Web-based supporting materials for "A Stochastic Block Model for Multilevel Networks: Application to the Sociology of Organisations"

Saint-Clair Chabert-Liddell
Pierre Barbillon
Sophie Donnet
UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 75005 Paris, France
E-mail: saint-clair.chabert-liddell@agroparistech.fr
Emmanuel Lazega
Institut d'Études Politiques de Paris, France

A. Proof of Proposition 1

PROPOSITION 1. *In the MLVSBM, the two following properties are equivalent: [1.]: Z^I is independent on Z^O, [2.]: $\gamma_{kl} = \gamma_{kl'} \quad \forall l, l' \in \{1, \dots, Q_O\}$ and imply that: [3.]: X^I and X^O are independent.*

PROOF. We first derive an expression for $\ell_\gamma(Z^I) = \ell_\gamma(Z^I|A)$:

$$\begin{aligned}
\ell_\gamma(Z^I|A) &= \int_{Z^O} \ell_\gamma(Z^I|A, Z^O) d\mathbb{P}(Z^O) \\
&= \sum_{l_1,\dots,l_{n_O}} \ell_\gamma(Z^I|A, Z_1^O = l_1, \dots, Z_{n_O}^O = l_{n_O}) \mathbb{P}(Z_1^O = l_1, \dots, Z_{n_O}^O = l_{n_O}) \\
&= \sum_{l_1,\dots,l_{n_O}} \prod_j \left(\prod_i \ell_\gamma(Z_i^I|A, Z_{A_i}^O = l_{A_i}) \right) \mathbb{P}(Z_j^O = l_j) \\
&= \sum_{l_1,\dots,l_{n_O}} \prod_j \left(\prod_{i,k} \gamma_{kl_j}^{\mathbb{1}_{Z_i^I = k} A_{ij}} \right) \pi_{l_j}^O = \prod_j \sum_l \prod_{i,k} \gamma_{kl}^{A_{ij} \mathbb{1}_{Z_i^I = k}} \pi_l^O
\end{aligned}$$

where $A_i = \{j : A_{ij} = 1\}$.

2. $\Rightarrow$ 1.: Assume that $\gamma_{kl} = \gamma_{kl'} \quad \forall l, l' \in \{1, \dots, Q_O\}$, then:

$$\begin{aligned}
\ell_\gamma(Z^I|Z^O, A) &= \prod_{k,l} \gamma_{kl}^{\sum_{i,j} A_{ij} \mathbb{1}_{Z_i^I = k} \mathbb{1}_{Z_j^O = l}} = \prod_k \gamma_{k1}^{\sum_{i,j} A_{ij} \mathbb{1}_{Z_i^I = k} \sum_l \mathbb{1}_{Z_j^O = l}} \\
&= \prod_{i,k} \gamma_{k1}^{\mathbb{1}_{Z_i^I = k}},
\end{aligned}$$

and

$$\ell_\gamma(Z^I|A) = \prod_j \sum_l \prod_{i,k} \gamma_{kl}^{A_{ij}\mathbb{1}_{Z_i^I=k}} \pi_l^O$$
$$= \prod_j \prod_{i,k} \gamma_{k1}^{A_{ij}\mathbb{1}_{Z_i^I=k}} \sum_l \pi_l^O = \prod_{i,k} \gamma_{k1}^{\mathbb{1}_{Z_i^I=k}},$$

hence $\ell_\gamma(Z^I|Z^O, A) = \ell_\gamma(Z^I|A)$.

1. $\Rightarrow$ 2.: Assume that $\ell_\gamma(Z^I|Z^O, A) = \ell_\gamma(Z^I|A)$ for any values of Z^I, Z^O, then in particular $\ell_\gamma(Z_1^I|Z^O, A) = \ell_\gamma(Z_1^I|A)$. Assuming that individual 1 belongs to organisation j, we can write, for any k:

$$\mathbb{P}(Z_1^I = k|Z_j^O, A_{ij} = 1) = \gamma_{kZ_j^O}.$$

However, this quantity does not depend on Z_j^O so $\gamma_{kZ_j^O} = \gamma_{k\cdot}$ for any value of k and Z_j^O. And so we have $\gamma_{k\ell} = \gamma_{k\ell'}$ for any (ℓ, ℓ').

1. $\Rightarrow$ 3.:

$$\ell_{\alpha^I,\alpha^O}(X^I, X^O|A) = \int_{z^I,z^O} \ell_{\alpha^I,\alpha^O}(X^I, X^O|A, Z^I = z^I, Z^O = z^O)\mathbb{P}(Z^I = z^I, Z^O = z^O)\mathrm{d}z^I\mathrm{d}z^O$$
$$= \int_{z^I,z^O} \ell_{\alpha^I}(X^I|Z^I = z^r)\mathbb{P}(Z^I = z^I|A, Z^O = z^O)\ell_{\alpha^O}(X^O|Z^O = z^O)\mathbb{P}(Z^O = z^O))\mathrm{d}z^I\mathrm{d}z^O$$
$$= \int_{z^I} \ell_{\alpha^I}(X^I|Z^I = z^r)\mathbb{P}(Z^I = z^I)\mathrm{d}z^I \int_{z^O} \ell_{\alpha^O}(X^O|Z^O = z^O)\mathbb{P}(Z^O = z^O))\mathrm{d}z^O$$
$$= \ell_{\alpha^I}(X^I)\ell_{\alpha^O}(X^O)$$

which is the definition of the independence.

B. Proof of Proposition 2

PROPOSITION 2. *The stochastic block model for multilevel networks is identifiable up to label switching under the following assumptions:*

$\mathcal{A}1$. *All coefficients of $\alpha^I \cdot \gamma \cdot \pi^O$ are distinct and all coefficients of $\alpha^O \cdot \pi^O$ are distinct.*

$\mathcal{A}2$. *$n_I \geq 2Q_I$ and $n_O \geq \max(2Q_O, Q_O + Q_I - 1)$.*

$\mathcal{A}3$. *At least $2Q_I$ organizations contain one individual or more.*

PROOF. Let $\theta = \{\pi^O, \gamma, \alpha^I, \alpha^O\}$ be the set of parameters and $\mathbb{P}_X$ the distribution of the observed data. We will prove that there is a unique θ corresponding to $\mathbb{P}_X$. More precisely, in what follows, we will compute the probabilities of some particular events, from which we will derive a unique expression for the unknown parameters. The beginning of the proof –identifiability of π^O and α^O– is mimicking the one given in Celisse et al. (2012). The last steps of the proof are original work.

Notations. For the sake of simplicity, in what follows, we use the following shorten notation:

$$x_{i:k} := (x_i, \dots, x_k), \quad X_{j,i:k} = (X_{ji}, \dots, X_{jk}) .$$

Moreover, $\{X_{j,i:k} = 1\}$ stands for $\{X_{ji} = 1, \dots, X_{jk} = 1\}$.

Identifiability of π^O. For any $l = 1, \dots, Q_O$, let τ_l be the following probability:

$$\tau_l = \mathbb{P}(X^O_{ij} = 1 | Z^O_i = l) = \sum_{l'} \alpha^O_{ll'} \pi^O_{l'} = (\alpha^O \cdot \pi^O)_l, \quad \forall (i,j). \tag{B.1}$$

Moreover, a quick computation proves that

$$\mathbb{P}(X^O_{i,j:(j+k)} = 1 | Z^O_i = l) = \tau_l^{k+1} \tag{B.2}$$

According to Assumption $\mathcal{A}1$, the coordinates of vector $(\tau_1, \dots, \tau_{Q_O})$ are all different. Hence, the Vandermonde matrix R^O of size $Q_O \times Q_O$ such that

$$R^O_{il} = (\tau_l)^{i-1}, \quad 1 \leq i \leq Q_O, \quad 1 \leq l \leq Q_O$$

is invertible. We define u^O_i as follows :

$$\begin{aligned} u^O_i &= \mathbb{P}_{\mathbf{X},\theta}(X^O_{1,2:(i+1)} = 1) \quad \text{for } 1 \leq i \leq 2Q_O - 1 \\ u^O_0 &= 1. \end{aligned}$$

The existence of $(u^O_i)_{i=0,\dots,2Q_O-1}$ comes from Assumption $\mathcal{A}2$ ($n_O \geq 2Q_O$). Moreover, the $(u^O_i)_{i=0,\dots,2Q_O-1}$ are calculated from the marginal distribution $\mathbb{P}_X$. We will use these quantities to identify the parameters (π^O, α^O).

First we have, for $1 \leq i \leq 2Q_O - 1$:

$$u^O_i = \sum_{l=1}^{Q_O} \mathbb{P}(X^O_{1,2:(i+1)} = 1 | Z^O_1 = l) \mathbb{P}(Z^O_1 = l) = \sum_{l=1}^{Q_O} \tau_l^i \pi^O_l,$$

using equation (B.2). Now, let us define M^O a $(Q_O + 1) \times Q_O$ matrix such that:

$$M^O_{ij} = u^O_{i+j-2} = \sum_{l=1}^{Q_O} \tau_l^{i-1} \pi^O_l \tau_l^{j-1}, \quad 1 \leq i \leq Q_O + 1, \quad 1 \leq j \leq Q_O. \tag{B.3}$$

For $k \in \{1, \dots, Q_O + 1\}$, we define δ_k as $\delta_k = \mathrm{Det}(M^O_{-k})$ where M^O_{-k} is the square matrix corresponding to M^O without the k-th row. Let B^O be the polynomial function defined as:

$$B^O(x) = \sum_{k=0}^{Q_O} (-1)^{k+Q_O} \delta_{k+1} x^k. \tag{B.4}$$

- B^O is of degree Q_O. Indeed, $\delta_{Q_O+1} = \det(M^O_{-(Q_O+1)})$ and $M_{-(Q_O+1)} = R^O D_{\pi^O} R^{O'}$ where $D_{\pi^O} = \mathrm{diag}(\pi^O)$. As a consequence, $M^O_{-(Q_O+1)}$ is the product of invertible matrices then $\delta_{Q_O+1} \neq 0$ and we can conclude.

- Moreover, $\forall l = 1, \ldots, Q_O$, $B^O(\tau_l) = 0$. Indeed, $B^O(\tau_l) = \det(N_l^O)$ where N_l^O is the concatenated matrix $N_l^O = \left(M^O \,|\, V_l\right)$ with $V_l = [1, \tau_l, \ldots, \tau_l^{Q_O}]'$ (computation of the determinant development against the last column). However, from Equation (B.3), we have $M_{\bullet j}^O = \sum_l \tau_l^{j-1} \pi_l^O V_l$, i.e. each column vector of M^O is a linear combination of $V_1, \ldots, V_{Q_O}$. As a consequence, $\forall l = 1, \ldots, Q_O$, N_l^O is of rank $< Q_O + 1$, and so $B^O(\tau_l) = 0$.

The $(\tau_l)_{l=1,\ldots,Q_O}$ being the roots of B, they can be expressed in a unique way (up to label switching) as functions of $(\delta_k)_{k=0,\ldots,Q_O}$, which themselves are derived from $\mathbb{P}_{\mathbf{X},\theta}$. As a consequence, the identifiability of R^O is derived from the identifiability of $(\tau_l)_{l=1,\ldots,Q_O}$. Using the fact that $D_{\pi^O} = R^{O^{-1}} M_{-Q_O}^O R^{O'^{-1}}$, we can identify π^O in a unique way.

Identifiability of α^O. For $1 \leq i, j \leq Q_O$, we define U_{ij} as follows:

$$U_{ij}^O = \mathbb{P}(X_{1,2:(i+1)}^O = 1, X_{2,(n_O-j+2):n_O}^O = 1)$$

with $U_{i1}^O = \mathbb{P}(X_{1,2:(i+1)}^O = 1)$. Then, we can write:

$$U_{i,j}^O = \sum_{l_1,l_2} \tau_{l_1}^{i-1} \pi_{l_1}^O \alpha_{l_1 l_2}^O \pi_{l_2}^O (\tau_{l_2})^{j-1}, \quad \forall 1 \leq i, j \leq Q_O,$$

and as consequence $U^O = R^O D_{\pi^O} \alpha^O D_{\pi^O} R^{O'}$. D_{π^O} and R^O being invertible, we get: $\alpha^O = D_{\pi^O}^{-1} R^{O^{-1}} U^O R^{O'^{-1}} D_{\pi^O}^{-1}$. And so U_O is uniquely derived from $\mathbb{P}_X$, so α^O is identified.

Identifiability of α^I. To identify α^I, we have to take into account the affiliation matrix A. Without loss of generality, we reorder the entries of both levels such that the affiliation matrix A has its $2Q_I \times 2Q_I$ top left block being an identity matrix (Assumption $\mathcal{A}3$).

- For any $k = 1 \ldots, Q_I$ and for $i = 2, \ldots, 2Q_I$, let σ_k be the probability $\mathbb{P}(X_{1i}^I = 1 | Z_1^I = k, A)$, A being such that $A_{jj} = 1, \forall j = 1, \ldots, 2Q_I$.

$$\begin{aligned}\sigma_k &= \mathbb{P}(X_{1i}^I = 1 | Z_1^I = k, A) \\ &= \sum_{k'} \mathbb{P}(X_{1i}^I = 1 | Z_1^I = k, Z_i^I = k') \mathbb{P}(Z_i^I = k' | Z_1^I = k, A)\,.\end{aligned}$$

Moreover,

$$\begin{aligned}\mathbb{P}(Z_i^I = k' | Z_1^I = k, A) &= \sum_l \mathbb{P}(Z_i^I = k' | Z_i^O = l, Z_1^I = k, A) \mathbb{P}(Z_i^O = l | Z_1^I = k, A) \\ &= \sum_l \gamma_{kl} \mathbb{P}(Z_i^O = l | Z_1^I = k, A)\,. \end{aligned} \qquad \text{(B.5)}$$

However, by Bayes' formula

$$\mathbb{P}(Z_i^O = l | Z_1^I = k, A) = \frac{\mathbb{P}(Z_1^I = k | Z_i^O = l, A)\mathbb{P}(Z_i^O = l)}{\mathbb{P}(Z_1^I = k, A)}.$$

Taking into the fact that $i \neq 1$ and A is such that 1 belongs to organisation 1 and i to organisation i, we have: $\mathbb{P}(Z_1^I = k | Z_i^O = l, A) = \mathbb{P}(Z_1^I = k | A)$. And so

$$\mathbb{P}(Z_i^O = l | Z_1^I = k, A) = \mathbb{P}(Z_i^O = l | A) = \pi_l^O.$$

Consequently, from equation (B.5), we have:

$$\mathbb{P}(Z_i^I = k' | Z_1^I = k, A) = \sum_l \gamma_{k'l} \pi_k^O$$

and so:

$$\begin{aligned}
\sigma_k &= \sum_{k'} \mathbb{P}(X_{1i}^I = 1 | Z_1^I = k, Z_i^I = k') \sum_l \gamma_{k'l} \pi_k^O \\
&= \sum_{k'l} \alpha_{kk'}^I \gamma_{k'l} \pi_l^O = (\alpha^I \cdot \gamma \cdot \pi^O)_k \\
&= (\alpha^I \cdot \pi^I)_k, \qquad \text{where } \pi^I = \gamma \cdot \pi^O.
\end{aligned}$$

- Now, we prove that $\forall i = 1, \dots, 2Q_I - 1$,

$$\mathbb{P}(X_{1,2:(i+1)}^I = 1 | Z_1^I = k, A) = \sigma_k^i. \tag{B.6}$$

Indeed,

$$\begin{aligned}
&\mathbb{P}(X_{1,2:(i+1)}^I = 1 | Z_1^I = k, A) \\
&= \sum_{k_{2:(i+1)}} \mathbb{P}(X_{1,2:(i+1)}^I = 1 | Z_{1:(i+1)}^I = (k, k_{2:(i+1)}), Z_1^I = k) \mathbb{P}(Z_{2:(i+1)}^I = k_{2:i+1} | Z_1^I = k, A) \\
&= \sum_{k_{2:(i+1)}} \mathbb{P}(X_{1,2:(i+1)}^I = 1 | Z_{1:(i+1)}^I = (k, k_{2:(i+1)})) \mathbb{P}(Z_{2:(i+1)}^I = k_{2:i+1} | A) \\
&= \sum_{k_{2:(i+1)}} \mathbb{P}(X_{1,2:(i+1)}^I = 1 | Z_{1:(i+1)}^I = (k, k_{2:(i+1)})) \sum_{l_{2:(i+1)}} \mathbb{P}(Z_{2:(i+1)}^I = k_{2:(i+1)}, Z_{2:(i+1)}^O = l_{2:(i+1)}, A).
\end{aligned}$$

Note that, to go from line 2 to line 3, we used the fact that $\mathbb{P}(Z_{2:(i+1)}^I = k_{2:i+1} | Z_1^I = k, A) = \mathbb{P}(Z_{2:(i+1)}^I = k_{2:i+1} | A)$, which is due the the particular structure of A (left diagonal block of size at least $2Q_I$, i.e. for any $i' = 1, \dots, 2Q_I$, individual i' belongs to organisation i'). Moreover, we can write:

$$\begin{aligned}
&\mathbb{P}(Z_{2:(i+1)}^I = k_{2:(i+1)}, Z_{2:(i+1)}^O = l_{2:i+1} | A) \\
&= \left[\prod_{\lambda=2,\dots i+1} \mathbb{P}(Z_\lambda^I = k_\lambda | Z_\lambda^O = l_\lambda) \mathbb{P}(Z_\lambda^O = l_\lambda) \right] \\
&= \left[\prod_{\lambda=2,\dots i+1} \gamma_{k_\lambda l_\lambda} \pi_\lambda^O \right].
\end{aligned}$$

Moreover, by conditional independence of the entries of the matrix X^I given the clustering we have:

$$\mathbb{P}(X^I_{1,2:(i+1)} = 1 | Z^I_1 = k, Z^I_{2:(i+1)} = k_{2:(i+1)}) = \prod_{\lambda=2,\dots i+1} \alpha^I_{kk_\lambda}.$$

As a consequence,

$$\begin{aligned}\mathbb{P}(X^I_{1,2:(i+1)} = 1 | Z^I_1 = k, A) &= \sum_{k_{2:(i+1)}, l_{2:(i+1)}} \prod_{\lambda=2,\dots i+1} \alpha^I_{kk_\lambda} \gamma_{k_\lambda l_\lambda} \pi^O_\lambda \\ &= \prod_{\lambda=2,\dots i+1} \sum_{k_\lambda, l_\lambda} \alpha^I_{kk_\lambda} \gamma_{k_\lambda l_\lambda} \pi^O_\lambda = \sigma^i_k\end{aligned}$$

- Then we define $(u^I_i)_{i=0,\dots,2Q_I-1}$, such that $u^I_0 = 1$ and $\forall 1 \leq i \leq 2Q_I - 1$:

$$\begin{aligned}u^I_i &= \mathbb{P}(X^I_{1,2:(i+1)} = 1 | A) \\ &= \sum_{k,l} \mathbb{P}(X^I_{1,2:(i+1)} = 1 | Z^I_1 = k) \mathbb{P}(Z^I_1 = k | Z^O_1 = l, A) \mathbb{P}(Z^O_1 = l) \\ &= \sum_k \sigma^i_k \underbrace{\sum_l \gamma_{kl} \pi^O_l}_{=\pi^I_k} \\ &= \sum_k \sigma^i_k \pi^I_k.\end{aligned}$$

  Note that the (u^I)'s can be defined because $n_I \geq 2Q_I$ (assumption $\mathcal{A}2$).

- To conclude we use the same arguments as the ones used for the identifiability of α^O, i.e. we define M^I a $(Q_I + 1) \times Q_I$ matrix such that $M^I_{ij} = u^I_{i+j-2}$ together with the matrices M^I_{-k} and the polynomial function B^I (see equation (B.4)). Let R^I be a $Q_I \times Q_I$ matrix such that $R^I_{ik} = \sigma^{i-1}_k$. R^I is an invertible Vandermonde matrix because of assumption $\mathcal{A}1$ on $\alpha^I \cdot \gamma \cdot \pi^O$. As before, R^I can be identified in unique way from B^I. Then, noting that $M^I_{-(Q_I+1)} = R^I D_{\pi^I} R^{I'}$ where $D_{\pi^I} = \text{diag}(\pi^I) = \text{diag}(\gamma \cdot \pi^O)$, we obtain: $D_{\pi^I} = (R^I)^{-1} M^I_{-Q_I} (R^{I'-1})$, which is uniquely defined by $\mathbb{P}_X$. Now, let us introduce

$$U^I_{ij} = \mathbb{P}(X^I_{1,2:(i+1)} = 1, X^I_{2,(n_I-j+2):n_I} = 1)$$

  with $U^I_{i1} = \mathbb{P}(X^I_{1,2:(i+1)} = 1)$. Then we have $U^I = R^I D_{\pi^I} \alpha^I D_{\pi^I} R^{I'}$ and so $\alpha^I = D^{-1}_{\pi^I} (R^I)^{-1} U^I (R^I)'^{-1} D^{-1}_{\pi^I}$. As a consequence, α^I is uniquely identified from $\mathbb{P}_X$.

Identifiability of γ. For any $2 \leq i \leq Q_I$ and $2 \leq j \leq Q_O$, let $U^{IO}_{i,j}$ be the probability that $X^I_{1,2:i} = 1$ and $X^O_{1,(i+1):(i+j-1)} = 1$. Note that the $U^{IO}_{i,j}$ can be defined because $n_O \geq Q_I + Q_O - 1$ and $n_I \geq Q_I$ (assumption $\mathcal{A}2$).

- Then, for all $2 \leq i \leq Q_I$ and $2 \leq j \leq Q_O$,

$$
\begin{aligned}
U_{ij}^{IO} &= \mathbb{P}(X_{1,2:i}^I = 1, X_{1,(i+1):(i+j-1)}^O = 1|A) \\
&= \sum_{k,l} \mathbb{P}(X_{1,2:i}^I = 1, X_{1,(i+1):(i+j-1)}^O = 1|A, Z_1^I = k, Z_1^O = l) \qquad \text{(B.7)} \\
&\quad \times \mathbb{P}(Z_1^I = k, Z_1^O = l, A)\,. \qquad \text{(B.8)}
\end{aligned}
$$

- We first prove that :

$$
\mathbb{P}(X_{1,2:i}^I = 1, X_{1,i+1:i+j-1}^O = 1|A, Z_1^I = k, Z_1^O = l) = \sigma_k^{i-1}\tau_l^{j-1}\,. \qquad \text{(B.9)}
$$

Indeed,

$$
\begin{aligned}
&\mathbb{P}(X_{1,2:i}^I = 1, X_{1,(i+1):(i+j-1)}^O = 1|A, Z_1^I = k, Z_1^O = l) = \\
&= \sum_{k_{2:i}, l_{2:n_O}} \mathbb{P}(X_{1,2:i}^I = 1, X_{1,(i+1):(i+j-1)}^O = 1|Z_{1:i}^I = (k, k_{2:i}), Z^O = (l, l_{2:n_O}), A) \\
&\qquad \times \ \mathbb{P}(Z_{2:i}^I = k_{2:i}, Z_{2:n_O}^O = l_{2:n_O}|Z_1^I = k, Z_1^O = l, A) \\
&= \sum_{k_{2:i}, l_{2:n_O}} \mathbb{P}(X_{1,2:i}^I = 1|Z_{1:i}^I = (k, k_{2:i})) \\
&\qquad \times \ \mathbb{P}(X_{1,(i+1):(i+j-1)}^O = 1|Z_1^O = l, Z_{(i+1):(i+j-1)}^O = l_{(i+1):(i+j-1)}) \\
&\qquad \times \ \mathbb{P}(Z_{2:i}^I = k_{2:i}, Z_{2:n_O}^O = l_{2:n_O}|Z_1^I = k, Z_1^O = l, A)\,. \qquad \text{(B.10)}
\end{aligned}
$$

Moreover, let us have a look at $\mathbb{P}(Z_{2:i}^I = k_{2:i}, Z^O = l_{2:n_O}|Z_1^I = k, Z_1^O = l, A)$:

$$
\begin{aligned}
&\mathbb{P}(Z_{2:i}^I = k_{2:i}, Z_{2:n_O}^O = l_{2:n_O}|Z_1^I = k, Z_1^O = l, A) \\
= \ &\mathbb{P}(Z_{2:i}^I = k_{2:i}|Z_{2:n_O}^O = l_{2:n_O}, Z_1^I = k, Z_1^O = l, A) \times \mathbb{P}(Z_{2:n_O}^O = l_{2:n_O}|Z_1^I = k, Z_1^O = l, A)\,.
\end{aligned}
$$

Because A has a diagonal block of size $\geq Q_I$, we have, for any $i = 1, \ldots, Q_I$, $A_{ij} = 1$ if $j = i$, 0 otherwise, we have

- $\mathbb{P}(Z_{2:i}^I = k_{2:i}|Z_{2:n_O}^O = l_{2:n_O}, Z_1^I = k, Z_1^O = l, A) = \mathbb{P}(Z_{2:i}^I = k_{2:i}|Z_{2:i}^O = l_{2:i})$,
- $\mathbb{P}(Z_{2:n_O}^O = l_{2:n_O}|Z_1^I = k, Z_1^O = l, A) = \mathbb{P}(Z_{2:n_O}^O = l_{2:n_O})\,.$

As a consequence,

$$
\begin{aligned}
&\mathbb{P}(Z_{2:i}^I = k_{2:i}, Z_{2:n_O}^O = l_{2:n_O}|Z_1^I = k, Z_1^O = l, A) = \\
&\quad \mathbb{P}(Z_{2:i}^I = k_{2:i}|Z_{2:i}^O = l_{2:i})\mathbb{P}(Z_{2:i}^O = l_{2:i})\mathbb{P}(Z_{(i+1):(i+j-1)}^O = l_{(i+1):(i+j-1)}) \\
&\times \mathbb{P}(Z_{(i+j):n_O}^O = l_{(i+j):n_O})\,.
\end{aligned}
$$

Going back to equation (B.10) and decomposing the summation we obtain:

$$
\begin{aligned}
& \mathbb{P}(X^I_{1,2:i} = X^O_{1,(i+1):(i+j-1)} = 1 | A, Z^I_1 = k, Z^O_1 = l) \\
= & \sum_{k_{2:i}, l_{2:n_O}} \mathbb{P}(X^I_{1,2:i} = 1 | Z^I_{1:i} = (k, k_{2:i})) \\
& \quad \times \mathbb{P}(X^O_{1,(i+1):(i+j-1)} = 1 | Z^O_1 = l, Z^O_{(i+1):(i+j-1)} = l_{(i+1):(i+j-1)}) \\
& \quad \times \mathbb{P}(Z^I_{2:i} = k_{2:i} | Z^O_{2:i} = l_{2:i}) \mathbb{P}(Z^O_{2:i} = l_{2:i}) \mathbb{P}(Z^O_{(i+1):(i+j-1)} = l_{(i+1):(i+j-1)}) \\
& \quad \times \mathbb{P}(Z^O_{(i+j):n_O} = l_{(i+j):n_O}) \\
= & \sum_{k_{2:i}} \mathbb{P}(X^I_{1,2:i} = 1 | Z^I_{1:i} = (k, k_{2:i})) \sum_{l_{2:i}} \mathbb{P}(Z^I_{2:i} = k_{2:i} | Z^O_{2:i} = l_{2:i}) \mathbb{P}(Z^O_{2:i} = l_{2:i}) \\
& \quad \sum_{l_{(i+1):(i+j-1)}} \mathbb{P}(X^O_{1,(i+1):(i+j-1)} = 1 | Z^O_1 = l, Z^O_{(i+1):(i+j-1)} = l_{(i+1):(i+j-1)})) \\
& \quad \times \underbrace{\mathbb{P}(Z^O_{(i+1):(i+j-1)} = l_{(i+1):(i+j-1)})}_{=\mathbb{P}(Z^O_{(i+1):(i+j-1)} = l_{(i+1):(i+j-1)} | Z^O_1 = l)} \underbrace{\sum_{l_{(i+j):n_O}} \mathbb{P}(Z^O_{(i+j):n_O} = l_{(i+j):n_O})}_{=1} \\
= & \sum_{k_{2:i}} \mathbb{P}(X^I_{1,2:i} = 1 | Z^I_1 = k, Z^I_{2:i} = k_{2:i}) \mathbb{P}(Z^I_{2:i} = k_{2:i} | A) \times \mathbb{P}(X^O_{1,(i+1):(i+j-1)} = 1 | Z^O_1 = l) \\
= & \sum_{k_{2:i}} \mathbb{P}(X^I_{1,2:i} = 1 | Z^I_1 = k, Z^I_{2:i} = k_{2:i}) \mathbb{P}(Z^I_{2:i} = k_{2:i} | Z^I_1 = k, A) \\
& \times \mathbb{P}(X^O_{1,(i+1):(i+j-1)} = 1 | Z^O_1 = l) \\
= & \mathbb{P}(X^I_{1,2:i} = 1 | Z^I_1 = k, A) \mathbb{P}(X^O_{1,(i+1):(i+j-1)} = 1 | Z^O_1 = l) \, .
\end{aligned}
$$

Finally, we have :

$$
\begin{aligned}
\mathbb{P}(X^I_{1,2:i} = 1 | Z^I_1 = k, A) &= \sigma_k^{i-1}, \quad \text{from equation (B.6)} \\
\mathbb{P}(X^O_{1,(i+1):(i+j-1)} = 1 | Z^O_1 = l) &= \tau_l^{j-1},
\end{aligned}
$$

and so, we have proved equality (B.9).

- Now, $A_{11} = 1$ implies $\mathbb{P}(Z^I_1 = k, Z^O_1 = l | A) = \gamma_{kl} \pi^O_l$ and combining this result with equations (B.9) and (B.7) leads to: $U^{IO}_{ij} = \sum_{k,l} \sigma_k^{i-1} \gamma_{kl} \pi^O_l \tau_l^{j-1}$. Setting

$$
\begin{aligned}
U^{IO}_{1j} &= \mathbb{P}(X^O_{1,i+1} = 1, \ldots, X^O_{1,i+j-1} = 1 | A) = \sum_{k,l} \gamma_{kl} \pi^O_l \tau_l^{j-1}, \quad \text{for } j > 1 \\
U^{IO}_{i1} &= \mathbb{P}(X^I_{12} = \cdots = X^I_{1,i} = 1 | A) = \sum_{k,l} \gamma_{kl} \pi^O_l, \quad \text{for } i > 1 \\
U^{IO}_{11} &= 1
\end{aligned}
$$

we obtain the following matrix expression for U^{IO}: $U^{IO} = R^I \gamma D_{\pi^O} R^{O'}$ where U^{IO} is completely defined by $\mathbb{P}_{X,\theta}$ and the other terms have been identified before. Thus $\gamma = (R^I)^{-1} U^{IO} (R^{O'})^{-1} D_{\pi^O}^{-1}$ and γ is identified.

C. Details of the Variational EM

The variational bound for the stochastic block model for multilevel network can be written as follows:

$$\begin{aligned}
\mathcal{I}_\theta(\mathcal{R}(Z^I, Z^O|A)) \quad = \quad & \sum_{j,l} \tau^O_{jl} \log \pi^O_l + \sum_{i,k} \tau^I_{ik} \sum_{j,l} A_{ij} \tau^O_{jl} \log \gamma_{kl} \\
& + \frac{1}{2} \sum_{i' \neq i} \sum_{k,k'} \tau^I_{ik} \tau^I_{i'k'} \log \phi\left(X^I_{ii'}, \alpha^I_{kk'}\right) + \frac{1}{2} \sum_{j' \neq j} \sum_{l,l'} \tau^O_{jl} \tau^O_{j'l'} \log \phi\left(X^O_{jj'} \alpha^O_{ll'}\right) \\
& - \sum_{i,k} \tau^I_{ik} \log \tau^I_{ik} - \sum_{j,l} \tau^O_{jl} \log \tau^O_{jl}
\end{aligned}$$

The variational EM algorithm then consists on iterating the two following steps. At iteration $(t+1)$:

VE step compute

$$\begin{aligned}
\{\tau^I, \tau^O\}^{(t+1)} &= \arg\max_{\tau^I, \tau^O} \mathcal{I}_{\theta^{(t)}}(\mathcal{R}(Z^I, Z^O|A)) \\
&= \arg\min_{\tau^I, \tau^O} KL\left(\mathcal{R}(Z^I, Z^O|A) \| \mathbb{P}_{\theta^{(t)}}(Z^I, Z^O|X^I, X^O, A)\right) .
\end{aligned}$$

M step compute

$$\theta^{(t+1)} = \arg\max_\theta \mathcal{I}_\theta(\mathcal{R}^{(t+1)}(Z^I, Z^O|A)).$$

The variational parameters are sought by solving the equation:

$$\Delta_{\tau^I, \tau^O}\left(\mathcal{I}_\theta(\mathcal{R}(Z^I, Z^O|A) + L(\tau^I, \tau^O)\right) = 0,$$

where $L(\tau^I, \tau^O)$ are the Lagrange multipliers for τ^I_i, τ^O_j for all $i \in \{1, \ldots, n_I\}$, $j \in \{1, \ldots, n_I\}$. There is no closed-form formula but when computing the derivatives, we obtain that the variational parameters follow the fixed point relationships:

$$\begin{aligned}
\widehat{\tau^O_{jl}} &\propto \pi^O_l \prod_{i,k} \gamma_{kl}^{A_{il}\widehat{\tau^I_{ik}}} \prod_{j' \neq j} \prod_{l'} \phi(X^O_{jj'}, \alpha^O_{ll'})^{\widehat{\tau^O_{j'l'}}} \\
\widehat{\tau^I_{jl}} &\propto \prod_{j,l} \gamma_{kl}^{A_{il}\widehat{\tau^O_{jl}}} \prod_{i' \neq i} \prod_{k'} \phi(X^I_{ii'}, \alpha^I_{kk'})^{\widehat{\tau^I_{i'k'}}} ,
\end{aligned}$$

which are used in the VE step to update the τ^I_i's and τ^O_j's.

On each update, the variational parameters of a certain level depend on both the parameter γ and the variational parameters of the other level, which emphasises the dependency structure of this multilevel model and the role of γ as the dependency parameter of the model. Notice also that when $\gamma_{kl} = \gamma_{kl'} = \pi^I_k$ for all l, l', that is the case of independence between the two levels then we can rewrite the fixed point relationships as follows:

$$\widehat{\tau^O_{jl}} \propto \pi^O_l \widehat{\tau^I_{ik}} \prod_{j' \neq j} \prod_{l'} \phi(X^O_{jj'}, \alpha^O_{ll'})^{\widehat{\tau^O_{j'l'}}} \quad \text{and} \quad \widehat{\tau^I_{jl}} \propto \quad \pi^I_k \widehat{\tau^O_{jl}} \prod_{i' \neq i} \prod_{k'} \phi(X^I_{ii'}, \alpha^I_{kk'})^{\widehat{\tau^I_{i'k'}}},$$

which is exactly the expression of the fixed point relationship of two independent SBMs. Then, for the M step, we derive the following closed-form formulae:

$$\widehat{\pi^O_l} = \frac{1}{n_O}\sum_j \widehat{\tau^O_{jl}} \qquad \widehat{\alpha^O_{ll'}} = \frac{\sum_{j'\neq j} \widehat{\tau^O_{jl}} X^O_{jj'} \widehat{\tau^O_{j'l'}}}{\sum_{j'\neq j} \widehat{\tau^I_{jl}} \widehat{\tau^I_{j'l'}}}$$

$$\widehat{\gamma_{kl}} = \frac{\sum_{i,j} \widehat{\tau^I_{ik}} A_{ij} \widehat{\tau^O_{jl}}}{\sum_{i,j} A_{ij} \widehat{\tau^O_{jl}}} \qquad \widehat{\alpha^I_{kk'}} = \frac{\sum_{i'\neq i} \widehat{\tau^I_{ik}} X^I_{ii'} \widehat{\tau^I_{i'k'}}}{\sum_{i'\neq i} \widehat{\tau^I_{ik}} \widehat{\tau^I_{i'k'}}}$$

for which the gradient

$$\Delta_\theta \left(\mathcal{I}_\theta(\mathcal{R}(Z^I, Z^O|A)) + L(\pi^O, \gamma)\right),$$

is null. The term $L(\pi^O, \gamma)$ contains the Lagrange multipliers for π^O and $\gamma_{k\cdot}$ for all $k \in \{1, \ldots, Q_I\}$.

Model parameters have natural interpretations. π^O_l is the mean of the posterior probabilities for the organisations to belong to cluster l. $\alpha^I_{kk'}$ (resp. $\alpha^O_{ll'}$) is ratio of existing links over possible links between blocks k and k' (resp. l and l'). γ_{kl} is the ratio of the number of individuals in cluster k that are affiliated to any organisations of cluster l on the number of individuals that are affiliated to any organisations of cluster l. If γ is such that the levels are independent, then any column of γ represents the proportion of individuals in the different blocks:

$$\pi^I_k = \gamma_{k1} = \frac{1}{n_I}\sum_i \widehat{\tau^I_{ik}}.$$

D. Details of the ICL criterion

We now derive an expression for the Integrated Complete Likelihood (ICL) model selection criterion. Following Daudin et al. (2008), the ICL is based on the integrated complete likelihood i.e. the likelihood of the observations and the latent variables where the parameters have been integrating out against a prior distribution. The latent variables (Z^I, Z^O) being unobserved, they are imputed using the maximum a posteriori (MAP) or $\hat{\tau}$. We denote by $\widehat{Z^O}$ and $\widehat{Z^I}$ the inputed latent variables. After imputation of the latent variables, an asymptotic approximation of this quantity leads to the ICL criterion given in the paper (Equation (8)) and recalled here:

$$\begin{aligned} ICL(Q_I, Q_O) = {} & \log \ell_{\widehat{\theta}}(X^I, X^O, \widehat{Z^I}, \widehat{Z^O}|A, Q_I, Q_O) \\ & - \frac{1}{2}\frac{Q_I(Q_I+1)}{2}\log\frac{n_I(n_I-1)}{2} - \frac{Q_O(Q_I-1)}{2}\log n_I \\ & - \frac{1}{2}\frac{Q_O(Q_O+1)}{2}\log\frac{n_O(n_O-1)}{2} - \frac{Q_O-1}{2}\log n_O. \end{aligned}$$

Let $\Theta = \Pi^O \times \mathcal{A}^I \times \mathcal{A}^O \times \Gamma$ be the space of the model parameters. We set a prior distribution on θ:

$$p(\theta|Q_I, Q_O) = p(\gamma|Q_I, Q_O)p(\pi^O|Q_O)p(\alpha^I|Q_I)p(\alpha^O|Q_O)$$

where $p(\pi^O|Q_O)$ is a Dirichlet distribution of hyper-parameter $(1/2, \cdots, 1/2)$ and $p(\alpha^I|Q_I)$ and $p(\alpha^O|Q_O)$ are independent Beta distributions.
The marginal complete likelihood is written as follows:

$$\begin{aligned}
\log \ell_\theta(\mathbf{X}, \mathbf{Z}|A, Q_I, Q_O) &= \log\left(\int_\Theta \ell_\theta(X^I, X^O, Z^I, Z^O|\theta, A, Q_I, Q_O) p(\theta|Q_I, Q_O) \mathrm{d}\theta\right) \\
&= \log \ell_{\alpha^I}(X^I|Z^I, Q_I) \qquad \text{(D.11)} \\
&\quad + \log \ell_\gamma(Z^I|A, Z^O, Q_I, Q_O) \qquad \text{(D.12)} \\
&\quad + \log \ell_{\alpha^O, \pi^O}(X^O, Z^O|Q_O)\,. \qquad \text{(D.13)}
\end{aligned}$$

The quantity defined in (D.13) evaluated at $Z^O := \widehat{Z^O}$ is approximated as in Daudin et al. (2008) by

$$\begin{aligned}
\log \ell_{\alpha^O}(X^O, \widehat{Z^O}, Q_O) &\approx_{n_O \to \infty} \log \ell_{\widehat{\alpha^O}, \widehat{\pi^O}}(X^O, \widehat{Z^O}|Q_O) - \mathrm{pen}(\pi^O, \alpha^O, Q_O) \\
\mathrm{pen}(\pi^O, \alpha^O, Q_O) &= \tfrac{Q_O - 1}{2} \log n_O + \tfrac{1}{2} \tfrac{Q_I(Q_I+1)}{2} \log \tfrac{n_I(n_I-1)}{2}
\end{aligned} \;. \qquad \text{(D.14)}$$

This approximation results from a BIC-type approximation of $\log \ell_{\widehat{\alpha^O}}(X^O|\widehat{Z^O}, Q_O)$ and a Stirling approximation of $\log \ell_{\pi^O}(\widehat{Z^O}, Q_O)$.
The same BIC-type approximation on $\log \ell_{\alpha^I}(X^I|\widehat{Z^I}, Q_I)$ (Equation (D.11)) leads to:

$$\begin{aligned}
\log \ell_{\alpha^I}(X^I|\widehat{Z^I}, Q_I) &=_{n_I \to \infty} \log \ell_{\widehat{\alpha^I}}(X^I|\widehat{Z^I}, Q_I) + \mathrm{pen}(\alpha^I, Q_I) \\
\text{with } \mathrm{pen}(\alpha^I, Q_I) &= \tfrac{1}{2} \tfrac{Q_I(Q_I+1)}{2} \log \tfrac{n_I(n_I-1)}{2}
\end{aligned} \;. \qquad \text{(D.15)}$$

For quantity (D.12) depending on γ and Z^I given (Q_I, Q_O), we have to adapt the calculus. Let us set independent Dirichlet prior distributions of order Q_I $\mathcal{D}(1/2, \ldots, 1/2)$ on the columns $\gamma_{\cdot l}$. We are able to derive an exact expression of $\log \ell_\gamma(Z^I|A, Z^O, Q_I, Q_O)$:

$$\begin{aligned}
\ell_\gamma(Z^I|A, Z^O, Q_I, Q_O) &= \int \ell(Z^I|A, Z^O, \gamma, Q_I, Q_O) p(\gamma, Q_I, Q_O) \mathrm{d}\gamma \\
&= \prod_i \int \prod_{j,k,l} \gamma_{kl}^{A_{ij} Z^I_{ik} Z^O_{jl}} p(\gamma_{kl}) \mathrm{d}\gamma_{kl} \\
&= \prod_l \int \prod_k \gamma_{kl}^{N_{kl}} p(\gamma_{kl}) \mathrm{d}\gamma_{kl}, \quad \text{where} \quad N_{kl} = \sum_{ij} A_{ij} Z^I_{ik} Z^O_{jl} \\
&= \prod_l \int \prod_k \gamma_{k,l}^{N_{kl}+a-1} \frac{\Gamma(1/2 \cdot Q_I)}{\Gamma(1/2)^{Q_I}} \mathrm{d}\gamma_{kl} \\
&= \frac{\Gamma(1/2 Q_I)^{Q_O}}{\Gamma(1/2)^{Q_O + Q_I}} \prod_l \frac{\prod_k \Gamma(N_{kl} + 1/2)}{\Gamma(1/2 Q_I + \sum_k N_{kl})}\,.
\end{aligned}$$

Now, using the fact that $\log \Gamma(n+1) \overset{n \to \infty}{\sim} (n + 1/2) \log n + n$, we obtain:

$$\begin{aligned}
\log \ell_\gamma(Z^I|A, Z^O, Q_I, Q_O) \approx_{(n_O, n_I) \to \infty} & \textstyle\sum_{k,l} (N_{kl} \log N_{kl} + N_{kl}) \\
& - \textstyle\sum_l \left(\frac{Q_I - 1}{2} + \sum_k N_{kl}\right) \log\left(\sum_k N_{kl}\right) - \sum_{k,l} N_{kl}.
\end{aligned} \qquad \text{(D.16)}$$

The quantity (D.16) evaluated at $(Z^I, Z^O) := (\widehat{Z^I}, \widehat{Z^O})$ can be reformulated in the following way:

$$\begin{aligned} \log \ell_\gamma(\widehat{Z^I}|A, \widehat{Z^O}, Q_I, Q_O) &\approx_{(n_O,n_I)\to\infty} \log \ell_{\hat\gamma}(\widehat{Z^I}|A, \widehat{Z^O}, Q_I, Q_O) - \tfrac{Q_I-1}{2} \textstyle\sum_l \log \sum_{i,j} A_{ij}\widehat{Z^O_{jl}} \\ \text{with } \hat\gamma_{kl} &= \frac{\sum_{i,j} \widehat{Z^I_{ik}} A_{ij} \widehat{Z^O_{jl}}}{\sum_{i,j} A_{ij}\widehat{Z^O_{jl}}} \end{aligned}$$

Noticing that $\log \sum_{i,j} A_{ij}\widehat{Z^O_{jl}} = \log n_I + \log \frac{\sum_{i,j} A_{ij}\widehat{Z^O_{jl}}}{n_I} = O(\log n_I)$ leads to

$$\log \ell_\gamma(\widehat{Z^I}|A, \widehat{Z^O}, Q_I, Q_O) \approx_{(n_O,n_I)\to\infty} \log \ell_{\hat\gamma}(\widehat{Z^I}|A, \widehat{Z^O}, Q_I, Q_O) - \frac{Q_I-1}{2} Q_O \log n_I\,. \tag{D.17}$$

Combining Equations (D.14), (D.15) and (D.17) we obtain the given expression.

References

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

Daudin, J.-J., F. Picard, and S. Robin (2008). A mixture model for random graphs. *Statistics and computing 18*(2), 173–183.