Chapter 5
Random Graphs for Modeling Network Data

Qing Zhou

UCLA Department of Statistics

Stats 201C Advanced Modeling and Inference
Lecture Notes

Outline

1. Network data
2. Latent space models
3. Stochastic block models
4. Variational EM
5. Community detection
6. Extensions and discussions

Examples & applications

- Social networks.
- Protein-protein interaction networks.
- Biomedical data with family history.

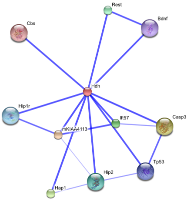

Figure sources: (left) forbes.com; (right) UW Madison.

Network Data

Observed data: A network (graph) among n nodes.

- Each node corresponds to an individual $i \in \{1, \ldots, n\} := V$.
- Connections among the nodes are given by an adjacency matrix, $A = (Y_{ij})_{n \times n}$ (symmetric):

$$Y_{ij} = 0 : \text{no edge between } i \text{ and } j;$$
$$Y_{ij} = 1 : \text{there is edge between } i \text{ and } j.$$

  If $Y_{ij} \in \mathbb{R} \setminus \{0\}$ when there is an edge, weighted graph.

- Build a probabilistic model on the random graph A; an observed network (y_{ij}) is a realization of A.

Modeling heterogeneity: nodes that share a large number of connections form a community (Matias and Robin 2014).

Latent space models

Reference: Hoff et al. (2002).

- Each node $i \in V$ is associated with an independent latent variable $Z_i \in \mathbb{R}^q$. The space for Z_i is the latent space.
- The distribution of the edge Y_{ij} depends on $\|Z_i - Z_j\|$ (distance between Z_i and Z_j in the latent space).
- Conditional distribution $[Y_{ij}|Z_i, Z_j]$ (assuming binary graph):

$$Y_{ij} = Y_{ji} \sim \text{Bern}(\gamma_{ij})$$
$$\text{logit}\{\gamma_{ij}\} = \alpha - \|Z_i - Z_j\|.$$

  If $\|Z_i - Z_j\|$ is small, then $\mathbb{P}(Y_{ij} = 1|Z_i, Z_j)$ is large (more likely to connect i and j).

- Predict Z_i and cluster them to detect communities.

Latent space models

Other related models:

- Graphon: latent variables $U_i \sim \text{Unif}(0,1)$.

$$\begin{aligned} Y_{ij} &\sim \text{Bern}(\gamma_{ij}) \\ \gamma_{ij} &= g(U_i, U_j), \end{aligned}$$

  g is a symmetric function, called a graphon: Nonparametric estimation.

- Stochastic block model (SBM): $Z_i \in \{1, \ldots, K\}$.

Stochastic block models

Model structure:

- Assume K communities (clusters) among the n nodes.
- Latent cluster labels $Z_i = (Z_{i1}, \dots, Z_{iK}) \in \{e_1, \dots, e_K\}$

$$Z_i = (Z_{i1}, \dots, Z_{iK}) \sim_{iid} M(1, \pi),$$

  where $\pi = (\pi_1, \dots, \pi_K)$ are cell probabilities.
- Given Z_i and Z_j, the edge $Y_{ij} = Y_{ji}$ is drawn independently:

$$Y_{ij} \mid Z_{im} = 1, Z_{j\ell} = 1 \sim f(\cdot; \gamma_{m\ell}).$$

  The matrix $\gamma = (\gamma_{m\ell})_{K \times K}$ contains all parameters for connection probabilities among the K communities.

Stochastic block models

Formulate as a hidden variable model:

- Parameters: $\theta = (\pi, \gamma)$.
- Hidden variables (missing data): $Z = (Z_1, \dots, Z_n)$.
- Observed data: $A = (Y_{ij})_{n \times n}$.
- To be concrete, assume

$$
\begin{aligned}
&Y_{ij} \mid Z_{im} = 1, Z_{j\ell} = 1 \sim \mathrm{Bern}(\gamma_{m\ell}) \\
&f(y; \gamma_{m\ell}) = \gamma_{m\ell}^{y}(1 - \gamma_{m\ell})^{1-y}, \quad y \in \{0, 1\}.
\end{aligned}
$$

Using EM for MLE:

- MLE $\hat{\theta}$ is the solution to:

$$\max_{\theta}\left\{\log \mathbb{P}(Y;\theta) = \log\left[\sum_{Z_1}\cdots\sum_{Z_n}\mathbb{P}(Y, Z_1, \ldots, Z_n;\theta)\right]\right\}.$$

- Complete-data log-likelihood

$$\begin{aligned}\ell(\theta \mid Y, Z) &= \log \mathbb{P}(Y, Z;\theta) \\ &= \sum_{i=1}^{n}\sum_{m} Z_{im}\log \pi_m + \frac{1}{2}\sum_{i\neq j}\sum_{m,\ell} Z_{im}Z_{j\ell}\log f(Y_{ij};\gamma_{m\ell}). \quad (1)\end{aligned}$$

- E-step needs $\mathbb{E}(Z_{im} \mid Y;\theta^{(t)})$ and $\mathbb{E}(Z_{im}Z_{j\ell} \mid Y;\theta^{(t)})$.

Stochastic block models

Difficulty:

- E-step is intractable, since $\mathbb{P}(Z_1, \ldots, Z_n \mid Y; \theta^{(t)})$ does not factorize in any way.
- Z_i, Z_j are dependent given Y_{ij} for all $i, j \Rightarrow Z_1, \ldots, Z_n$ are all dependent given $A = (Y_{ij})$.
- Compare:
  (1) Mixture modeling, $Z_i \perp Z_j \mid Y$.
  (2) HMM, $(Z_1, \ldots, Z_n \mid Y)$ is a Markov chain.

Variational EM algorithm

An iterative maximization view of EM:

$$\ell(\theta|Y) := \log \mathbb{P}(Y;\theta) = \log \mathbb{P}(Y,Z;\theta) - \log \mathbb{P}(Z \mid Y;\theta).$$

Take expectation wrt a distribution F over Z:

$$\ell(\theta|Y) = \mathbb{E}_F \{\log \mathbb{P}(Y,Z;\theta)\} + H(F) + KL(F \| \mathbb{P}(Z \mid Y;\theta)), \quad (2)$$

where $H(F) = \mathbb{E}_F\{-\log F(Z)\}$ is the entropy of F and $KL \geq 0$ is the Kullback-Leibler divergence. Thus, for any F

$$\ell(\theta|Y) \geq \mathbb{E}_F \{\log \mathbb{P}(Y,Z;\theta)\} + H(F) := L(\theta, F).$$

$L(\theta, F)$: evidence lower bound (ELBO),
F: variational distribution.

Variational EM algorithm

EM iterates between two maximization steps to

$$\max_{F,\theta} \{L(\theta, F) = \mathbb{E}_F \{\log \mathbb{P}(Y, Z; \theta)\} + H(F)\}.$$

- E-step: Given $\theta^{(t)}$, $\max_F L(\theta^{(t)}, F)$, due to (2), $\Leftrightarrow$

$$\min_F KL(F \| \mathbb{P}(Z \mid Y; \theta^{(t)})) \Rightarrow F^{(t)} = \mathbb{P}(Z \mid Y; \theta^{(t)}).$$

- M-step: Given $F^{(t)}$, $\max_\theta L(\theta, F^{(t)}) \Leftrightarrow$

$$\begin{aligned} \max_\theta \mathbb{E}_{F^{(t)}} \{\log \mathbb{P}(Y, Z; \theta)\} &= \max_\theta \mathbb{E} \left\{\log \mathbb{P}(Y, Z; \theta) \mid Y; \theta^{(t)}\right\} \\ &= \max_\theta Q(\theta \mid \theta^{(t)}) \Rightarrow \theta^{(t+1)}. \end{aligned}$$

Note that $L(\theta, F^{(t)})$ is the minorization function in the MM view of EM.

Variational EM algorithm

Variational EM maximizes $L(\theta, F)$ within a restricted class of $F \in \mathcal{F}$ so that E-step is tractable.

- E-step: Given $\theta^{(t)}$

$$\max_{F \in \mathcal{F}} \mathbb{E}_F \left\{ \log \mathbb{P}(Y, Z; \theta^{(t)}) \right\} + H(F) \Rightarrow F^{(t)} \in \mathcal{F}.$$

- M-step: Given $F^{(t)}$

$$\max_{\theta} \mathbb{E}_{F^{(t)}} \{ \log \mathbb{P}(Y, Z; \theta) \} \Rightarrow \theta^{(t+1)}.$$

Note that $L(\theta, F)$ always a lower bound of $\ell(\theta \mid Y)$ for any F.

Variational EM algorithm

Reference Daudin et al. (2008).
Assume $F(Z) = \prod_{i=1}^{n} h(Z_i; \tau_i)$, and $Z_i \sim M(1, \tau_i)$ under h.

- $\mathbb{E}_F(Z_{im}Z_{j\ell}) = \mathbb{E}_F(Z_{im})\mathbb{E}_F(Z_{j\ell}) = \tau_{im}\tau_{j\ell}$.
- Then plug into complete-date log-likelihood (1) and $H(F)$:

$$L(\theta, F) = \sum_{i=1}^{n} \sum_{m} \tau_{im} \log \pi_m + \frac{1}{2} \sum_{i \neq j} \sum_{m,\ell} \tau_{im}\tau_{j\ell} \log f(Y_{ij}; \gamma_{m\ell})$$
$$- \sum_{i=1}^{n} \sum_{m} \tau_{im} \log \tau_{im} := L(\theta, \tau).$$

- Variational EM iteratively maximize $L(\theta, \tau)$ over τ (E-step) and θ (M-step).

Variational EM algorithm

E-step:

- Given $\theta^{(t)}$, $\max_\tau L(\theta^{(t)}, \tau)$ subject to $\sum_m \tau_{im} = 1$ for all i.

$$\max_\tau L(\theta^{(t)}, \tau) + \sum_{i=1}^{n} \lambda_i \left(1 - \sum_m \tau_{im}\right)$$

$$\Rightarrow \log \pi_m^{(t)} - \log \tau_{im} + \sum_{j \neq i} \sum_{\ell} \tau_{j\ell} \log f(Y_{ij}; \gamma_{m\ell}^{(t)}) = \lambda_i + 1,$$

  by taking derivative wrt τ_{im}.

- No closed form, $\tau^{(t)}$ is given by the fixed point of

$$\tau_{im} \propto \pi_m^{(t)} \prod_{j \neq i} \prod_{\ell=1}^{K} \left\{ f(Y_{ij}; \gamma_{m\ell}^{(t)}) \right\}^{\tau_{j\ell}},$$

  subject to $\sum_m \tau_{im} = 1$ for each i. Use this as an iterative algorithm to obtain $\tau^{(t)}$.

Some intuition behind the update

$$\tau_{im} \propto \pi_m^{(t)} \prod_{j \neq i} \prod_{\ell=1}^{K} \left\{ f(Y_{ij}; \gamma_{m\ell}^{(t)}) \right\}^{\tau_{j\ell}}.$$

Consider a Gibbs sampler for $[Z \mid Y]$ by iteratively sampling from $[Z_i \mid Y, Z_{-i}]$ for $i = 1, \ldots, n$

$$\begin{aligned} \mathbb{P}(Z_{im} = 1 \mid Y, Z_{-i}) &\propto \mathbb{P}(Z_{im} = 1 \mid Z_{-i}) \mathbb{P}(Y \mid Z_{im} = 1, Z_{-i}) \\ &= \pi_m^{(t)} \prod_{j \neq i} \prod_{\ell=1}^{K} \left\{ f(Y_{ij}; \gamma_{m\ell}^{(t)}) \right\}^{Z_{j\ell}}, \end{aligned}$$

given the current parameter $\theta^{(t)}$.

Variational EM algorithm

M-step:

- Given $\tau^{(t)}$, $\max_\tau L(\theta, \tau^{(t)})$ subject to $\sum_m \pi_m = 1$.

$$\pi_m^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} \tau_{im}^{(t)}$$

$$\gamma_{m\ell}^{(t+1)} = \frac{\sum_{i \neq j} \tau_{im}^{(t)} \tau_{j\ell}^{(t)} Y_{ij}}{\sum_{i \neq j} \tau_{im}^{(t)} \tau_{j\ell}^{(t)}}.$$

- $\tau_{im}^{(t)}$ approximates $\mathbb{P}(Z_{im} = 1 \mid Y, \theta^{(t)})$, weight of node i in cluster m.
- $\tau_{im}^{(t)} \tau_{j\ell}^{(t)}$ approximates $\mathbb{P}(Z_{im} = 1, Z_{j\ell} = 1 \mid Y, \theta^{(t)})$, weight of node i in cluster m and j in cluster ℓ (Y_{ij} indicates an edge between the two clusters).

Variational EM algorithm

Consistency of variational estimator (Bickel et al. 2013):

- MLE $\hat{\theta}^{\mathsf{ML}} = \operatorname{argmax}_\theta \ell(\theta \mid Y)$.
- Variational estimator $\hat{\theta}^{\mathsf{VR}} = \operatorname{argmax}_\theta \max_\tau L(\theta, \tau)$.
- Bound $\max_\tau L(\theta, \tau)$ by two log-likelihood functions:

$$\log \mathbb{P}(Y, Z = z; \theta) \leq \max_\tau L(\theta, \tau) \leq \ell(\theta \mid Y), \tag{3}$$

  for any z.
- Asymptotic normality for both estimators as $n \to \infty$.

Variational EM algorithm

Logit transformation of parameters:

$$\omega_m = \log\{\pi_m/\pi_K\}, \quad m = 1, \ldots, K-1,$$
$$\nu_{m\ell} = \log\{\gamma_{m\ell}/(1-\gamma_{m\ell})\}, \quad m, \ell = 1, \ldots, K.$$

Theorem 1

Assume the true parameter is $\theta^* = (\pi^*, \gamma^*)$, where γ^* has no identical columns. Let $\lambda_n = \mathbb{E}(\text{degree}) = n\mathbb{P}_{\theta^*}(Y_{ij} = 1)$. If $\lambda_n / \log n \to \infty$, then

$$\sqrt{n}(\hat{\omega} - \omega^*) \xrightarrow{d} \mathcal{N}(0, \Sigma_1),$$
$$\sqrt{n\lambda_n}(\hat{\nu} - \nu^*) \xrightarrow{d} \mathcal{N}(0, \Sigma_2),$$

for both $\hat{\theta}^{\mathrm{VR}}$ and $\hat{\theta}^{\mathrm{ML}}$, where Σ_1 and Σ_2 are functions of θ^*.

Community detection

Clustering nodes: predict Z.

- Posterior distribution $\mathbb{P}(Z \mid Y, \hat{\theta})$. Celisse et al. (2012) establish

$$\frac{\sum_{z \neq z^*} \mathbb{P}(Z = z \mid Y; \hat{\theta})}{\mathbb{P}(Z = z^* \mid Y; \hat{\theta})} \xrightarrow{p} 0,$$

  where z^* is the true cluster labels.

- Spectral clustering (von Luxburg 2007) also achieves vanishing clustering error rate (Rohe et al. 2011):

$$\frac{\#\text{ of misclustered nodes}}{n} \to 0, \quad a.s.$$

Community detection

Spectral clustering of $A = (Y_{ij})_{n \times n}$ (Rohe et al. 2011):

Define normalized graph Laplacian $L = D^{-1/2} A D^{-1/2}$, where $D = \mathrm{diag}(d_1, \dots, d_n)$ and $d_i = \sum_j Y_{ij}$ is the degree of node i.

1. Find $X = [X_1 \mid \cdots \mid X_K] \in \mathbb{R}^{n \times K}$, X_j's are the orthogonal eigenvectors corresponding to the largest K eigenvalues of L (in absolute value).
2. Treat each row of X as a data point in $\mathbb{R}^K$, apply k-means to cluster the n rows into K clusters, $C_1, \dots, C_K$ (partition of $\{1, \dots, n\}$).

Output: $\hat{Z}_{im} = 1$ if $i \in C_m$.

Community detection

Why does spectral clustering work?

- Define population version of A: $\mathcal{A} = (\mathcal{A}_{ij})_{n \times n}$,

$$\mathcal{A}_{ij} = \mathbb{E}(Y_{ij} \mid Z) = \mathbb{P}(Y_{ij} = 1 \mid Z).$$

- Let $B = (\gamma_{m\ell})_{K \times K}$ and $Z = (Z_{im})_{n \times K}$, then $\mathcal{A} = ZBZ^{\mathsf{T}}$.
- Define the graph Laplacian of $\mathcal{A}$ similarly: $\mathcal{L} = \mathcal{D}^{-1/2}\mathcal{A}\mathcal{D}^{-1/2}$, where $\mathcal{D}_{ii} = \sum_j \mathcal{A}_{ij}$.
- Then the eigenvectors of L converge to the eigenvectors of $\mathcal{L}$.
- $\mathcal{L}$ has K nonzero eigenvalues, the associated eigenvectors $\mathcal{U} = (u_{ij}) = [U_1 \mid \cdots \mid U_K] \in \mathbb{R}^{n \times K}$ satisfies:

$$u_i = u_j \Leftrightarrow Z_i = Z_j,$$

where u_i is the ith row of $\mathcal{U}$.

Community detection

Example of $\mathcal{L}$ and $\mathcal{U}$:

```
> B
     [,1] [,2]
[1,]  0.8  0.1
[2,]  0.1  0.7
> Z
     [,1] [,2]
[1,]    1    0
[2,]    1    0
[3,]    0    1
[4,]    0    1
> A
     [,1] [,2] [,3] [,4]
[1,]  0.8  0.8  0.1  0.1
[2,]  0.8  0.8  0.1  0.1
[3,]  0.1  0.1  0.7  0.7
[4,]  0.1  0.1  0.7  0.7
> L
           [,1]       [,2]       [,3]       [,4]
[1,] 0.44444444 0.44444444 0.05892557 0.05892557
[2,] 0.44444444 0.44444444 0.05892557 0.05892557
[3,] 0.05892557 0.05892557 0.43750000 0.43750000
[4,] 0.05892557 0.05892557 0.43750000 0.43750000
> eigen(L)
$values
[1] 1.0000000 0.7638889 0.0000000 0.0000000

$vectors
           [,1]       [,2]       [,3]          [,4]
[1,] -0.5144958  0.4850713  0.0000000  7.071068e-01
[2,] -0.5144958  0.4850713  0.0000000 -7.071068e-01
[3,] -0.4850713 -0.5144958 -0.7071068 -1.665335e-16
[4,] -0.4850713 -0.5144958  0.7071068 -1.387779e-16
```

Extensions and discussions

- Weighted graphs, e.g., $Y_{ij} \mid Z_{im} = 1, Z_{j\ell} = 1 \sim \text{Poiss}(\gamma_{m\ell})$.
- Degree-corrected block model:

$$Y_{ij} \mid Z_{im} = 1, Z_{j\ell} = 1 \sim \text{Poiss}(\gamma_{m\ell}\kappa_i\kappa_j),$$

  κ_i controls expected degree of node i.
- Accounting for covariates
  1. Nodewise covariates x_i, $i = 1, \ldots, n$:

$$Z_i \sim M(1, \pi(x_i)).$$

  2. Edgewise covariates x_{ij}, $i \neq j$. Bernoulli model:

$$\text{logit}\,\{\mathbb{P}(Y_{ij} = 1 \mid Z_{im} = 1, Z_{j\ell} = 1)\} = x_{ij}^{\mathsf{T}}\beta + \gamma_{m\ell}.$$

  Poisson model:

$$Y_{ij} \mid Z_{im} = 1, Z_{j\ell} = 1 \sim \text{Poiss}(\exp(x_{ij}^{\mathsf{T}}\beta + \gamma_{m\ell})).$$

Graphons

Hereafter, consider simple graphs: unweighted and symmetric.

Recall the definition of a *graphon*, $g : [0,1]^2 \to [0,1]$. We define a random simple graph $(Y_{ij}) \in \{0,1\}^{n \times n}$ given a graphon g:

1. Draw $U_i \sim \text{Unif}(0,1)$ for $i = 1, \ldots, n$.
2. Draw $Y_{ij} = Y_{ji} \sim \text{Bern}(g(U_i, U_j))$ for all $i \neq j$.

Graphons

SBM as a graphon model:

- Partition $(0, 1)$ into K intervals, J_m for $m = 1, \ldots, K$, so that $|J_m| = \pi_m$.
- Let $g(u, v) = \gamma_{m\ell}$ if $u \in J_m$ and $v \in J_\ell$ (block-wise constant).
- Then the graphon is equivalent to the SBM.
  Let $Z_{im} = I(U_i \in J_m)$. If $Z_{im} = 1, Z_{j\ell} = 1$, then

$$g(U_i, U_j) = \gamma_{m\ell}$$
$$Y_{ij} \sim \text{Bern}(g(U_i, U_j)) = \text{Bern}(\gamma_{m\ell}).$$

Graphons

- Exchangeable graphs: A random graph G is said to be exchangeable if its distribution is invariant to any relabeling (or permutation) of its vertex set.
- An equivalent definition is that its adjacency matrix $(Y_{ij})_{n\times n}$ is a jointly exchangeable random array, i.e.

$$\mathbb{P}(Y_{ij} \in A_{ij}, \forall i, j \in [n]) = \mathbb{P}(Y_{\pi(i)\pi(j)} \in A_{ij}, \forall i, j \in [n]) \quad (4)$$

for every permutation π of $\{1, \dots, n\}$ and every collection of measurable sets $\{A_{ij}\}$. We write $(Y_{ij}) \overset{d}{=} (Y_{\pi(i)\pi(j)})$ when (4) holds.

Graphons

Theorem 2 (Aldous-Hoover)

A random array $(X_{ij}), X_{ij} \in \Omega, i, j \in \mathbb{N}$, is jointly exchangeable if and only if there is a random function $F : [0,1]^3 \to \Omega$ such that

$$(X_{ij}) \overset{d}{=} (F(U_i, U_j, U_{ij})), \tag{5}$$

where $(U_i)_{i\in\mathbb{N}}$ and $(U_{ij})_{i,j\in\mathbb{N}}$ are, respectively, an infinite sequence and array of i.i.d. Unif$[0,1]$ independent of F.

A few remarks:

1. $(X_{ij})_{i,j\in\mathbb{N}}$ is an infinite two-way array, $i = 1, 2, \ldots$ and $j = 1, 2, \ldots$. Exchangeability of X is an assumption on the data source.
2. A exchangeable graph G on n nodes is regarded as a sample of finite size from this data source.

Graphons

- Apply Theorem 2 to $(Y_{ij})_{\mathbb{N}\times\mathbb{N}}$ with $\Omega = \{0,1\}$: $F(x,y,u) \in \{0,1\}$ for all $x, y, u \in [0,1]$. Assume F is symmetric in (x,y).
- Define a function $g : [0,1]^2 \to [0,1]$ by $g(x,x) = 0$ and

$$g(x,y) := \mathbb{P}(F(x,y,U) = 1 \mid F),$$

  where $U \sim \text{Unif}[0,1]$ and is independent of F.
- Then g is a random symmetric function and

$$(Y_{ij}) \overset{d}{=} (F(U_i, U_j, U_{ij})) \overset{d}{=} (I(U_{ij} < g(U_i, U_j))). \tag{6}$$

  This is because (Y_{ij}) are independent given (U_i) and F and

$$\begin{aligned}\mathbb{P}(Y_{ij} = 1 \mid U_i, U_j, F) &= g(U_i, U_j) \quad \text{(by definition of } g) \\ &= \mathbb{P}(U_{ij} < g(U_i, U_j) \mid U_i, U_j, F).\end{aligned}$$

Graphons

Corollary 1

A random simple graph G with vertex set $\mathbb{N}$ is exchangeable if and only if there is a random function $g : [0,1]^2 \to [0,1]$ such that its adjacency matrix

$$(Y_{ij}) \overset{d}{=} (I(U_{ij} < g(U_i, U_j))), \tag{7}$$

where (U_i) and (U_{ij}) are i.i.d. Unif$[0,1]$ and independent of g.

The random function g is called a graphon.

Graphons

Every exchangeable random simple graph G on $\mathbb{N}$ is represented by a random graphon g:

1. Draw g from a distribution ν (over functions $[0,1]^2 \to [0,1]$).
2. Draw $U_i, i \in \mathbb{N}$ independently from Unif$[0,1]$.
3. For every pair $i < j \in \mathbb{N}$, draw

$$Y_{ij} \mid g, U_i, U_j \sim \mathrm{Bern}(g(U_i, U_j)).$$

Remarks:

1. The distribution of G is determined by ν.
2. Statistical modeling of exchangeable simple graphs is parameterized by graphons g.

A review article: Orbanz and Roy (2015).

References I

P. Bickel, Choi D., X. Chang, and H. Zhang. Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *Annals of Statistics*, 41:1922–1943, 2013.

A. Celisse, J.J. Daudin, and L. Pierre. Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electronic Journal of Statistics*, 6:1847–1899, 2012.

J.J. Daudin, F. Picard, and S. Robin. A mixture model for random graphs. *Statistics and Computing*, 18:173–183, 2008.

P.D. Hoff, A.E. Raftery, and M.S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97:1090–1098, 2002.

References II

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

P. Orbanz and D.M. Roy. Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37:437–461, 2015.

K. Rohe, S. Chatterjee, and B. Yu. Spectral clustering and the high-dimensional stochastic block model. *Annals of Statistics*, 39:1878–1915, 2011.

U. von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17:395 – 416, 2007.