Estimation and Prediction for Stochastic Blockstructures

Krzysztof Nowicki and Tom A. B. Snijders, 2001

Andrea Teruzzi, DSBA Bocconi '22

July 4, 2022

Table of Contents

Introduction and Preview

- The paper extends the approach of Nowicki and Snijders (1997) to a larger class of relational data.

Introduction and Preview

- The paper extends the approach of Nowicki and Snijders (1997) to a larger class of relational data.
- The goal of the paper is defining a general approach for partitioning a set of individuals into classes, given a set of pairwise relations.

Introduction and Preview

- The paper extends the approach of Nowicki and Snijders (1997) to a larger class of relational data.
- The goal of the paper is defining a general approach for partitioning a set of individuals into classes, given a set of pairwise relations.
- In order to do so we need:

Introduction and Preview

- The paper extends the approach of Nowicki and Snijders (1997) to a larger class of relational data.
- The goal of the paper is defining a general approach for partitioning a set of individuals into classes, given a set of pairwise relations.
- In order to do so we need:

Introduction and Preview

- The paper extends the approach of Nowicki and Snijders (1997) to a larger class of relational data.
- The goal of the paper is defining a general approach for partitioning a set of individuals into classes, given a set of pairwise relations.
- In order to do so we need:
  - Set

Introduction and Preview

- The paper extends the approach of Nowicki and Snijders (1997) to a larger class of relational data.
- The goal of the paper is defining a general approach for partitioning a set of individuals into classes, given a set of pairwise relations.
- In order to do so we need:
  - Set
  - Probability law

Introduction and Preview

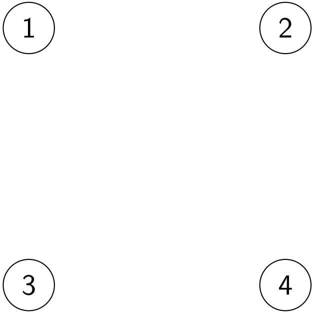
- The paper extends the approach of Nowicki and Snijders (1997) to a larger class of relational data.
- The goal of the paper is defining a general approach for partitioning a set of individuals into classes, given a set of pairwise relations.
- In order to do so we need:
  - Set
  - Probability law
  - Inference on the probability law (Bayesian)

Table of Contents

Example

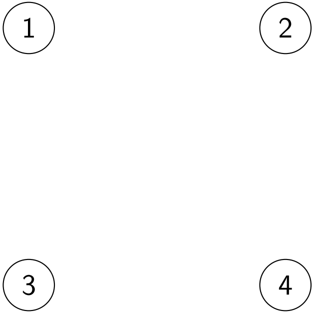
Example

Set of vertices V

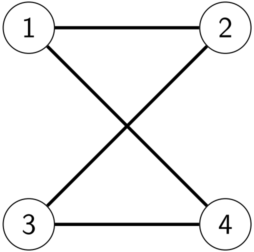


Example

Set of vertices V

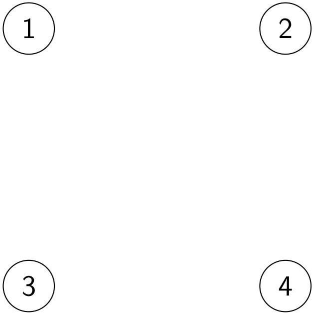


Graph G (e.g. assistance)

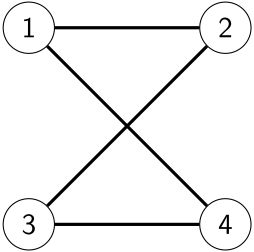


Example

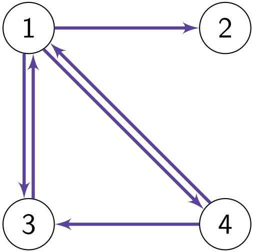
Set of vertices V



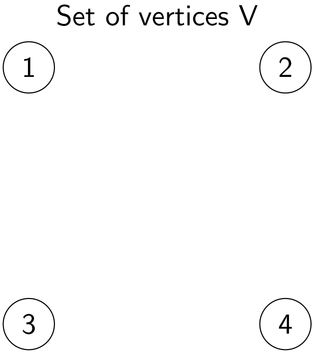
Graph G (e.g. assistance)

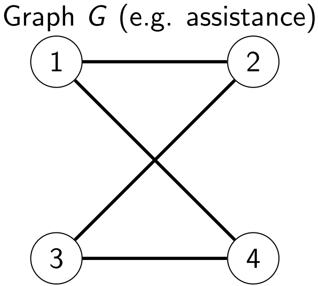


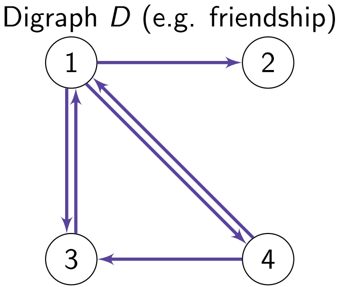
Digraph D (e.g. friendship)

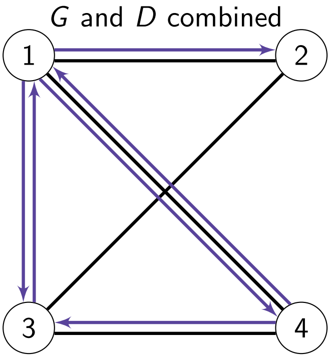


Example


Set of vertices V
1
2
3
4

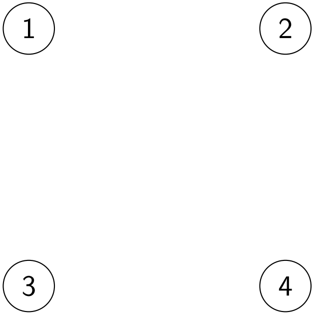

Graph G (e.g. assistance)
1
2
3
4


Digraph D (e.g. friendship)
1
2
3
4

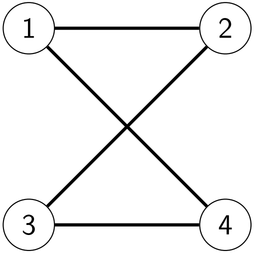

G and D combined
1
2
3
4

Example

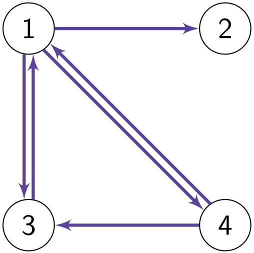
Set of vertices V



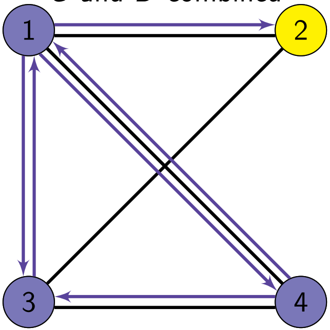
Graph G (e.g. assistance)



Digraph D (e.g. friendship)



G and D combined



Relational Data: dyadic relations I

- Given a set of n vertices labeled $1, \ldots, n$, we define the **set of pairs** for which the probability model applies:

$$\mathcal{N} \subset \mathcal{N}_0 = \{(i,j) \in \{1, \ldots, n\}^2 | i \neq j\}$$

Relational Data: dyadic relations I

- Given a set of n vertices labeled $1, \ldots, n$, we define the **set of pairs** for which the probability model applies:

$$\mathcal{N} \subset \mathcal{N}_0 = \{(i,j) \in \{1, \ldots, n\}^2 | i \neq j\}$$

Relational Data: dyadic relations I

- Given a set of n vertices labeled $1, \ldots, n$, we define the **set of pairs** for which the probability model applies:

$$\mathcal{N} \subset \mathcal{N}_0 = \{(i,j) \in \{1, \ldots, n\}^2 | i \neq j\}$$

$$e.g. \quad \mathcal{N} = \{(1,2), (1,3), (1,4), (2,3), (2,4), (3,4)\}$$

Relational Data: dyadic relations I

- Given a set of n vertices labeled $1, \ldots, n$, we define the **set of pairs** for which the probability model applies:

$$\mathcal{N} \subset \mathcal{N}_0 = \{(i,j) \in \{1, \ldots, n\}^2 | i \neq j\}$$

$$e.g. \quad \mathcal{N} = \{(1,2),(1,3),(1,4),(2,3),(2,4),(3,4)\}$$

We assume $\mathcal{N}$ to be symmetric i.e. $(i,j) \in \mathcal{N} \iff (j,i) \in \mathcal{N}$.

Relational Data: dyadic relations I

- Given a set of n vertices labeled $1, \ldots, n$, we define the **set of pairs** for which the probability model applies:

$$\mathcal{N} \subset \mathcal{N}_0 = \{(i,j) \in \{1, \ldots, n\}^2 | i \neq j\}$$

$$e.g. \quad \mathcal{N} = \{(1,2), (1,3), (1,4), (2,3), (2,4), (3,4)\}$$

  We assume $\mathcal{N}$ to be symmetric i.e. $(i,j) \in \mathcal{N} \iff (j,i) \in \mathcal{N}$.

- The relation from vertex i to vertex j takes a value from the the finite set:

$$\boldsymbol{\alpha} = \{a_1, a_2, \ldots, a_R\}$$

Relational Data: dyadic relations I

- Given a set of n vertices labeled $1, \ldots, n$, we define the **set of pairs** for which the probability model applies:

$$\mathcal{N} \subset \mathcal{N}_0 = \{(i,j) \in \{1, \ldots, n\}^2 | i \neq j\}$$

$$e.g. \quad \mathcal{N} = \{(1,2),(1,3),(1,4),(2,3),(2,4),(3,4)\}$$

We assume $\mathcal{N}$ to be symmetric i.e. $(i,j) \in \mathcal{N} \iff (j,i) \in \mathcal{N}$.

- The relation from vertex i to vertex j takes a value from the the finite set:

$$\boldsymbol{\alpha} = \{a_1, a_2, \ldots, a_R\}$$

Relational Data: dyadic relations I

- Given a set of n vertices labeled $1, \ldots, n$, we define the **set of pairs** for which the probability model applies:

$$\mathcal{N} \subset \mathcal{N}_0 = \{(i,j) \in \{1, \ldots, n\}^2 | i \neq j\}$$

$$e.g. \quad \mathcal{N} = \{(1,2),(1,3),(1,4),(2,3),(2,4),(3,4)\}$$

  We assume $\mathcal{N}$ to be symmetric i.e. $(i,j) \in \mathcal{N} \iff (j,i) \in \mathcal{N}$.

- The relation from vertex i to vertex j takes a value from the the finite set:

$$\boldsymbol{\alpha} = \{a_1, a_2, \ldots, a_R\}$$

$$e.g. \quad \boldsymbol{\alpha} = \{0, A, F, AF\} \quad R = 4$$

Relational Data: dyadic relations I

- Given a set of n vertices labeled $1, \dots, n$, we define the **set of pairs** for which the probability model applies:

$$\mathcal{N} \subset \mathcal{N}_0 = \{(i,j) \in \{1, \dots, n\}^2 | i \neq j\}$$

$$e.g. \quad \mathcal{N} = \{(1,2),(1,3),(1,4),(2,3),(2,4),(3,4)\}$$

  We assume $\mathcal{N}$ to be symmetric i.e. $(i,j) \in \mathcal{N} \iff (j,i) \in \mathcal{N}$.

- The relation from vertex i to vertex j takes a value from the the finite set:

$$\boldsymbol{\alpha} = \{a_1, a_2, \dots, a_R\}$$

$$e.g. \quad \boldsymbol{\alpha} = \{0, A, F, AF\} \quad R = 4$$

  The aims of our probability model is take into account the mutual dependence of relation from i to j and from j to i. We define the **set of dyadic relations**:

$$\boldsymbol{\alpha}^2 = \{\mathrm{a} = (a_t, a_v) | a_t, a_v \in \boldsymbol{\alpha}\}$$

Relational Data: dyadic relations II

- We call $\mathcal{A} \subset \alpha^2$ the **alphabet** of pairwise relations. For every $(i, j) \in \mathcal{N}$ exists $\mathrm{a} = (a_t, a_v) \in \mathcal{A}$ describing the dyadic relation of between i and j.

Relational Data: dyadic relations II

- We call $\mathcal{A} \subset \boldsymbol{\alpha}^2$ the **alphabet** of pairwise relations. For every $(i, j) \in \mathcal{N}$ exists $\mathrm{a} = (a_t, a_v) \in \mathcal{A}$ describing the dyadic relation of between i and j.
- An important operator on $\mathcal{A}$ is the **reflection operator**:

$$\pi(x, y) = (y, x)$$

Relational Data: dyadic relations II

- We call $\mathcal{A} \subset \alpha^2$ the **alphabet** of pairwise relations. For every $(i, j) \in \mathcal{N}$ exists $\mathrm{a} = (a_t, a_v) \in \mathcal{A}$ describing the dyadic relation of between i and j.
- An important operator on $\mathcal{A}$ is the **reflection operator**:

$$\pi(x, y) = (y, x)$$

  - $\mathcal{A}$ is **closed under reflection**

Relational Data: dyadic relations II

- We call $\mathcal{A} \subset \boldsymbol{\alpha}^2$ the **alphabet** of pairwise relations. For every $(i,j) \in \mathcal{N}$ exists $\mathrm{a} = (a_t, a_v) \in \mathcal{A}$ describing the dyadic relation of between i and j.
- An important operator on $\mathcal{A}$ is the **reflection operator**:

$$\pi(x, y) = (y, x)$$

  - $\mathcal{A}$ is **closed under reflection**
  - $\mathcal{A}_0 = \{\mathrm{a} \in \mathcal{A} | \pi(\mathrm{a}) = \mathrm{a}\}$ is the set of symmetric relations

Relational Data: dyadic relations II

- We call $\mathcal{A} \subset \boldsymbol{\alpha}^2$ the **alphabet** of pairwise relations. For every $(i,j) \in \mathcal{N}$ exists $\mathrm{a} = (a_t, a_v) \in \mathcal{A}$ describing the dyadic relation of between i and j.
- An important operator on $\mathcal{A}$ is the **reflection operator**:

$$\pi(x,y) = (y,x)$$

  - $\mathcal{A}$ is **closed under reflection**
  - $\mathcal{A}_0 = \{\mathrm{a} \in \mathcal{A} | \pi(\mathrm{a}) = \mathrm{a}\}$ is the set of symmetric relations
  - $\mathcal{A}_1 = \{\mathrm{a} \in \mathcal{A} | \pi(\mathrm{a}) \neq \mathrm{a}\}$ is the set of asymmetric relations

Relational Data: dyadic relations II

- We call $\mathcal{A} \subset \boldsymbol{\alpha}^2$ the **alphabet** of pairwise relations. For every $(i, j) \in \mathcal{N}$ exists $\mathrm{a} = (a_t, a_v) \in \mathcal{A}$ describing the dyadic relation of between i and j.
- An important operator on $\mathcal{A}$ is the **reflection operator**:

$$\pi(x, y) = (y, x)$$

  - $\mathcal{A}$ is **closed under reflection**
  - $\mathcal{A}_0 = \{\mathrm{a} \in \mathcal{A} | \pi(\mathrm{a}) = \mathrm{a}\}$ is the set of symmetric relations
  - $\mathcal{A}_1 = \{\mathrm{a} \in \mathcal{A} | \pi(\mathrm{a}) \neq \mathrm{a}\}$ is the set of asymmetric relations
  - $\mathcal{A} = \mathcal{A}_0 \cup \mathcal{A}_1$ and $\mathcal{A}_0 \cap \mathcal{A}_1 = \varnothing$

Relational Data: dyadic relations II

- We call $\mathcal{A} \subset \alpha^2$ the **alphabet** of pairwise relations. For every $(i, j) \in \mathcal{N}$ exists $\mathrm{a} = (a_t, a_v) \in \mathcal{A}$ describing the dyadic relation of between i and j.
- An important operator on $\mathcal{A}$ is the **reflection operator**:

$$\pi(x, y) = (y, x)$$

  - $\mathcal{A}$ is **closed under reflection**
  - $\mathcal{A}_0 = \{\mathrm{a} \in \mathcal{A} | \pi(\mathrm{a}) = \mathrm{a}\}$ is the set of symmetric relations
  - $\mathcal{A}_1 = \{\mathrm{a} \in \mathcal{A} | \pi(\mathrm{a}) \neq \mathrm{a}\}$ is the set of asymmetric relations
  - $\mathcal{A} = \mathcal{A}_0 \cup \mathcal{A}_1$ and $\mathcal{A}_0 \cap \mathcal{A}_1 = \varnothing$
  - To avoid redundancy, $\mathcal{A}_1$ can be partitioned into 2 disjoint half subsets such that $\mathcal{A}_1 = \mathcal{A}_{10} \cup \mathcal{A}_{11}$, $\mathcal{A}_{10} \cap \mathcal{A}_{11} = \varnothing$ and $\pi(\mathcal{A}_{10}) = \mathcal{A}_{11}$

Relational Data: dyadic relations II

- We call $\mathcal{A} \subset \boldsymbol{\alpha}^2$ the **alphabet** of pairwise relations. For every $(i,j) \in \mathcal{N}$ exists $\mathrm{a} = (a_t, a_v) \in \mathcal{A}$ describing the dyadic relation of between i and j.
- An important operator on $\mathcal{A}$ is the **reflection operator**:

$$\pi(x,y) = (y,x)$$

  - $\mathcal{A}$ is **closed under reflection**
  - $\mathcal{A}_0 = \{\mathrm{a} \in \mathcal{A} | \pi(\mathrm{a}) = \mathrm{a}\}$ is the set of symmetric relations
  - $\mathcal{A}_1 = \{\mathrm{a} \in \mathcal{A} | \pi(\mathrm{a}) \neq \mathrm{a}\}$ is the set of asymmetric relations
  - $\mathcal{A} = \mathcal{A}_0 \cup \mathcal{A}_1$ and $\mathcal{A}_0 \cap \mathcal{A}_1 = \varnothing$
  - To avoid redundancy, $\mathcal{A}_1$ can be partitioned into 2 disjoint half subsets such that $\mathcal{A}_1 = \mathcal{A}_{10} \cup \mathcal{A}_{11}$, $\mathcal{A}_{10} \cap \mathcal{A}_{11} = \varnothing$ and $\pi(\mathcal{A}_{10}) = \mathcal{A}_{11}$
  - We define $\mathcal{A}' = \mathcal{A}_0 \cup \mathcal{A}_{10}$

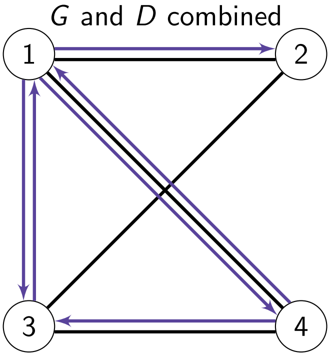
Relational Data: dyadic relations II

- We call $\mathcal{A} \subset \boldsymbol{\alpha}^2$ the **alphabet** of pairwise relations. For every $(i,j) \in \mathcal{N}$ exists $\mathrm{a} = (a_t, a_v) \in \mathcal{A}$ describing the dyadic relation of between i and j.
- An important operator on $\mathcal{A}$ is the **reflection operator**:

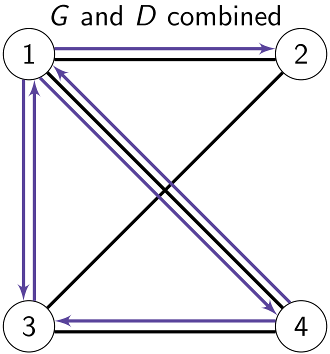
$$\pi(x,y) = (y,x)$$

  - $\mathcal{A}$ is **closed under reflection**
  - $\mathcal{A}_0 = \{\mathrm{a} \in \mathcal{A} | \pi(\mathrm{a}) = \mathrm{a}\}$ is the set of symmetric relations
  - $\mathcal{A}_1 = \{\mathrm{a} \in \mathcal{A} | \pi(\mathrm{a}) \neq \mathrm{a}\}$ is the set of asymmetric relations
  - $\mathcal{A} = \mathcal{A}_0 \cup \mathcal{A}_1$ and $\mathcal{A}_0 \cap \mathcal{A}_1 = \varnothing$
  - To avoid redundancy, $\mathcal{A}_1$ can be partitioned into 2 disjoint half subsets such that $\mathcal{A}_1 = \mathcal{A}_{10} \cup \mathcal{A}_{11}$, $\mathcal{A}_{10} \cap \mathcal{A}_{11} = \varnothing$ and $\pi(\mathcal{A}_{10}) = \mathcal{A}_{11}$
  - We define $\mathcal{A}' = \mathcal{A}_0 \cup \mathcal{A}_{10}$
  - We denote $r = |\mathcal{A}|$, $r_0 = |\mathcal{A}_0|$, $r_1 = \frac{1}{2}|\mathcal{A}_1| = |\mathcal{A}_{10}| = |\mathcal{A}_{11}|$

Example

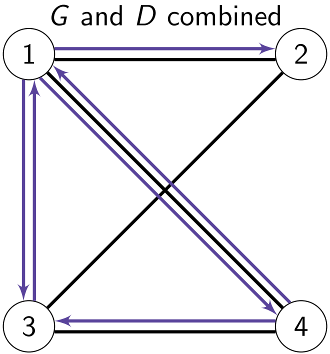


Example



$$\mathcal{N} = \{(1,2),(1,3),(1,4),(2,3),(2,4),(3,4)\}$$
$$\boldsymbol{\alpha} = \{0, A, F, AF\}$$

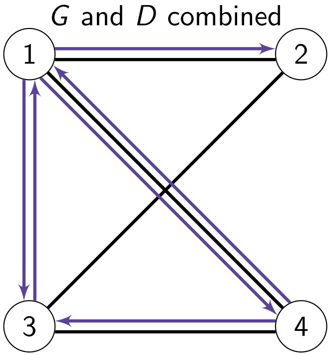
Example



$$\mathcal{N} = \{(1,2),(1,3),(1,4),(2,3),(2,4),(3,4)\}$$
$$\boldsymbol{\alpha} = \{0, A, F, AF\}$$

$$\mathcal{A}_0 = \{(0,0),(A,A),(F,F),(AF,AF)\}$$
$$\mathcal{A}_1 = \{(0,F),(F,0)(A,AF),(AF,A)\}$$
$$\mathcal{A}_{10} = \{(0,F),(A,AF)\}$$
$$\mathcal{A}_{11} = \{(F,0),(AF,A)\}$$

Example



$$\mathcal{N} = \{(1,2),(1,3),(1,4),(2,3),(2,4),(3,4)\}$$
$$\alpha = \{0, A, F, AF\}$$

$$\mathcal{A}_0 = \{(0,0),(A,A),(F,F),(AF,AF)\}$$
$$\mathcal{A}_1 = \{(0,F),(F,0)(A,AF),(AF,A)\}$$
$$\mathcal{A}_{10} = \{(0,F),(A,AF)\}$$
$$\mathcal{A}_{11} = \{(F,0),(AF,A)\}$$

$$r = |\mathcal{A}| = 8$$
$$r_0 = |\mathcal{A}_0| = 4$$
$$r_1 = 1/2|\mathcal{A}_1| = |\mathcal{A}_{10}| = |\mathcal{A}_{11}| = 2$$

Relational Data: Colored Relational Structure

The second aspect of the structure is a **discrete vertex characteristic**. The set of vertices $\{1, \dots, n\}$ is partitioned into c colors $1, \dots, c$ and we denote the set of colors $\mathcal{C} = \{1, \dots, c\}$.

Relational Data: Colored Relational Structure

The second aspect of the structure is a **discrete vertex characteristic**. The set of vertices $\{1, \dots, n\}$ is partitioned into c colors $1, \dots, c$ and we denote the set of colors $\mathcal{C} = \{1, \dots, c\}$.

Definition

It is called **colored relational structure** a set $\mathcal{N}$ of ordered pairs of vertices between which relations are given, with the vertices belonging to c categories stored in the vector $\mathbf{x} = (x_i)_{i=1}^n$ with $x_i \in \mathcal{C}$ and with a relational structure where the dyadic relations $\mathbf{y} = (y_{ij})_{(i,j)\in\mathcal{N}}$ take values in $\mathcal{A}$.
We consider a stochastic setting and we denote the random adjacency matrix and the random vector of colors by $(\mathbf{Y}, \mathbf{X})$.

Relational Data: Colored Relational Structure

The second aspect of the structure is a **discrete vertex characteristic**. The set of vertices $\{1, \ldots, n\}$ is partitioned into c colors $1, \ldots, c$ and we denote the set of colors $\mathcal{C} = \{1, \ldots, c\}$.

Definition

It is called **colored relational structure** a set $\mathcal{N}$ of ordered pairs of vertices between which relations are given, with the vertices belonging to c categories stored in the vector $\mathbf{x} = (x_i)_{i=1}^n$ with $x_i \in \mathcal{C}$ and with a relational structure where the dyadic relations $\mathbf{y} = (y_{ij})_{(i,j)\in\mathcal{N}}$ take values in $\mathcal{A}$.
We consider a stochastic setting and we denote the random adjacency matrix and the random vector of colors by $(\mathbf{Y}, \mathbf{X})$.

In our particular framework, we will consider $\mathbf{Y} = \mathbf{y}$ observed, but $\mathbf{X}$ unknown (posteriori blockmodelling).

Table of Contents

Stochastic blockmodel: $P(X)$

The ultimate goal of our analysis is specify a probability model for colored relational structures. In particular we are interested in:

- The predictive density $P(\mathbf{X}|\mathbf{y})$
- Estimate the vector of parameters of the models.

Stochastic blockmodel: $P(X)$

The ultimate goal of our analysis is specify a probability model for colored relational structures. In particular we are interested in:

- The predictive density $P(\mathbf{X}|\mathbf{y})$
- Estimate the vector of parameters of the models.

Assume that the random colors X_i are iid random variables with probability:

$$P(X_i = k) = \theta_k \quad k \in \mathcal{C}$$

Stochastic blockmodel: $P(X)$

The ultimate goal of our analysis is specify a probability model for colored relational structures. In particular we are interested in:

- The predictive density $P(\mathbf{X}|\mathbf{y})$
- Estimate the vector of parameters of the models.

Assume that the random colors X_i are iid random variables with probability:

$$P(X_i = k) = \theta_k \quad k \in \mathcal{C}$$

and the joint distribution of $\mathbf{X}$ to be:

$$P(X = x) = \theta_1^{m_1} \dots \theta_c^{m_c}$$
$$m_k = \sum_{i=1}^{n} I(x_i = k)$$

The vector $(m_k)_{k \in \mathcal{C}}$ is **sufficient statistic** for the probability law of X.

Stochastic blockmodel: $P(Y|x)$ I

Given the vector of colors $\mathbf{X} = \mathbf{x}$, the random vectors $\mathbf{Y}_{ij}$ for $(i, j) \in \mathcal{N}$ with $i < j$ are **independent** with probabilities:

$$P(\mathbf{Y}_{ij} = \mathrm{a} | \mathbf{X} = \mathbf{x}) = \eta_{\mathrm{a}}(x_i, x_j) \quad \text{for} \;\; \mathrm{a} \in \mathcal{A} \text{ and } x_i, x_j \in \mathcal{C}$$

Stochastic blockmodel: $P(Y|x)$ I

Given the vector of colors $\mathbf{X} = \mathbf{x}$, the random vectors $\mathbf{Y}_{ij}$ for $(i, j) \in \mathcal{N}$ with $i < j$ are **independent** with probabilities:

$$P(\mathbf{Y}_{ij} = \mathrm{a} | \mathbf{X} = \mathbf{x}) = \eta_{\mathrm{a}}(x_i, x_j) \quad \text{for} \;\; \mathrm{a} \in \mathcal{A} \text{ and } x_i, x_j \in \mathcal{C}$$

$$e.g. \quad \eta_{(AF,A)}(x_2 = blue, x_3 = red)$$

Stochastic blockmodel: $P(Y|x)$ I

Given the vector of colors $\mathbf{X} = \mathbf{x}$, the random vectors $\mathbf{Y}_{ij}$ for $(i,j) \in \mathcal{N}$ with $i < j$ are **independent** with probabilities:

$$P(\mathbf{Y}_{ij} = \mathrm{a} | \mathbf{X} = \mathbf{x}) = \eta_{\mathrm{a}}(x_i, x_j) \quad \text{for } \ \mathrm{a} \in \mathcal{A} \text{ and } x_i, x_j \in \mathcal{C}$$

$$e.g. \quad \eta_{(AF,A)}(x_2 = blue, x_3 = red)$$

and the array $\boldsymbol{\eta}$ of **color-dependent dyad probabilities** satisfies:

$$\sum_{\mathrm{a} \in \mathcal{A}} \eta_{\mathrm{a}}(k, h) = 1 \quad \text{for all } k, h \in \mathcal{C}$$

Stochastic blockmodel: $P(Y|x)$ I

Given the vector of colors $\mathbf{X} = \mathbf{x}$, the random vectors $\mathbf{Y}_{ij}$ for $(i,j) \in \mathcal{N}$ with $i < j$ are **independent** with probabilities:

$$P(\mathbf{Y}_{ij} = \mathrm{a} | \mathbf{X} = \mathbf{x}) = \eta_{\mathrm{a}}(x_i, x_j) \quad \text{for} \;\; \mathrm{a} \in \mathcal{A} \text{ and } x_i, x_j \in \mathcal{C}$$

$$e.g. \quad \eta_{(AF,A)}(x_2 = blue, x_3 = red)$$

and the array $\boldsymbol{\eta}$ of **color-dependent dyad probabilities** satisfies:

$$\sum_{\mathrm{a} \in \mathcal{A}} \eta_{\mathrm{a}}(k,h) = 1 \quad \text{for all } k,h \in \mathcal{C}$$

Because $y_{ij} = \pi(y_{ji})$, the probabilities $\boldsymbol{\eta}$ must be **invariant** with respect to the reflection operator :

$$\eta_{\mathrm{a}}(k,h) = \eta_{\pi(\mathrm{a})}(h,k)$$

Stochastic blockmodel: $P(Y|x)$ I

Given the vector of colors $\mathbf{X} = \mathbf{x}$, the random vectors $\mathbf{Y}_{ij}$ for $(i,j) \in \mathcal{N}$ with $i < j$ are **independent** with probabilities:

$$P(\mathbf{Y}_{ij} = \mathrm{a} | \mathbf{X} = \mathbf{x}) = \eta_{\mathrm{a}}(x_i, x_j) \quad \text{for} \quad \mathrm{a} \in \mathcal{A} \text{ and } x_i, x_j \in \mathcal{C}$$

$$e.g. \quad \eta_{(AF,A)}(x_2 = blue, x_3 = red)$$

and the array $\boldsymbol{\eta}$ of **color-dependent dyad probabilities** satisfies:

$$\sum_{\mathrm{a} \in \mathcal{A}} \eta_{\mathrm{a}}(k,h) = 1 \quad \text{for all } k,h \in \mathcal{C}$$

Because $y_{ij} = \pi(y_{ji})$, the probabilities $\boldsymbol{\eta}$ must be **invariant** with respect to the reflection operator :

$$\eta_{\mathrm{a}}(k,h) = \eta_{\pi(\mathrm{a})}(h,k)$$

which implies a **redundancy in parameters** $\eta_a(k,h)$.

Stochastic blockmodel: $P(Y|x)$ II

A non-redundant parameterization is obtained as follows:

$$\begin{cases} \eta_{\mathrm{a}}(k,h) & \text{for a} \in \mathcal{A} \text{ and } k < h \\ \eta_{\mathrm{a}}(k,k) & \text{for a} \in \mathcal{A}' \text{ and } k \in \mathcal{C} \end{cases}$$

Stochastic blockmodel: $P(Y|x)$ II

A non-redundant parameterization is obtained as follows:

$$\begin{cases} \eta_{\mathrm{a}}(k,h) & \text{for } \mathrm{a} \in \mathcal{A} \text{ and } k < h \\ \eta_{\mathrm{a}}(k,k) & \text{for } \mathrm{a} \in \mathcal{A}' \text{ and } k \in \mathcal{C} \end{cases}$$

The **conditional joint distribution** $P(\mathbf{y}|\mathbf{x})$ is given by:

$$P(y|\mathbf{x},\boldsymbol{\theta},\boldsymbol{\eta}) = \left(\prod_{\mathrm{a}\in\mathcal{A}} \prod_{1\leq k<h\leq c} (\eta_{\mathrm{a}}(k,h))^{e_{\mathrm{a}}(k,h)}\right) \times \left(\prod_{\mathrm{a}\in\mathcal{A}'} \prod_{k=1}^{c} (\eta_{\mathrm{a}}(k,k))^{e_{\mathrm{a}}(k,k)}\right)$$

where $e_{\mathrm{a}}(k,h)$ counts the number of relations of type a from vertices of color k to vertices of color h and is a **sufficient statistic**.

Stochastic blockmodel

The stochastic blockmodel is then given by the joint distribution of $(\mathbf{Y}, \mathbf{X})$:

$$
\begin{aligned}
P(\mathbf{y}, \mathbf{x} | \boldsymbol{\theta}, \boldsymbol{\eta}) &= P(\mathbf{x}) \times P(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\eta}) \\
&= \theta_1^{m_1} \times \cdots \times \theta_c^{m_c} \\
&\quad \times \left(\prod_{\mathrm{a} \in \mathcal{A}} \prod_{1 \le k < h \le c} (\eta_{\mathrm{a}}(k, h)^{e_{\mathrm{a}}(k,h)}) \right) \\
&\quad \times \left(\prod_{\mathrm{a} \in \mathcal{A}'} \prod_{k=1}^{c} (\eta_{\mathrm{a}}(k, k))^{e_{\mathrm{a}}(k,k)} \right)
\end{aligned}
$$

Stochastic blockmodel

The stochastic blockmodel is then given by the joint distribution of $(\mathbf{Y}, \mathbf{X})$:

$$
\begin{aligned}
P(\mathbf{y}, \mathbf{x} | \boldsymbol{\theta}, \boldsymbol{\eta}) &= P(\mathbf{x}) \times P(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\eta}) \\
&= \theta_1^{m_1} \times \cdots \times \theta_c^{m_c} \\
&\quad \times \left(\prod_{\mathrm{a} \in \mathcal{A}} \prod_{1 \le k < h \le c} (\eta_{\mathrm{a}}(k,h)^{e_{\mathrm{a}}(k,h)}) \right) \\
&\quad \times \left(\prod_{\mathrm{a} \in \mathcal{A}'} \prod_{k=1}^{c} (\eta_{\mathrm{a}}(k,k))^{e_{\mathrm{a}}(k,k)} \right)
\end{aligned}
$$

Because our model consider a stochastic structure $(\mathbf{Y}, \mathbf{X})$, with $\mathbf{Y}$ observed and $\mathbf{X}$ latent variable, the probability of observing edge pattern $\mathbf{y}$ is:

$$
P(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\eta}) = \sum_{\mathbf{x} \in \mathcal{C}^n} P(\mathbf{y}, \mathbf{x} | \boldsymbol{\theta}, \boldsymbol{\eta})
$$

Bayesian posteriori blockmodelling

From the stochastic blockmodel:

Bayesian posteriori blockmodelling

From the stochastic blockmodel:

- We assume a prior density $f(\boldsymbol{\theta}, \boldsymbol{\eta})$ for the parameters $(\boldsymbol{\theta}, \boldsymbol{\eta})$, which we can update using $\mathbf{y}$ and **inference about the parameters**:

$$f(\boldsymbol{\theta}, \boldsymbol{\eta} \mid \mathbf{y}) = \sum_{x} f(\boldsymbol{\theta}, \boldsymbol{\eta}, \mathbf{x} \mid \mathbf{y})$$

Bayesian posteriori blockmodelling

From the stochastic blockmodel:

- We assume a prior density $f(\boldsymbol{\theta}, \boldsymbol{\eta})$ for the parameters $(\boldsymbol{\theta}, \boldsymbol{\eta})$, which we can update using $\mathbf{y}$ and **inference about the parameters**:

$$f(\boldsymbol{\theta}, \boldsymbol{\eta} | \mathbf{y}) = \sum_{x} f(\boldsymbol{\theta}, \boldsymbol{\eta}, \mathbf{x} | \mathbf{y})$$

- The **recovery of the block structure** is based on the posterior predictive distribution of of **X**:

$$P(x | y) = \int f(\boldsymbol{\theta}, \boldsymbol{\eta}, \mathbf{x} | \mathbf{y}) d\boldsymbol{\theta} d\boldsymbol{\eta}$$

Bayesian posteriori blockmodelling

From the stochastic blockmodel:

- We assume a prior density $f(\boldsymbol{\theta}, \boldsymbol{\eta})$ for the parameters $(\boldsymbol{\theta}, \boldsymbol{\eta})$, which we can update using **y** and **inference about the parameters**:

$$f(\boldsymbol{\theta}, \eta | \mathbf{y}) = \sum_{x} f(\boldsymbol{\theta}, \boldsymbol{\eta}, \mathbf{x} | \mathbf{y})$$

- The **recovery of the block structure** is based on the posterior predictive distribution of of **X**:

$$P(x | y) = \int f(\boldsymbol{\theta}, \boldsymbol{\eta}, \mathbf{x} | \mathbf{y}) d\boldsymbol{\theta} d\boldsymbol{\eta}$$

$\Longrightarrow$ **Gibbs sampling** is used to obtain the conditional distribution $f(\boldsymbol{\theta}, \boldsymbol{\eta}, \mathbf{x} | \mathbf{y})$, which we need to solve both problems.

Table of Contents

Identifiability

The parameters $(\boldsymbol{\theta}, \boldsymbol{\eta})$ in the joint distribution $f(\boldsymbol{\theta}, \boldsymbol{\eta}, \mathbf{x}, \mathbf{y})$ **are not identifiable**. What matters is the partition defined by $\mathbf{x}$, not the colors labels $1, \ldots, c$.

Identifiability

The parameters $(\boldsymbol{\theta}, \boldsymbol{\eta})$ in the joint distribution $f(\boldsymbol{\theta}, \boldsymbol{\eta}, \mathbf{x}, \mathbf{y})$ **are not identifiable**. What matters is the partition defined by $\mathbf{x}$, not the colors labels $1, \ldots, c$.

In particular, let $\mathcal{S}$ denote the group of permutation of $\{1, \ldots, c\}$ and define $h_s(\cdot)$ and $s(\cdot)$ such that:

$$P(\mathbf{Y} = \mathbf{y} | \, \boldsymbol{\theta}, \boldsymbol{\eta}) = P(\mathbf{Y} = \mathbf{y} | \, h_s(\boldsymbol{\theta}, \boldsymbol{\eta}))$$
$$s(\mathbf{x}) = (s(x_1), \ldots, s(x_n))$$

for every permutation $s \in \mathcal{S}$.

Identifiability

The parameters $(\boldsymbol{\theta}, \boldsymbol{\eta})$ in the joint distribution $f(\boldsymbol{\theta}, \boldsymbol{\eta}, \mathbf{x}, \mathbf{y})$ **are not identifiable**. What matters is the partition defined by $\mathbf{x}$, not the colors labels $1, \ldots, c$.

In particular, let $\mathcal{S}$ denote the group of permutation of $\{1, \ldots, c\}$ and define $h_s(\cdot)$ and $s(\cdot)$ such that:

$$P(\mathbf{Y} = \mathbf{y} | \, \boldsymbol{\theta}, \boldsymbol{\eta}) = P(\mathbf{Y} = \mathbf{y} | \, h_s(\boldsymbol{\theta}, \boldsymbol{\eta}))$$
$$s(\mathbf{x}) = (s(x_1), \ldots, s(x_n))$$

for every permutation $s \in \mathcal{S}$. The transformation preserves the partition of the vertices and it implies:

$$P(\mathbf{x} | \, \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\eta}) = P(s(\mathbf{x}) | \, \mathbf{y}, h_s(\boldsymbol{\theta}, \boldsymbol{\eta}))$$

Invariance I

Because the model is invariant under the relabelling of the colors, **when the prior distribution is invariant, then the posterior is invariant**:

$$P(X_i = k|\,\mathbf{y}) = 1/c$$

independently from k.

Invariance I

Because the model is invariant under the relabelling of the colors, **when the prior distribution is invariant, then the posterior is invariant**:

$$P(X_i = k | \mathbf{y}) = 1/c$$

independently from k. We can deal with the invariance problem in some ways:

Invariance I

Because the model is invariant under the relabelling of the colors, **when the prior distribution is invariant, then the posterior is invariant**:

$$P(X_i = k|\,\mathbf{y}) = 1/c$$

independently from k. We can deal with the invariance problem in some ways:

1. Put **restrictions** on the parameters

$$e.g. \quad \theta_1 < \theta_2 \cdots < \theta_c$$

   However this approach is not always applicable and it may require some prior information (informative prior).

Invariance II

2. Consider the **posterior distribution of functions of** $(\theta, \eta, \mathbf{X})$ **that are invariant** with respect to relabelling.

Invariance II

2. Consider the **posterior distribution of functions of** $(\theta, \eta, \mathbf{X})$ **that are invariant** with respect to relabelling. In particular we consider:
   - $I\{X_i = X_j\}$

Invariance II

2. Consider the **posterior distribution of functions of** $(\theta, \eta, \mathbf{X})$ **that are invariant** with respect to relabelling. In particular we consider:
   - $I\{X_i = X_j\}$
   - $\eta_a(X_i, X_j) = \sum_{1 \leq k, h \leq c} \eta_a(k, h) I\{X_i = k, X_j = h\}$

Invariance II

2. Consider the **posterior distribution of functions of $(\boldsymbol{\theta}, \boldsymbol{\eta}, \mathbf{X})$ that are invariant** with respect to relabelling. In particular we consider:
   - $I\{X_i = X_j\}$
   - $\eta_a(X_i, X_j) = \sum_{1 \le k, h \le c} \eta_a(k, h) I\{X_i = k, X_j = h\}$

which are invariant w.r.t. the transformations $h_s(\boldsymbol{\theta}, \boldsymbol{\eta})$ and $s(\mathbf{X})$.

Invariance II

2. Consider the **posterior distribution of functions of** $(\boldsymbol{\theta}, \boldsymbol{\eta}, \mathbf{X})$ **that are invariant** with respect to relabelling. In particular we consider:

   - $I\{X_i = X_j\}$
   - $\eta_a(X_i, X_j) = \sum_{1 \le k,h \le c} \eta_a(k,h) I\{X_i = k, X_j = h\}$

   which are invariant w.r.t. the transformations $h_s(\boldsymbol{\theta}, \boldsymbol{\eta})$ and $s(\mathbf{X})$.

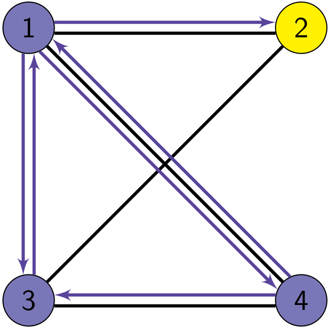
   We can apply the expectation to these functions and find the following matrix and the three-way array :

$$\left(P(X_i = X_j | \, \mathbf{y}) \right)_{1 \le i \ne j \le n}$$

$$\left(\mathbb{E}(\eta_a(X_i, X_j)|y \right)_{1 \le i \ne j \le n, \, a \in \mathcal{A}'}$$

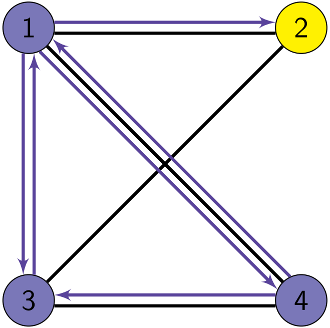
Example

G and D combined



Example

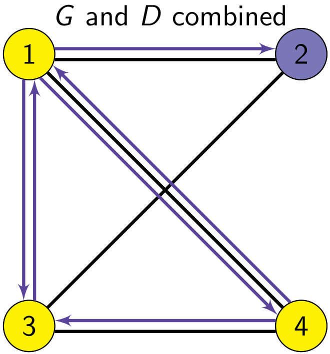
G and D combined



$$I\{X_i = X_j\}$$

X_i \ X_j	*1*	*2*	*3*	*4*
1	-	0	1	1
2	0	-	0	0
3	1	0	-	1
4	1	0	1	-

Example



$$I\{X_i = X_j\}$$

X_i \ X_j	*1*	*2*	*3*	*4*
1	-	0	1	1
2	0	-	0	0
3	1	0	-	1
4	1	0	1	-

Table of Contents

Gibbs Sampler

Definition

Gibbs Sampling is a simulation method used to approximate a target posterior distribution. It is an iterative simulation scheme that, given a set of unknown random vectors, consists in drawing each random vector from its probability law conditioned on the values of all of the other random vectors.

Gibbs Sampler

Definition

Gibbs Sampling is a simulation method used to approximate a target posterior distribution. It is an iterative simulation scheme that, given a set of unknown random vectors, consists in drawing each random vector from its probability law conditioned on the values of all of the other random vectors.

We apply this scheme to $((\boldsymbol{\theta}, \boldsymbol{\eta}), X_1, X_2, \ldots, X_n)$.

Gibbs Sampler

Definition

Gibbs Sampling is a simulation method used to approximate a target posterior distribution. It is an iterative simulation scheme that, given a set of unknown random vectors, consists in drawing each random vector from its probability law conditioned on the values of all of the other random vectors.

We apply this scheme to $((\boldsymbol{\theta}, \boldsymbol{\eta}), X_1, X_2, \ldots, X_n))$. At each iteration, given the current values $(\boldsymbol{\theta}^{(p)}, \boldsymbol{\eta}^{(p)}, \mathbf{X}^{(p)})$:

1. Draw $(\boldsymbol{\theta}^{(p+1)}, \boldsymbol{\eta}^{(p+1)})$ from $f(\boldsymbol{\theta}, \boldsymbol{\eta} | \, \mathbf{X}^{(p)}, \mathbf{y})$

Gibbs Sampler

Definition

Gibbs Sampling is a simulation method used to approximate a target posterior distribution. It is an iterative simulation scheme that, given a set of unknown random vectors, consists in drawing each random vector from its probability law conditioned on the values of all of the other random vectors.

We apply this scheme to $((\boldsymbol{\theta}, \boldsymbol{\eta}), X_1, X_2, \ldots, X_n))$. At each iteration, given the current values $(\boldsymbol{\theta}^{(p)}, \boldsymbol{\eta}^{(p)}, \mathbf{X}^{(p)})$:

1. Draw $(\boldsymbol{\theta}^{(p+1)}, \boldsymbol{\eta}^{(p+1)})$ from $f(\boldsymbol{\theta}, \boldsymbol{\eta} | \mathbf{X}^{(p)}, \mathbf{y})$
2. For each $i = 1, \ldots, n$: draw $X_i^{(p+1)}$ from $f(X_i | \boldsymbol{\theta}^{(p+1)}, \boldsymbol{\eta}^{(p+1)}, (X_j^{(p+1)})_{j<i}, (X_k^{(p)})_{k>j})$

Gibbs Sampler

Definition

Gibbs Sampling is a simulation method used to approximate a target posterior distribution. It is an iterative simulation scheme that, given a set of unknown random vectors, consists in drawing each random vector from its probability law conditioned on the values of all of the other random vectors.

We apply this scheme to $((\boldsymbol{\theta}, \boldsymbol{\eta}), X_1, X_2, \ldots, X_n))$. At each iteration, given the current values $(\boldsymbol{\theta}^{(p)}, \boldsymbol{\eta}^{(p)}, \mathbf{X}^{(p)})$:

1. Draw $(\boldsymbol{\theta}^{(p+1)}, \boldsymbol{\eta}^{(p+1)})$ from $f(\boldsymbol{\theta}, \boldsymbol{\eta} | \mathbf{X}^{(p)}, \mathbf{y})$
2. For each $i = 1, \ldots, n$: draw $X_i^{(p+1)}$ from $f(X_i | \boldsymbol{\theta}^{(p+1)}, \boldsymbol{\eta}^{(p+1)}, (X_j^{(p+1)})_{j<i}, (X_k^{(p)})_{k>j})$

We are left to specify the full conditional distribution for these two steps.

Step 1: Prior choice I

For step 1, we need to specify the **prior distribution** $f(\theta, \eta)$. If we do not have any particular prior information, it is also reasonable to assume **independence** between $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$.

Step 1: Prior choice I

For step 1, we need to specify the **prior distribution** $f(\boldsymbol{\theta}, \boldsymbol{\eta})$. If we do not have any particular prior information, it is also reasonable to assume **independence** between $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$.

- For $\boldsymbol{\theta}$ we use a **Dirichlet** $D(T_1, \dots, T_c)$.

Step 1: Prior choice I

For step 1, we need to specify the **prior distribution** $f(\boldsymbol{\theta}, \boldsymbol{\eta})$. If we do not have any particular prior information, it is also reasonable to assume **independence** between $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$.

- For $\boldsymbol{\theta}$ we use a **Dirichlet** $D(T_1, \ldots, T_c)$.
    - $T_k = 1$ corresponds to uniform distribution

Step 1: Prior choice I

For step 1, we need to specify the **prior distribution** $f(\theta, \eta)$. If we do not have any particular prior information, it is also reasonable to assume **independence** between θ and η.

- For θ we use a **Dirichlet** $D(T_1, \ldots, T_c)$.
  - $T_k = 1$ corresponds to uniform distribution
  - $T_k = T$ small when we think the true model has less classes than c

Step 1: Prior choice I

For step 1, we need to specify the **prior distribution** $f(\theta, \eta)$. If we do not have any particular prior information, it is also reasonable to assume **independence** between $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$.

- For $\boldsymbol{\theta}$ we use a **Dirichlet** $D(T_1, \ldots, T_c)$.
  - $T_k = 1$ corresponds to uniform distribution
  - $T_k = T$ small when we think the true model has less classes than c
  - $T_k = T$ large when we think the true model has c classes (e.g. $T = 100c$)

Step 1: Prior choice I

For step 1, we need to specify the **prior distribution** $f(\boldsymbol{\theta}, \boldsymbol{\eta})$. If we do not have any particular prior information, it is also reasonable to assume **independence** between $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$.

- For $\boldsymbol{\theta}$ we use a **Dirichlet** $D(T_1, \dots, T_c)$.
  - $T_k = 1$ corresponds to uniform distribution
  - $T_k = T$ small when we think the true model has less classes than c
  - $T_k = T$ large when we think the true model has c classes (e.g. $T = 100c$)

- For $\boldsymbol{\eta}$ use a **Dirichlet** of parameters $E_{\text{a}}(k, h)$.

Step 1: Prior choice I

For step 1, we need to specify the **prior distribution** $f(\boldsymbol{\theta}, \boldsymbol{\eta})$. If we do not have any particular prior information, it is also reasonable to assume **independence** between $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$.

- For $\boldsymbol{\theta}$ we use a **Dirichlet** $D(T_1, \dots, T_c)$.
  - $T_k = 1$ corresponds to uniform distribution
  - $T_k = T$ small when we think the true model has less classes than c
  - $T_k = T$ large when we think the true model has c classes (e.g. $T = 100c$)
- For $\boldsymbol{\eta}$ use a **Dirichlet** of parameters $E_{\mathrm{a}}(k, h)$.
  - For $k < h$, $\boldsymbol{\eta}(k, h) = (\eta_{\mathrm{a}}(k, h))_{\mathrm{a} \in \mathcal{A}}$ is an unconstrained probability vector of dimension $r = |\mathcal{A}|$.

Step 1: Prior choice II

- For $k = k$, $\boldsymbol{\eta}(k, k)$ is constrained to $\boldsymbol{\eta}_{\mathsf{a}}(k, k) = \boldsymbol{\eta}_{\pi(\mathsf{a})}(k, k)$.

Step 1: Prior choice II

- For $k = k$, $\boldsymbol{\eta}(k,k)$ is constrained to $\boldsymbol{\eta}_{\mathrm{a}}(k,k) = \boldsymbol{\eta}_{\pi(\mathrm{a})}(k,k)$.We define:

$$\boldsymbol{\eta}^{(0)}_{\mathrm{a}}(k,k) = \begin{cases} \boldsymbol{\eta}_{\mathrm{a}}(k,k) & \mathrm{a} \in \mathcal{A}_0 \\ 2\boldsymbol{\eta}_{\mathrm{a}}(k,k) & \mathrm{a} \in \mathcal{A}_{10} \end{cases}$$

transforming $\boldsymbol{\eta}(k,k)$ in a $(r_0 + r_1)$-dimensional (note $r_0 + r_1 < r$) vectors without redundant elements for every $k \in \mathcal{C}$.

Step 1: Prior choice II

- For $k = k$, $\boldsymbol{\eta}(k, k)$ is constrained to $\boldsymbol{\eta}_{\mathrm{a}}(k, k) = \boldsymbol{\eta}_{\pi(\mathrm{a})}(k, k)$.We define:

$$\boldsymbol{\eta}_{\mathrm{a}}^{(0)}(k, k) = \begin{cases} \boldsymbol{\eta}_{\mathrm{a}}(k, k) & \mathrm{a} \in \mathcal{A}_0 \\ 2\boldsymbol{\eta}_{\mathrm{a}}(k, k) & \mathrm{a} \in \mathcal{A}_{10} \end{cases}$$

transforming $\boldsymbol{\eta}(k, k)$ in a $(r_0 + r_1)$-dimensional (note $r_0 + r_1 < r$) vectors without redundant elements for every $k \in \mathcal{C}$.

- Set $E_{\mathrm{a}}(k, h) = 1$ for an invariant prior.

Step 1: Prior choice II

- For $k = k$, $\boldsymbol{\eta}(k,k)$ is constrained to $\boldsymbol{\eta}_{\mathsf{a}}(k,k) = \boldsymbol{\eta}_{\pi(\mathsf{a})}(k,k)$.We define:

$$\boldsymbol{\eta}^{(0)}_{\mathsf{a}}(k,k) = \begin{cases} \boldsymbol{\eta}_{\mathsf{a}}(k,k) & \mathsf{a} \in \mathcal{A}_0 \\ 2\boldsymbol{\eta}_{\mathsf{a}}(k,k) & \mathsf{a} \in \mathcal{A}_{10} \end{cases}$$

  transforming $\boldsymbol{\eta}(k,k)$ in a $(r_0 + r_1)$-dimensional (note $r_0 + r_1 < r$) vectors without redundant elements for every $k \in \mathcal{C}$.
- Set $E_{\mathsf{a}}(k,h) = 1$ for an invariant prior.

The posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\eta})$ given $(\mathbf{y}, \mathbf{x})$ is is largely tractable:

$$\begin{aligned} &(m_k + T_k)_{k \in \mathcal{C}} && \text{for } \boldsymbol{\theta} \\ &(e_{\mathsf{a}}(k,h) + E_{\mathsf{a}}(k,h))_{\mathsf{a} \in \mathcal{A}} && \text{for } \boldsymbol{\eta}(k,h),\ 1 \leq k < h \leq c \\ &(e_{\mathsf{a}}(k,k) + E_{\mathsf{a}}(k,k))_{\mathsf{a} \in \mathcal{A}'} && \text{for } \boldsymbol{\eta}^{(0)}(k,k),\ 1 \leq k \leq c \end{aligned}$$

Step 1: Prior choice II

- For $k = k$, $\boldsymbol{\eta}(k,k)$ is constrained to $\boldsymbol{\eta}_{\mathrm{a}}(k,k) = \boldsymbol{\eta}_{\pi(\mathrm{a})}(k,k)$.We define:

$$\boldsymbol{\eta}_{\mathrm{a}}^{(0)}(k,k) = \begin{cases} \boldsymbol{\eta}_{\mathrm{a}}(k,k) & \mathrm{a} \in \mathcal{A}_0 \\ 2\boldsymbol{\eta}_{\mathrm{a}}(k,k) & \mathrm{a} \in \mathcal{A}_{10} \end{cases}$$

  transforming $\boldsymbol{\eta}(k,k)$ in a $(r_0 + r_1)$-dimensional (note $r_0 + r_1 < r$) vectors without redundant elements for every $k \in \mathcal{C}$.
- Set $E_{\mathrm{a}}(k,h) = 1$ for an invariant prior.

The posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\eta})$ given $(\mathbf{y}, \mathbf{x})$ is is largely tractable:

$$\begin{aligned} (m_k + T_k)_{k \in \mathcal{C}} \quad & \text{for } \boldsymbol{\theta} \\ (e_{\mathrm{a}}(k,h) + E_{\mathrm{a}}(k,h))_{\mathrm{a} \in \mathcal{A}} \quad & \text{for } \boldsymbol{\eta}(k,h),\ 1 \le k < h \le c \\ (e_{\mathrm{a}}(k,k) + E_{\mathrm{a}}(k,k))_{\mathrm{a} \in \mathcal{A}'} \quad & \text{for } \boldsymbol{\eta}^{(0)}(k,k),\ 1 \le k \le c \end{aligned}$$

We can approximate $\boldsymbol{\eta}_{\mathrm{a}}(X_i, X_j)$ using the average of $\boldsymbol{\eta}_{\mathrm{a}}^{(p)}(X_i, X_j)$ drawn from the Gibbs sampler.

Step 2: $f(X_i|\boldsymbol{\theta}, \boldsymbol{\eta}, \mathbf{x}_{-i})$

From the probability law of the stochastic block-model $P(\mathbf{y}, \mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\eta})$ we can derive:

$$P(X_i = k|\, \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\eta}, \{X_j\}_{j\neq i}) = Q\theta_k \prod_{\mathrm{a}\in\mathcal{A}} \prod_{h=1}^{c} (\boldsymbol{\eta}_{\mathrm{a}}(k,h))^{d_{\mathrm{a}}(i,h)}$$

with $d_{\mathrm{a}}(i,k) = \sum_{j:(i,j)\in\mathcal{N}} I\{y_{ij} = \mathrm{a}\} I\{x_j = k\}$ counting number of dyadic relations of type a from vertex i to vertices of color k and with Q a constant not depending on k.

Step 2: $f(X_i|\boldsymbol{\theta},\boldsymbol{\eta},\mathbf{x}_{-i})$

From the probability law of the stochastic block-model $P(\mathbf{y},\mathbf{x}|\boldsymbol{\theta},\boldsymbol{\eta})$ we can derive:

$$P(X_i = k|\,\mathbf{y},\boldsymbol{\theta},\boldsymbol{\eta},\{X_j\}_{j\neq i}) = Q\theta_k \prod_{\mathrm{a}\in\mathcal{A}} \prod_{h=1}^{c} (\boldsymbol{\eta}_{\mathrm{a}}(k,h))^{d_{\mathrm{a}}(i,h)}$$

with $d_{\mathrm{a}}(i,k) = \sum_{j:(i,j)\in\mathcal{N}} I\{y_{ij} = \mathrm{a}\} I\{x_j = k\}$ counting number of dyadic relations of type a from vertex i to vertices of color k and with Q a constant not depending on k.

The **relative frequency** of $X_i^{(p)} = X_j^{(p)}$ over the Gibbs sampler can be used to **approximate the matrix of posterior predictive probabilities** $(P(X_i = X_j|\,\mathbf{y}))_{1\leq i\neq j\leq n}$

Adequacy of class structure

We can build some parameters for evaluating the adequacy of the obtained class structure:

Adequacy of class structure

We can build some parameters for evaluating the adequacy of the obtained class structure:

- I_y is the information of the observed relations with $c \geq 2$:

$$I_y = -\frac{2}{|\mathcal{N}|} \sum_{(i,j)\in\mathcal{N}, i<j} log(\eta_{y_{ij}}(X_i, X_j))$$

This value is small when $\boldsymbol{\eta}$ and $\mathbf{X}$ determine to a large extent of the observed relations.

Adequacy of class structure

We can build some parameters for evaluating the adequacy of the obtained class structure:

- I_y is the information of the observed relations with $c \geq 2$:

$$I_y = -\frac{2}{|\mathcal{N}|} \sum_{(i,j)\in\mathcal{N}, i<j} log(\eta_{y_{ij}}(X_i, X_j))$$

  This value is small when $\boldsymbol{\eta}$ and $\mathbf{X}$ determine to a large extent of the observed relations.

- H_x measures the extent to which the distribution of $\mathbf{X}$ defines one clear-cut partition of vertices into classes.

$$H_x = \frac{4}{n(n-1)} \sum_{i,j=1}^{n} \pi_{ij}(1 - \pi_{ij})$$

  with $\pi_{ij} = P(X_i = X_j | \, \mathbf{y})$. If H_x is small it relatively clear if pair of vertices are in the same class or not.

Example – Adequacy parameters

c	I_y	H_x
2	.94	.24
3	.91	.21
4	.89	.26
5	.89	.27

Figure 1: Parameter for the class structure for Kapferer's dataset

Convergence and class assignment

To check to convergence of the Gibbs sampler:

Convergence and class assignment

To check to convergence of the Gibbs sampler:

- Run multiple Gibbs sampler with different independent starting points and check that parameters H_x and I_y have similar values.

Convergence and class assignment

To check to convergence of the Gibbs sampler:

- Run multiple Gibbs sampler with different independent starting points and check that parameters H_x and I_y have similar values.
- For invariant prior, check that the the posterior predictive distribution of the sequence $\mathbf{X}^{(p)}$ is uniform.

Convergence and class assignment

To check to convergence of the Gibbs sampler:

- Run multiple Gibbs sampler with different independent starting points and check that parameters H_x and I_y have similar values.
- For invariant prior, check that the the posterior predictive distribution of the sequence $\mathbf{X}^{(p)}$ is uniform.

Assign vertices to classes:

Convergence and class assignment

To check to convergence of the Gibbs sampler:

- Run multiple Gibbs sampler with different independent starting points and check that parameters H_x and I_y have similar values.
- For invariant prior, check that the the posterior predictive distribution of the sequence $\mathbf{X}^{(p)}$ is uniform.

Assign vertices to classes:

- For uniform prior, arbitrarily assign labels.

Convergence and class assignment

To check to convergence of the Gibbs sampler:

- Run multiple Gibbs sampler with different independent starting points and check that parameters H_x and I_y have similar values.
- For invariant prior, check that the the posterior predictive distribution of the sequence $\mathbf{X}^{(p)}$ is uniform.

Assign vertices to classes:

- For uniform prior, arbitrarily assign labels.
- For nonuniform prior, labels are identified by the prior.

Example – Estimated posterior probabilities I

i	Posterior probability $P(X_i = X_j \mid \mathbf{y})$																																					
1	—	9	9	9	9	6	9	9	9	8	9	6	9	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
2	9	—	9	9	9	6	9	9	9	8	9	6	9	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
3	9	9	—	9	9	6	9	9	9	9	9	6	9	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
5	9	9	9	—	9	6	9	9	9	9	9	6	9	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
7	9	9	9	9	—	6	9	9	9	8	9	6	9	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
9	6	6	6	6	6	—	6	6	6	5	5	7	6	4	5	5	4	4	4	5	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0
12	9	9	9	9	9	6	—	9	9	8	9	6	9	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
13	9	9	9	9	9	6	9	—	9	9	9	6	9	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
14	9	9	9	9	9	6	9	9	—	8	9	6	9	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
16	8	8	9	9	8	5	8	9	8	—	9	5	9	9	7	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
19	9	9	9	9	9	5	9	9	9	9	—	5	9	9	7	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
21	6	6	6	6	6	7	6	6	6	5	5	—	6	4	7	6	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0
24	9	9	9	9	9	6	9	9	9	9	9	6	—	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
25	8	8	8	8	8	4	8	8	8	9	9	4	8	—	7	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
30	8	8	8	8	8	5	8	8	8	7	7	7	8	7	—	8	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0
34	7	7	7	7	7	5	7	7	7	6	6	6	7	6	8	—	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	0
4	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	—	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0
6	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	—	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0
8	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	—	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0
10	1	1	1	1	1	5	1	1	1	0	0	4	1	0	2	3	9	9	9	—	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0
15	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	—	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0
17	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	—	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0
18	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	—	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0
20	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	—	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0
22	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	—	9	9	9	9	9	9	9	9	9	9	9	9	9	0
23	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	—	9	9	9	9	9	9	9	9	9	9	9	9	0
26	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	—	9	9	9	9	9	9	9	9	9	9	9	0
27	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	—	9	9	9	9	9	9	9	9	9	9	0
28	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	—	9	9	9	9	9	9	9	9	9	0
29	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	—	9	9	9	9	9	9	9	9	0
31	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	—	9	9	9	9	9	9	9	0
32	1	1	1	1	1	4	1	1	1	1	1	4	1	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	—	9	9	9	9	9	9	0
33	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	—	9	9	9	9	9	0
35	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	—	9	9	9	9	0
36	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	—	9	9	9	0
37	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	—	9	9	0
38	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	—	9	0
39	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	—	0
11	0	0	0	0	0	0	0	0	0	1	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	—

Figure 2: Estimated posterior probabilities $P(X_i = X_j \mid \mathbf{y})$ for Kapferer's dataset

Example – Estimated posterior probabilities I

i	Posterior probability $P(X_i = X_j \mid \mathbf{y})$																																						
1	—	9	9	9	9	6	9	9	9	8	9	6	9	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
2	9	—	9	9	9	6	9	9	9	8	9	6	9	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
3	9	9	—	9	9	6	9	9	9	9	9	6	9	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
5	9	9	9	—	9	6	9	9	9	9	9	6	9	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
7	9	9	9	9	—	6	9	9	9	8	9	6	9	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
9	6	6	6	6	6	—	6	6	6	5	5	7	6	4	5	5	4	4	4	5	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0
12	9	9	9	9	9	6	—	9	9	8	9	6	9	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
13	9	9	9	9	9	6	9	—	9	9	9	6	9	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
14	9	9	9	9	9	6	9	9	—	8	9	6	9	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
16	8	8	9	9	8	5	8	9	8	—	9	5	9	9	7	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
19	9	9	9	9	9	5	9	9	9	9	—	5	9	9	7	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
21	6	6	6	6	6	7	6	6	6	5	5	—	6	4	7	6	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0
24	9	9	9	9	9	6	9	9	9	9	9	6	—	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
25	8	8	8	8	8	4	8	8	8	9	9	4	8	—	7	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
30	8	8	8	8	8	5	8	8	8	7	7	7	8	7	—	8	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0
34	7	7	7	7	7	5	7	7	7	6	6	6	7	6	8	—	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	0
4	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	—	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0
6	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	—	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0
8	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	—	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0
10	1	1	1	1	1	5	1	1	1	0	0	4	1	0	2	3	9	9	9	—	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0
15	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	—	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0
17	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	—	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0
18	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	—	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0
20	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	—	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0
22	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	—	9	9	9	9	9	9	9	9	9	9	9	9	9	0
23	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	—	9	9	9	9	9	9	9	9	9	9	9	9	0
26	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	—	9	9	9	9	9	9	9	9	9	9	9	0
27	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	—	9	9	9	9	9	9	9	9	9	9	0
28	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	—	9	9	9	9	9	9	9	9	9	0
29	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	—	9	9	9	9	9	9	9	9	0
31	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	—	9	9	9	9	9	9	9	0
32	1	1	1	1	1	4	1	1	1	1	1	4	1	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	—	9	9	9	9	9	9	0
33	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	—	9	9	9	9	9	0
35	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	—	9	9	9	9	0
36	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	—	9	9	9	0
37	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	—	9	9	0
38	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	—	9	0
39	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	—	0
11	0	0	0	0	0	0	0	0	0	1	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	—

Figure 2: Estimated posterior probabilities $P(X_i = X_j \mid \mathbf{y})$ for Kapferer's dataset

Example – Estimated posterior probabilities II

a	h	$k=1$	$k=2$	$k=3$
(0,0)	1	.27	.72	.14
	2	.72	.78	.31
	3	.14	.31	
(A,A)	1	.38	.19	.15
	2	.19	.15	.22
	3	.15	.22	
(F,F)	1	.03	.01	.07
	2	.01	.01	.04
	3	.07	.04	
(AF,AF)	1	.16	.03	.42
	2	.03	.05	.07
	3	.42	.07	
$(0,F)$	1	.01	.00	.05
	2	.03	.00	.27
	3	.04	.03	
(A,AF)	1	.07	.01	.08
	2	.02	.00	.04
	3	.05	.03	

Figure 3: Estimated posterior probabilities $\eta_a(X_i, X_j)$ for Kapferer's dataset

Bibliography I

Krzysztof Nowicki and Tom Snijders.
Estimation and prediction for stochastic blockstructures.
Journal of the American Statistical Association, 96, 02 2001.

Tom Snijders and Krzysztof Nowicki.
Estimation and prediction for stochastic block-structures for graphs with latent block structure.
Journal of Classification, 01 1997.