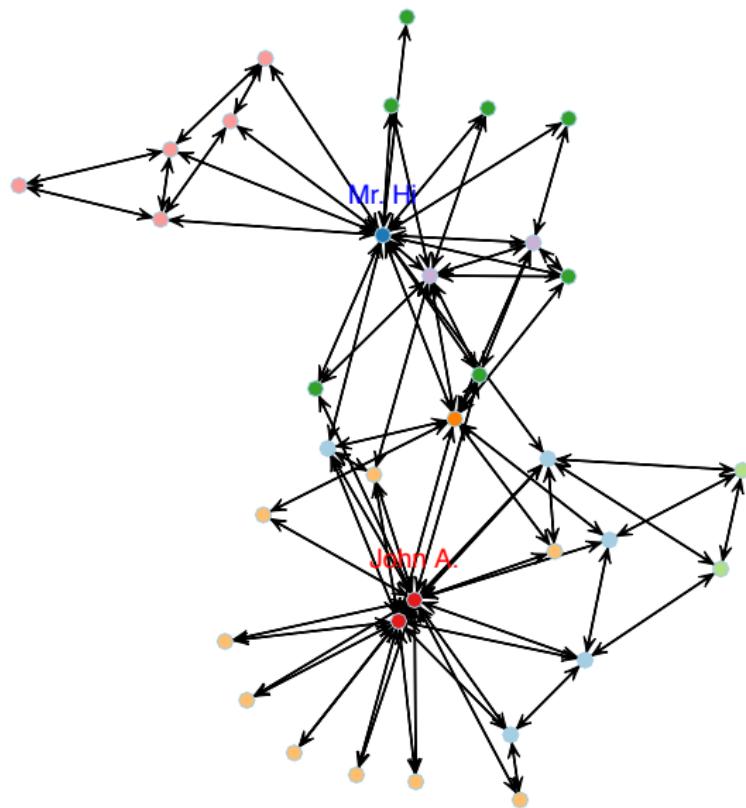
Experiments

- ▶ Simulate networks
- ▶ $N = 50, 100, 500, 1000$
- ▶ $K = 3$
- ▶ Unbalanced proportions : $\alpha_k \propto (1/2)^k$
 - ▶ $\alpha = (80.6, 16.1, 3.3)$
- ▶ $\lambda_{kl} = \lambda'$ and $\lambda_{lk} = (1/2)\lambda'$

Experiments

Network size	Model	$\hat{K}_n = 3$	$\hat{K}_n = 2$	$\hat{K}_n = 4$
$N = 50$	IPM	0.59	0.41	0.00
	PM	0.17	0.82	0.01
$N = 100$	IPM	0.96	0.04	0.00
	PM	0.90	0.07	0.03
$N = 500$	IPM	1.00	0.00	0.00
	PM	1.00	0.00	0.00
$N = 1000$	IPM	1.00	0.00	0.00
	PM	1.00	0.00	0.00

Real data : Zachary on UCINET



References

- ▶ K. Nowicki and T.A.B. Snijders (2001), Estimation and prediction for stochastic blockstructures. *96*, 1077-1087
- ▶ E.M. Airoldi, D.M. Blei, S.E. Fienberg, E.P. Xing (2008), Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, *9*, 1981-2014
- ▶ J-J. Daudin, F. Picard et S. Robin (2008), A mixture model for random graphs. *Statistics and Computing*, *18*, 2, 151-171
- ▶ P. Latouche, E. Birmelé, C. Ambroise (2011), Overlapping stochastic block models with application to the French political blogosphere network. *Annals of Applied Statistics*, *5*, 1, 309-336
- ▶ P. Latouche, E. Birmelé, C. Ambroise (2012), Variational Bayesian inference and complexity control for stochastic block models. *Statistical Modelling*, *12*, 1, 93-115

- ▶ **Log-likelihoods of the model :**

- ▶ Observed-data : $\log p(\mathbf{X} | \boldsymbol{\alpha}, \boldsymbol{\Pi}) = \log \{\sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z} | \boldsymbol{\alpha}, \boldsymbol{\Pi})\}$
 $\hookrightarrow K^N$ terms
- ▶ Expectation Maximization (EM) algorithm requires the knowledge of $p(\mathbf{Z} | \mathbf{X}, \boldsymbol{\alpha}, \boldsymbol{\Pi})$

Problem

$p(\mathbf{Z} | \mathbf{X}, \boldsymbol{\alpha}, \boldsymbol{\Pi})$ is not tractable (no conditional independence)

Approximations

Gibbs : Nowicki and Snijders (2001)

VEM : Daudin et al. (2008)

VBEM : Latouche et al. (2012)

- ▶ **Log-likelihoods of the model :**

- ▶ Observed-data : $\log p(\mathbf{X} | \boldsymbol{\alpha}, \boldsymbol{\Pi}) = \log \{\sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z} | \boldsymbol{\alpha}, \boldsymbol{\Pi})\}$
 $\hookrightarrow K^N$ terms
- ▶ Expectation Maximization (EM) algorithm requires the knowledge of $p(\mathbf{Z} | \mathbf{X}, \boldsymbol{\alpha}, \boldsymbol{\Pi})$

Problem

$p(\mathbf{Z} | \mathbf{X}, \boldsymbol{\alpha}, \boldsymbol{\Pi})$ is not tractable (no conditional independence)

Approximations

Gibbs : Nowicki and Snijders (2001)

VEM : Daudin et al. (2008)

VBEM : Latouche et al. (2012)

- ▶ **Log-likelihoods of the model :**

- ▶ Observed-data : $\log p(\mathbf{X} | \boldsymbol{\alpha}, \boldsymbol{\Pi}) = \log \{\sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z} | \boldsymbol{\alpha}, \boldsymbol{\Pi})\}$
 $\hookrightarrow K^N$ terms
- ▶ Expectation Maximization (EM) algorithm requires the knowledge of $p(\mathbf{Z} | \mathbf{X}, \boldsymbol{\alpha}, \boldsymbol{\Pi})$

Problem

$p(\mathbf{Z} | \mathbf{X}, \boldsymbol{\alpha}, \boldsymbol{\Pi})$ is not tractable (no conditional independence)

Approximations

Gibbs : Nowicki and Snijders (2001)

VEM : Daudin et al. (2008)

VBEM : Latouche et al. (2012)

Criteria

Since $\log p(\mathbf{X} | \boldsymbol{\alpha}, \boldsymbol{\Pi})$ is not tractable, we *cannot* rely on:

- ▶ $AIC = \log p(\mathbf{X} | \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\Pi}}) - C$
- ▶ $BIC = \log p(\mathbf{X} | \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\Pi}}) - \frac{C}{2} \log \frac{N(N-1)}{2}$

ICL

Biernacki et al. (2000) \leftrightarrow Daudin et al. (2008)

Variational Bayes EM \leftrightarrow $ILvb$

Latouche et al. (2012)

Exact ICL $\hookrightarrow ICL_{ex}$

Côme and Latouche (2013)

Criteria

Since $\log p(\mathbf{X} | \boldsymbol{\alpha}, \boldsymbol{\Pi})$ is not tractable, we *cannot* rely on:

- ▶ $AIC = \log p(\mathbf{X} | \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\Pi}}) - C$
- ▶ $BIC = \log p(\mathbf{X} | \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\Pi}}) - \frac{C}{2} \log \frac{N(N-1)}{2}$

ICL

Biernacki et al. (2000) \hookrightarrow Daudin et al. (2008)

Variational Bayes EM \hookrightarrow *ILvb*

Latouche et al. (2012)

Exact ICL \hookrightarrow ICL_{ex}

Côme and Latouche (2013)

Criteria

Since $\log p(\mathbf{X} | \boldsymbol{\alpha}, \boldsymbol{\Pi})$ is not tractable, we *cannot* rely on:

- ▶ $AIC = \log p(\mathbf{X} | \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\Pi}}) - C$
- ▶ $BIC = \log p(\mathbf{X} | \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\Pi}}) - \frac{C}{2} \log \frac{N(N-1)}{2}$

ICL

Biernacki et al. (2000) \hookrightarrow Daudin et al. (2008)

Variational Bayes EM \hookrightarrow *ILvb*

Latouche et al. (2012)

Exact ICL \hookrightarrow *ICL_{ex}*

Côme and Latouche (2013)