A Stochastic Block Model for Multilevel Networks: Application to the Sociology of Organisations

Saint-Clair Chabert-Liddell* Pierre Barbillon* Sophie Donnet*
Emmanuel Lazega†

Abstract

This work is motivated by the analysis of multilevel networks. We define a multilevel network as the junction of two interaction networks, one level representing the interactions between individuals and the other one the interactions between organisations. The levels are linked by an affiliation relationship, each individual belonging to a unique organisation. We design a Stochastic block model (SBM) suited to multilevel networks. SBM is a latent variable model for networks, where the connections between nodes depend on a latent clustering (blocks), thus modeling some connection heterogeneity. We prove the identifiability of our model. The parameters of the model are estimated with a variational EM algorithm. An Integrated Completed Likelihood criterion is developed not only to select the number of blocks but also to detect whether the two levels (individuals and organisations) are dependent or not. In a comprehensive simulation study, we exhibit the benefit of considering our approach, illustrate the robustness of our parameter estimation and highlight the reliability of our model selection criterion. Our approach is applied on a sociological dataset collected during a television programmers trade fare. The inter-organisational level is the economic networks between companies and the inter-individual level is the informal network between their representatives.

Keywords: Stochastic Block Model, Multilevel Network, Variational methods, Social Network, Sociology of Organisations, Hierarchical model

1 Introduction

The statistical analysis of network data has been a hot topic for the last decade. The last few years witnessed a growing interest for multilayer networks (see Kivelä et al., 2014; Bianconi 2018). A particular case of multilayer networks are multilevel networks.

Multilevel networks arise in sociology when willing to study jointly the social network of individuals and the interaction network of organisations the individuals belong to. Indeed, individuals not only interact with each others but are also members of interacting organisations. Following Lazega and Snijders (2015), one might think that these two

*UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 75005, Paris, France
†Institut d’Études Politiques de Paris, France
types of interactions (between individuals and between organisations) are interdependent, the individuals shaping there organisations and the organisations having an influence on the individuals.

In what follows, a multilevel network refers to an inter-individual network, an inter-organisational network and the affiliation on the individuals to the organisations. Besides, we assume that the individuals belong to a unique organisation (this point will be discussed at the end of the paper). Such a dataset is studied by Lazega et al. (2008), some researchers in cancerology being the individuals and their laboratories the organisations. Another dataset is dealt with by Brailly et al. (2016). It is concerned with the economic network of audiovisual firms and the informal network of their sales representatives. This latter dataset will be analysed in this contribution. From a general perspective in the sociology of organisations and collective action, analysing multilevel aims to understand how the two levels are interwined and how one level impacts the other. In the case of Brailly et al. (2016), it could unravel the tense competition and the cooperation dilemmas among actors.

In the last years, SBM (Holland et al., 1983) has become a popular tool to model the heterogeneity of connection in a network, assuming that the actors at stake are divided into blocks and that the members of a cluster (block) share a similar profile of connectivity. Compared to other graph clustering methods such as modularity maximisation, hierarchical clustering or spectral clustering (see Kolaczyk, 2009, and references therein), SBM is a generative model and can fit to a wide range of topologies since it clusters into blocks the nodes that are structurally equivalent. This includes but is not restricted to the detection of associative communities where the probability of connection within a block is higher than the probability of connection between blocks. SBMs have been extended to particular types of multilayer networks: Barbillon et al. (2017) propose a SBM for multiplex networks and Matias and Miele (2017) a SBM for time-evolving networks. In this paper, we propose a SBM suited to multilevel networks.

**Our contribution.** In a few words, we model the heterogeneity in the inter-individual and inter-organisational connections by introducing blocks of individuals and blocks of organisations, the blocks containing homogeneous groups of actors (individuals or organisations) with respect to their connectivities. The two levels are assumed to be interdependent through their latent blocks. More specifically, the latent blocks of the individual level depend on the latent blocks of the organisation level and the affiliation. Because of the dependence on the affiliation, our model states that two individuals belonging to organisations in the same block (of organisations) are more likely to be in the same block of individuals and so to share a similar pattern of connections.

Due to the latent variables, the estimation of the parameters is a complex task. We resort to a variational version of the Expectation-Maximisation (EM) algorithm. For SBM, the variational approach (Jordan et al., 1999; Blei et al., 2017) has proven its efficiency for deriving maximum likelihood estimates (Daudin et al., 2008; Mariadassou et al., 2010; Barbillon et al., 2017) and for Bayesian inference (Latouche et al., 2012).
In this paper, we obtain approximate maximum likelihood estimates by an ad-hoc version of the variational EM algorithm.

Another important task is the choice of the number of blocks (blocks). Several solutions adapted to SBM can be found in the literature. Daudin et al. (2008) uses the Integrated Complete Likelihood (ICL) which is an alternative to the Bayesian Information Criterion (BIC). ICL was firstly developed by Biernacki et al. (2000) for mixture models. ICL has proven its efficiency and relevance for various SBMs and their extensions such as multiplex network (Barbillon et al., 2017), dynamic SBM (Matias and Miele, 2017) or degree corrected SBM (Yan, 2016). We propose an ad-hoc version of the ICL criterion. Besides, a critical issue in sociology is to verify the multilevel hypothesis of a given dataset. We propose a criteria deciding whether the two levels (inter-individual and inter-organisational) are independent or not.

Related works. The term multilevel networks arises in the statistical literature for a wide variety of complex networks. For instance, Zijlstra et al. (2006) adapt the p2-model to handle multiple observations of a network. Sweet et al. (2014) extend the Mixed Membership Stochastic Block Model (Airoldi et al., 2008) to the hierarchical network models framework (Sweet et al., 2013) for the same type of data. Snijders (2017) discusses the use of the stochastic actor-oriented model (Snijders, 2001) for temporal and multivariate networks.

When dealing with the multilevel networks we defined before, Wang et al. (2013) adopt an exponential random graph model (ERGM) strategy that is used in applications across many fields such as environmental science (Hileman and Lubell, 2018) or sociology (Lazega and Snijders, 2015, chapter 10-11, 13-14). When focusing on a clustering approach, as far as we know, only two papers have been published, namely Žiberna (2014) and Barbillon et al. (2017). The first paper develops three general approaches for blockmodeling multilevel networks. First, the separate analysis consists in clustering the levels separately or using the clustering of one level on the other. Second, the conversion approach converts the level of the organisations into a new kind of interaction between individuals, the interactions are then aggregated into a single layer network; this is close to the approach taken by Barbillon et al. (2017) who transform the inter-organisational network into an inter-individual network thus adopting a multiplex network approach (the individuals interconnect directly or through the organisations they belong to). The third approach in Žiberna (2014) is called the true multilevel approach and is the closest to the one we propose on this paper. However, the cost function to be optimised requires a pre-specified blockmodeling and so is not a generative model. Moreover, it is not as flexible as the SBM since the type of topology has to be specified.

Also, note that the multiplex SBM approach applied to a multilevel network suggested by Barbillon et al. (2017) is only applicable when the number of individuals and organisations are not too different. Indeed it requires to duplicate the data of the inter-organisational level to fit the size of the inter-individual’s one. Furthermore it only provides a clustering on the individuals and not on both the individuals and the organisations. In contrast, our MLVSBM does not need to modify the data to obtain a
bi-clustering of the nodes.

If we release the constraint of the unique affiliation, then the inter-level can be modelled by a latent block model and we obtain a particular case of the multipartite SBM of Bar-Hen et al. (2018). However, then the interactions between individuals and organisations are considered on the same level as the affiliations, and the clustering might be strongly influenced by the number of individuals in each organisation.

Finally, our work is also different from the SBM with edges covariates (Mariadassou et al., 2010) with the individuals as nodes and the inter-organisational network as edges covariates. Indeed, in that case, the clustering obtained for the individuals is the remaining structure of the inter-individual level once the effect of the covariates has been taken into account. In addition this model does not provide a clustering of the organisations.

Outline of the paper. The paper is organised as follows. The Stochastic block model adapted to multilevel networks (MLVSBM) is defined in Section 2. We also give conditions guaranteeing the independence between levels and the identifiability of the parameters. The inference strategy and the model selection criterion are provided in Section 3. The proof of the independence between levels, of the identifiability and the details on variational EM and ICL criterion are postponed to the appendix. We present in Section 4 an extensive simulation study illustrating the relevance of our inference method, model selection criterion and procedure. Section 5 is dedicated to the application of our MLVSBM on the sociological dataset. Finally we discuss our contribution and future works in Section 6.

2 A multilevel stochastic block model (MLVSBM)

Dataset. Let us consider \( n_I \) individuals involved in \( n_O \) organisations. We encode the networks into adjacency matrices as follows. Let \( X^I \) be the binary \( n_I \times n_I \) matrix representing the inter-individual network. \( X^I \) is such that:

\[
X^I_{ii'} = \begin{cases} 
1 & \text{if there is an interaction from individual } i \text{ to individual } i' \\
0 & \text{otherwise.} 
\end{cases}
\]

\( X^O \) is the binary \( n_O \times n_O \) matrix representing the inter-organisational network, \( \forall (j, j') \in \{1, \ldots, n_O\} \) (\( j \neq j' \)):

\[
X^O_{jj'} = \begin{cases} 
1 & \text{if there is an interaction from organisation } j \text{ to organisation } j' \\
0 & \text{otherwise.} 
\end{cases}
\]

Remark. [1] and [2] correspond to directed interactions. For undirected interactions one would set:

\[ X^I_{ii'} = X^I_{i'i} \ \forall (i, i') \in \{1, \ldots, n_I\}^2 \text{ or/and } X^O_{jj'} = X^O_{j'j} \ \forall (j, j') \in \{1, \ldots, n_O\}^2. \]

In what follows, we present the results for undirected networks. However, all the results can be adapted to directed networks. Also note that no self-interaction is considered.
Let $A$ be the affiliation matrix. $A$ is a $n_I \times n_O$ matrix such that:

$$A_{ij} = \begin{cases} 
1 & \text{if individual } i \text{ belongs to organisation } j, \\
0 & \text{otherwise}
\end{cases}$$

$A$ is such that $\forall i = 1, \ldots, n_I, \sum_{j=1}^{n_O} A_{ij} = 1$ since we assume that any individual belongs to a unique organisation. A synthetic view of a generic dataset is provided in Figure 1.

We propose a joint modelling of the inter-individual and inter-organisational networks based on an extension of the SBM. More precisely, assume that the $n_O$ organisations are divided into $Q_O$ blocks and that the individuals are divided into $Q_I$ blocks. Let $Z^O = (Z^O_1, \ldots, Z^O_{n_O})$ and $Z^I = (Z^I_1, \ldots, Z^I_{n_I})$ be such that $Z^O_j = l$ if organisation $j$ belongs to cluster $l$ ($l \in \{1, \ldots, Q_O\}$) and $Z^I_i = k$ if individual $i$ belongs to cluster $k$ ($k \in \{1, \ldots, Q_I\}$).

Given these clusterings, we assume that the interactions between organisations and interactions between individuals are independent and distributed as follows:

$$\begin{align*}
\mathbb{P}(X^O_{ij'} = 1 | Z^O_j, Z^O_{i'}) &= a^O_{Z^O_j Z^O_{i'}} \\
\mathbb{P}(X^I_{ii'} = 1 | Z^I_i, Z^I_{i'}) &= a^I_{Z^I_i Z^I_{i'}}.
\end{align*}$$

(3)

As a consequence, the blocks gather nodes (blocks of individuals on the one hand and blocks of organisations on the other hand) sharing the same profiles of connectivity.

| Individual 1 | \cdots | Individual $n_I$ | Organisation 1 | \cdots | Organisation $n_O$ |
|---------------|--------|------------------|----------------|--------|------------------|
| $A_{ij}$      |        |                  | $X^I_{ii'}$    |        |                  |
| 1             |        |                  | 1              |        |                  |
| 0 – 1 – 0     |        |                  | $A_{ij}$       |        |                  |
| 1             |        |                  | 1              |        |                  |
|                |        |                  | 1              |        |                  |

Figure 1: Matrix representation of a multilevel network
In order to take into account the fact that organisations may shape the individual behaviours, we assume that the memberships of the individuals ($Z^I$) depend on the cluster of the organisations ($Z^O$) they are affiliated to. More precisely, we set:

$$P(Z^I_i = k | Z^O_j, A_{ij} = 1) = \gamma_{kZ^O_j} \quad \forall i \in \{1, \ldots, n_I\} \quad \forall k \in \{1, \ldots, Q_I\}$$  \hspace{1cm} (4)

where $\gamma$ is a $Q_I \times Q_O$ matrix such that $\sum_{k=1}^{Q_I} \gamma_{kl} = 1$ for any $l \in \{1, \ldots, Q_O\}$. The ($Z^O_j$) are assumed to be independent random variables distributed as

$$P(Z^O_j = l) = \pi^O_l \quad \forall j \in \{1, \ldots, n_O\} \quad \forall l \in \{1, \ldots, Q_O\}$$  \hspace{1cm} (5)

Equations (4) and (5) state that the clustering of an individual is not completely driven by his/her behaviour but is also shaped by the clustering of the organisation he/she belongs to. In particular, if $Q_O = Q_I$ and $\gamma$ is equal to the identity matrix (up to a reordering of the rows) then, the clustering of the individuals is completely determined by the clustering of the organisations. At the opposite, if all the columns of $\gamma$ are equal, then the clustering of the individuals is independent on the clustering of the organisations. This point will be developed hereafter.

Equations (3), (4) and (5) define a joint modelling of $X^I$ and $X^O$. In what follows, we set $\theta = \{\pi^O, \gamma, \alpha^I, \alpha^O\}$ the vector of the unknown parameters, $X = \{X^I, X^O\}$ are the observed variables and $Z = \{Z^I, Z^O\}$ the latent variables. The DAG of the MLVSBM is plotted in Figure 2.

Figure 2: DAG of the stochastic block model for multilevel network (MLVSBM)
Likelihood. From Equations \([3], [4], \text{and} [5]\), we derive the complete log-likelihood for a directed MLVSBM:

\[
\log \ell_\theta(X^I, X^O, Z^I, Z^O | A) = \log \ell_\pi (Z^O) + \log \ell_\gamma (Z^I | Z^O, A) + \log \ell_\alpha^I (X^I | Z^I) + \log \ell_\alpha^O (X^O | Z^O) = \sum_{j,l} \mathbb{1}_{Z^O_j = l} \log \pi^O_l + \sum_{i,k} \mathbb{1}_{Z^I_i = k} \sum_{j,l} A_{ij} \mathbb{1}_{Z^O_j = l} \log \gamma_{kl}
\]

\[+ \frac{1}{2} \sum_{i' \neq i} \sum_{k,k'} \mathbb{1}_{Z^I_i = k} \mathbb{1}_{Z^I_{i'} = k'} \log \phi(X^I_{ii'}, \alpha^I_{kk'}) + \frac{1}{2} \sum_{j' \neq j} \sum_{l,l'} \mathbb{1}_{Z^O_j = l} \mathbb{1}_{Z^O_{j'} = l'} \log \phi(X^O_{jj'}, \alpha^O_{ll'}), \]

where \(\phi(x, a) = a^x (1 - a)^{1-x}\).

Remark. Note that the factors \(1/2\) in Equation \((6)\) derive from the fact that we consider undirected networks. If one or both of the networks are directed, then the corresponding \(1/2\) disappears.

The log-likelihood of the observations \(\ell_\theta(X|A)\) is obtained by integrating out the latent variables \(Z\) in Equation \((6)\). As soon as \(n_O, n_I, Q_O, \text{or} Q_I\) increase, this summation over all the possible clusterings \(Z^I\) and \(Z^O\) cannot be performed within a reasonable computational time. As a consequence, we will resort to the EM algorithm to maximise this likelihood (see Section \(3\)).

Independence. We now derive conditions for the structural independence between levels in term of parameters equality.

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

The proof of this proposition is left in the appendix, Section \(A\). The above proposition can be interpreted as follows: in the case where the clustering of the individuals does not depend on the clustering of the organisations, all column vectors of \(\gamma\) are identical. Hence, under this restriction on \(\gamma\), the model for multilevel network can be rewritten as the product of two independent Stochastic Block Model, one for each level. Conversely, in the case of a strong dependence between the levels, each column of \(\gamma\) will have one coefficient, the value of which close to one. Therefore, individuals belonging to the same block will be affiliated to the same block of organisations. Even if the \(\gamma\)'s imply a dependent relationship between the two levels, the connections of the corresponding blocks at the two levels may have different connectivity patterns since there is no constrain on the corresponding connection parameters \(\alpha^O\) and \(\alpha^I\). These two models are illustrated on a small multilevel network in Figure \(3\).
Figure 3: On the left: MLVSBM with inter-organisational level on the top and inter-individual level on the bottom. On the right: the same network with independent levels. The various levels of blue depict the clustering of the individuals and the various levels of red depict the clustering of the organisations. The parameters $\alpha$ over the plain links between nodes are the probabilities of connections given the nodes colours (clustering). The outer circles around the nodes of the individuals represent the blocks of the organisations they are affiliated to. The dashed links stand for affiliations.
Identifiability. Conditions for the identifiability of the MLVSBM are given in the proposition below.

**Proposition 2.** The MLVSBM is identifiable up to label switching under the following assumptions:

- **A1.** All coefficients of $\alpha^I \cdot \gamma \cdot \pi^O$ are distinct and all coefficients of $\alpha^O \cdot \pi^O$ are distinct.
- **A2.** $n_I \geq 2Q_I$ and $n_O \geq \max(2Q_O, Q_O + Q_I - 1)$.
- **A3.** At least $2Q_I$ organizations contain one individual or more.

The set of parameters that does not verify the first assumption $A1$ is null set with respect to Lebesgue measure. Also, the maximum number of detectable blocks in the SBM grows in $\sqrt{\#\text{nodes}}$ (Choi et al., 2012) which makes assumption $A2$ very weak in practice. The last assumption $A3$, on the affiliation, means that at least some organisations must not be empty and enough individuals belong to different organisations. The proof of this proposition is provided in the appendix, Section B.

3 Statistical Inference

We now present a maximum likelihood procedure and a criterion for model selection.

3.1 Variational method for maximum likelihood estimation

As said before, the likelihood of $X$ $\ell_\theta(X|A)$ is obtained by integrating out the latent variables $Z$ in the complete data likelihood (6). However, this calculus becomes not computationally tractable as the number of nodes and blocks grow.

The Expectation-Maximisation algorithm (EM) (Dempster et al., 1977) is a popular solution maximising the likelihood of models with latent variables. However it requires the computation of $P(Z|X,A)$ which is also not tractable in our case. The variational version of the EM algorithm is a powerful solution for such cases. It was first used for SBM by Daudin et al. (2008).

In a few words, the variational EM algorithm maximises the so-called variational bound i.e. a lower bound of the log likelihood $\ell_\theta(X|A)$ denoted $I_\theta(R(Z|X))$ and defined as follows:

$$I_\theta(R(Z|A)) := \mathbb{E}_R [\ell_\theta(Z, X|A)] + \mathcal{H}(R(Z|A))$$

$$= \ell_\theta(X|A) - KL(R(Z|A)||P_\theta(Z|X, A)) \leq \ell_\theta(X|A)$$

where $KL$ is the Kullback-Leibler divergence, $\mathcal{H}$ is the Shannon entropy: $\mathcal{H}(P) = \mathbb{E}_P[-\log(P)]$ and $R(Z|A)$ is an approximation of the true distribution $P_\theta(Z|X, A)$. In our context, and following Daudin et al. (2008), we propose to chose $R(Z|A)$ in a family of factorized distributions, resulting into a mean field approximation $R(Z|A)$ defined as:

$$R(Z|A) = \prod_i h_{\tau^I_i}(Z_{i}^I) \prod_j h_{\tau^O_j}(Z_{j}^O),$$

(7)
where $h$ is the multinomial distribution.

The variational EM algorithm consists of iterating two steps. Step $VE$ maximises the variational bound with respect to the parameters of the approximate distribution defined in equation (7). This is equivalent to minimising the Kullbach-Leibler divergence term. Step $M$ maximises the variational bound with respect to the model parameters $\theta$. The details of the calculus and algorithm are developed in the appendix (Section C).

### 3.2 Model selection

#### 3.2.1 Selection of the number of blocks

Following Biernacki et al. (2000) and Daudin et al. (2008), we propose a model selection criterion to choose the unknown numbers of blocks $Q_I$ and $Q_O$. The ICL criterion is an integrated version of BIC applied to the complete likelihood. In other words, it is an asymptotic approximation of the complete likelihood integrated over its parameters and latent variables. It has proven its practical efficiency for the SBM and its extensions. Our criterion is equal to:

$$ICL(Q_I, Q_O) = \log \ell_{\hat{g}}(X^I, X^O, \hat{Z}^I, \hat{Z}^O | A, Q_I, Q_O) - \text{pen}(Q_I, Q_O),$$

where

$$\text{pen}(Q_I, Q_O) = \frac{1}{2} \frac{Q_I(Q_I + 1)}{2} \log \frac{n_I(n_I - 1)}{2} + \frac{Q_O(Q_O - 1)}{2} \log n_O + \frac{1}{2} \frac{Q_O(Q_O + 1)}{2} \log \frac{n_O(n_O - 1)}{2} + \frac{Q_O - 1}{2} \log n_O,$$ (9)

where $\hat{Z}^O$ and $\hat{Z}^I$ are the imputed latent variables using the maximum a posteriori (MAP) of $P_{\hat{g}}(Z | X, A; Q_I, Q_O)$. The proof is provided in the appendix, Section D. As for the variational inference, $P_{\hat{g}}(Z | X, A; Q_I, Q_O)$ is unknown and, in practice, we replace it by its mean-field approximation $R_{\hat{g}}(Z; Q_I, Q_O).

**Remark.** Once again, note that the penalty (9) is adapted to undirected networks. For instance, the term $\frac{Q_I(Q_I + 1)}{2} \log \frac{n_I(n_I - 1)}{2}$ would become $Q_I^2 \log n_I(n_I - 1)$ if $X^I$ were not symmetric.

**Remark.** We recall that the penalty of the ICL for a (unilevel) SBM is given by

$$\text{pen}_{SBM}(Q) = \frac{1}{2} \frac{Q(Q + 1)}{2} \log \frac{n(n - 1)}{2} + \frac{Q - 1}{2} \log n.$$ (10)

The penalty term in Equation (9) for the inter-organisational level is the same as the one given in Equation (10). For the inter-individual network, the factor in front $\log n_I$ is $Q_O(Q_O - 1)$ instead of $Q_I - 1$ for the SBM as in Equation (10), that is the penalty term which corresponds to the degree of freedom of $\gamma$. 

3.2.2 Determining the independence between levels

The ICL criterion can also be used to assess whether the two levels of interactions are independent or not. If we force every columns of $\gamma$ to be identical, then the penalty term on $\gamma$ becomes $\frac{1}{2}(Q_I - 1) \log n_I$ and, as a consequence:

$$ICL_{Ind}(Q_I, Q_O) = ICL_{SBM}^O(Q_I) + ICL_{SBM}^I(Q_O).$$

(11)

The ICL criterion favours independence if

$$\max_{(Q_I, Q_O)} ICL_{MLVSBM}(Q_I, Q_O) \leq \max_{Q_I} ICL_{SBM}^O(Q_I) + \max_{Q_O} ICL_{SBM}^I(Q_O).$$

If this is the case, then the gain in term of likelihood does not compensate the gain $\frac{1}{2}(Q_O - 1)(Q_I - 1) \log n_I$ in the penalty. This criterion focuses on the dependence between levels given by the inter-level. Note that, in the degenerate case where one of the two levels has no structure (a unique block), then the ICL of the independent model $ICL_{Ind}(Q_I, Q_O)$ and the non constrained model $ICL_{MLVSBM}(Q_I, Q_O)$ are the same and we are not able to decide.

3.2.3 Procedure for model selection

We now provide a procedure for model selection which seeks for the optimal number of blocks at a reasonable cost. As a by-product, it states whether the two levels are independent or not.

The practical choice of the model and the estimation of its parameters are computational intensive tasks. Indeed, we should compare all the possible models (one model corresponding to given numbers of individual and organisational blocks) trough the ICL criterion. However, for each model, the variational EM algorithm has be started at a large number of initialization points (due to its sensitivity to the starting point), resulting in an unreasonable computational cost. Instead, we propose to adopt a stepwise strategy, resulting in a faster exploration of the model space combined with efficient initializations of the VEM algorithm. The procedure we suggest is given in Algorithm 1.

Each step of the algorithm requires $O(\max\{Q_I, Q_O\}^2)$ variational EM algorithms. The local initialisation allows a convergence in a few iterations of the VEM algorithms. Inferring an independent SBM on each level beforehand is a fast way to start with good initialisation and allows us to state on the independence of the model at the same time as we just need to compare the sum of the ICL of the two SBMs used for initialisation with the ICL of the best multilevel model.

In practice we used the R package blockmodels [Leger, 2016] to infer the SBM. To simulate and infer the MLVSBM, we developed our own R package available at https://chabert-liddell.github.io/MLVSBM/. It features the simulation and inference of multilevel networks with symmetric and/or asymmetric adjacency matrices, model and independence selection as described in this paper. It also handles missing at random data [Rubin, 1976] on the adjacency matrices and link prediction.
Algorithm 1: Model selection algorithm

Data: \( \{X^I, X^O, A\} \), a multilevel network
Result: Inference of \( X \)

Procedure

• Infer independent SBMs on \( X^I \) and \( X^O \). Deduce \( \arg \max_{Q_I} ICL_{SBM}^I(Q_I) \) and \( \arg \max_{Q_O} ICL_{SBM}^O(Q_O) \). Compute the corresponding \( ICL_{ind} \) from formula (11).

• Set initial clustering of size \((Q_I^{init}, Q_O^{init}) = (\arg \max_{Q_I} ICL_{SBM}^I(Q_I), \arg \max_{Q_O} ICL_{SBM}^O(Q_O))\) with the one obtained from the SBMs.

while \( ICL \) is increasing do

| Fit a MLVSBM on all model of size \((Q_I \pm 1, Q_O \pm 1)\) initialised by merging 2 blocks or splitting a cluster with hierarchical clustering. |
| Select for each model size the one with the highest variational bound. |
| Keep among all selected models the one with the highest \( ICL \). |

return The MLVSBM with the highest \( ICL \)

4 Illustration on simulated data

In this section, we study the performances of the inference procedure for the MLVSBM including the ability to recover blocks, the selection of the numbers of blocks and the independence detection.

Remark. In order to evaluate the ability to recover blocks, we resort to the Adjusted Rand Index (ARI) [Hubert and Arabie, 1985], which is a comparison index between two clusterings with a correction for chance. This index is close to 0 when the two clusterings are independent and is 1 when the clusterings are identical (up to label switching).

4.1 Experimental design

In what follows, we set \( Q_O = Q_I = 3 \) and the number of nodes to \( n_O = 20Q_O = 60 \) for the inter-organisational level and \( n_I = 60Q_I = 180 \) for the inter-individual level.

Let \( d \) be a density parameter: the lower \( d \), the sparser the network and the harder the inference. \( \epsilon (\geq 1) \) is a parameter tuning the strength of the communities; when \( \epsilon \) is high, the communities are easily separable.

In the simulation study, we focus on the three following standard topologies:

• **Assortative communities.** The probability of connection within communities is higher than the probability of connection between communities: \( \alpha^I = d \begin{bmatrix} \epsilon & 1 & 1 \\ 1 & \epsilon & 1 \\ 1 & 1 & \epsilon \end{bmatrix} \).

• **Disassortative communities.** The probability of connection within communities is
lower than the probability of connection between communities: $\alpha = d^* \begin{bmatrix} 1 & \epsilon & \epsilon \\ \epsilon & 1 & \epsilon \\ \epsilon & \epsilon & 1 \end{bmatrix}$.

- **Core-periphery.** A core cluster is highly connected to the whole network while the probability of connection in the periphery is low: $\alpha = d^* \begin{bmatrix} \epsilon & \epsilon & 1 \\ \epsilon & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$.

We fix the topology of the inter-organisational level $X^O$ to be an assortative communities with $d = 0.1$, $\epsilon = 5$ and of communities of equal size on average. We expect this topology to be easy to infer and to obtain a perfect recovery of the clustering with high probability. For the inter-individual level, $d$ is set equal to 0.01, 0.05 or 0.1 while $\epsilon$ ranges from 1 to 10 by step of 0.5. When $\epsilon = 1$, the topology is the one of an Erdős-Rényi graph and the communities should be indistinguishable. A power-law distribution is chosen for the affiliation matrix in order to get different sizes of organisations. Other distributions were tried but the results (not reported here) show that their impact on inference is weak. Finally, $\delta$ is a parameter for the strength of the dependence between levels, ranging from 0 to 1. More precisely, we set:

$$\gamma = \begin{bmatrix} \frac{\delta}{2} & \frac{1}{2}(1 - \delta) & \frac{1}{2}(1 - \delta) \\ \frac{1}{2}(1 - \delta) & \delta & \frac{1}{2}(1 - \delta) \\ \frac{1}{2}(1 - \delta) & \frac{1}{2}(1 - \delta) & \delta \end{bmatrix}$$

where $\gamma$ has been defined in Equation 11. $\delta = 1$ implies a deterministic link between the clustering of the two levels, i.e. the cluster of an individual is fully determined by the cluster of his/her organisation. When $\delta = 0$, the diagonal coefficients of $\gamma$ are null so individuals cannot belong to a specific block of individuals. When $\delta = 1/Q_I$ the levels are independent.

### 4.2 Simulation results

First, we fix $\delta = 0.8$ and make $\epsilon$ varies. We compare the ability of our model to recover the true clustering of $Z^I$ from $X^I$. We compare our performances to the ones obtained by applying a standard (unilevel) SBM on $X^I$. The inferred clusterings are compared to the true one (simulated) through the ARI. Note that, the number of groups ($Q_I$) being assumed as unknown, two types of error may occur: one for not selecting the right number of blocks and one for assigning nodes to the wrong blocks. The results are displayed in Figure 4.

In Figure 4 (A), we plot–for 3 values of density $d$ and 3 topologies–the ARI when using MLSBM (plain line) and SBM (dashed line) as $\epsilon$ varies. We observe that, for any topology (assortative, core-periphery and disassortative), the MLVSBM starts to recover perfectly the clustering for a lower value of $\epsilon$ than the SBM. The difficulty of the inference increases as $\epsilon$ decreases: as can be seen in Figure 4 A, MLVSBM still performs well ($ARI > 0$) for small values of $\epsilon$ while standard SBM is unable to recover the clustering.
In Figures 4B and C we plot the number of blocks chosen by the MLSBM (B) and the SBM (C) for 3 values of density (rows) and 3 topologies (columns) (the true value being $Q_I = 3$). We observe that the number of blocks chosen by the ICL criterion for the inter-individual level is also closer to the true number of blocks for the multilevel model. It varies from 1, when no structure is detected to 3 which is the true number of blocks. It never selects more blocks than expected. This is coherent with prior knowledge that ICL tends to underestimate model size for the SBM (Hayashi et al., 2016; Brault, 2014).

Figure 4: Clustering and model selection for 3 different topologies on the inter-individual level, varying $\epsilon$ and density $d$. Each situation is simulated 50 times. A: ARI for the inter-individual level, comparing the model used for inference. B: Number of blocks for the inter-individual level chosen by the ICL criterion for the Multilevel SBM (in blue). C: Number of blocks for the inter-individual level chosen by the ICL criterion for the SBM (in red).

On the three topologies, with $\epsilon = 3$, depending on the density $d$, we obtain either a perfect recovery of the clustering or a random clustering or something in between for the inter-individual level, for both MLVSBM and SBM. In order to understand better this phenomenon, we fix $\epsilon$ to 3 and make $\delta$ varies. The results are reported in Figure 5. When $\delta = 1/3$ (yellow vertical line in Figure 5A), the two levels are independent and as shown in Figure 5A, the results in term of clustering is the same for the MLVSBM and the SBM on $X^I$. As soon as $\delta$ departs from this value, MLVSBM is able to recover some of
the structure of the lower level thanks to the inter-organisational level and this ability is observed even for very low density when $\delta$ gets closer to 1 (see Figure 5 A and B). Figure 5 C depicts the performances of the ICL criterion to state on the independence between the two levels. For $d = 0.01$, the inter-individual level is very sparse and no structure is detected on the inter-individual level and the chosen number of individuals blocks $Q_I$ is equal to 1. In this case, $ICL_{ind} = ICL_{MLVSBM}$ and we cannot choose between the two models. In the simulations, ICL were close for values of $\delta$ far from 1 but we decided to break ties in favour of the simpler model by saying that levels are dependent only if $ICL_{MLVSBM} \geq ICL_{ind} + 10^{-6}$. For higher densities, we see, as expected, that the further $\delta$ departs from $1/3$ the more the multilevel model is preferred, no matter if the two models supply the same clusterings or not. However the independent model is sometimes selected when $\delta$ is not too far from $1/3$. This is a consequence of the conservative nature of ICL, requiring strong evidence from the likelihood to select a more complex model.

![Figure 5](image)

Figure 5: Clustering and model selection for 3 different topologies on the inter-individual level, as function of $\delta$ and density $d$. Each situation is simulated 50 times. The yellow vertical lines represent a $\delta = 1/3$ (i.e. a $\gamma$ with uniform coefficients, resulting into independence between the two levels). A: ARI for the inter-individual level, comparing the model used for inference. B: Number of blocks for the inter-individual level chosen by the ICL criterion for the MLVSBM. C: Model selected by the ICL for the inter-level dependence.
Simulations gave similar results (not reported here) when we inverse the topologies of the inter-individual and inter-organisational level showing that information on structure transit in both ways. Moreover, when the number of nodes of the "easy-to-infer" level increases, it facilitates the inference the clustering on the "hard-to-infer" level.

5 Application to the multilevel network issued from a television programs trade fair

We apply our model to the data set (Brailly et al., 2016) described below.

5.1 Context and Description of the data set

Promoshow East is a pseudonym for the largest television programs trade fair for Eastern Europe. Sellers from Western Europe and the USA come to sell audiovisual products to regional and local buyers such as broadcasting companies. The audiovisual products are divided into many different categories and the data come from the sub-sector of animation and cartoons. From a sociological perspective, reconstituting and analysing multilevel (inter-individual and inter-organisational) networks in this industry is important. In economic sociology, it helps redefine the nature of markets (Brailly et al., 2016; Lazega and Mounier, 2002). In the sociology of culture, it helps understand, from a structural perspective, the mechanisms underlying contemporary globalisation and uniformisation of culture (Brailly et al., 2016; Favre et al., 2016). In the sociology of organisations and collective action, it helps understand the importance of multilevel relational infrastructures for the management of tense competition and cooperation dilemmas by various categories of actors (Lazega, 2019), in this case the (sophisticated) sales representatives of cultural industries.

Data of the interactions for both levels where collected by face-to-face interview. At the individual level, people were asked to select from a list the individuals from which they obtain advice or information during or before the trade fair. The level consists of 128 individuals and 710 directed interactions (density = 0.044). Those individuals were affiliated to 109 organisations, each one containing between one and six individuals. Two possible kinds of interactions at the organisation level can be considered: a deal network and a meeting network. The deal network was built by asking to the participants with whom they signed a deal since the last trade fair 12 months ago. Then, these answers were aggregated at the organisation level. The meeting network comes from the aggregation at the organisation level of the meetings planned by individuals on the trade fair’s website. The meeting network is symmetric and the deal network was symmetrised by duplicating the few interactions that did not have a symmetric counterpart. The respective densities are 0.067 and 0.059.
5.2 Statistical analysis

The MLVSBM is inferred on the two datasets (one dataset corresponding to the deal network at the inter-organisational level, the other dataset to the meeting network at the inter-organisational level). In both cases the ICL criterion favours dependence between the two levels and chooses 4 blocks of individuals. The number of blocks of organisations is equal to 3 for the deal network and 4 for the meeting network. In order to determine which is the most relevant inter-organisational network, we test the ability of the MLVSBM to predict dyads or links in the inter-individual network when the deal or the meeting networks are considered. To do so, we choose uniformly dyads and links to remove and try to predict them. More precisely, we remove a certain percentage of all dyads i.e. we set $X_{ii'} = \text{NA}$ (this percentage ranging from 5% to 40% by stepsize of 5%). We also propose to remove existing links (i.e. forcing $X_{ii'}^I = 0$ when $X_{ii'} = 1$ was observed, for some randomly chosen $(i,i')$). The percentage of removed existing links varies from 5% to 95% (with stepsize of 5%). We repeat the following procedure 100 times for each step:

1. Remove dyads or links uniformly at random
2. Infer the newly obtained network from scratch in order to obtain the probability of a link $P(X_{ii'}^I = 1; \hat{\theta})$ for each missing dyad or for each dyad such that $X_{ii'}^I = 0$
3. Predict link among all missing dyads or among all dyads such that $X_{ii'}^I = 0$.

Missing data are handled as Missing At Random (Tabouy et al., 2019) and the probability of existence of an edge is given by: $P(X_{ii'}^I = 1; \hat{\theta}) = \sum_{k,k'} \hat{\tau}_{ik} \hat{\alpha}_{kl} \hat{\tau}_{i'k'}$. Since the result of our procedure is equivalent to a binary classification problem, we compare those results by using the area under the ROC curve (AUC), with a random classification giving an AUC of 0.5. Figure 6 shows that using the MLVSBM compared to a single level SBM improves a lot the recovery of the inter-individual level for this dataset. This confirms the dependence between levels detected by the ICL. Also using the deal network gives better predictions for both missing dyads and missing links than the meeting network. We also considered a merged network at the inter-organisational level by making the union of links of the deal and the meeting network. The improvement in term of prediction over the deal network is not very significant and this composite network is much harder to analyse sociologically.

5.3 Analysis and comments

For the analysis, we use the model chosen by the ICL for the MLVSBM with the deal network as inter-organisational level. We estimate 3 blocks of organisations, 4 blocks of individuals and the inter-individual level and the inter-organisational level are dependent. This network is plotted in Figure 7 B.

In Figure 7 A, we plot a synthetic view of the blocks of this multilevel network and in Figure 7 we reordered the adjacency matrices of both levels by blocks. At the
inter-organisational level, the first cluster (in red) is a residual group composed of 61 organisations that are weakly connected to the rest of the organisations. The second cluster (in orange) consists of customers: broadcasters that come to the trade fair to buy programs and independent buyers who buy programs, planning to sell them later to broadcasters. We observe a non-null intra-cluster connection, but deals are mainly done between organisations of the second and third clusters (in yellow), the latter mostly containing distributors.

At the inter-individual level, the blocks 1 and 2 consist of buyers (exclusively for Block 1). They differ in their affiliations, both are affiliated to the second block of organisations but a larger proportion of the individuals of block 2 are affiliated to the residual cluster of organisations. They also differ in the way they connect to blocks 3 and 4. Block 4 is a residual group consisting of roughly half of the individuals. It does not exhibit any particular pattern in its affiliations and is weakly connected, mainly inward connection from block 2. Block 3 consists of sellers giving advices to individuals of block 2 and has reciprocal relationship with individuals of block 1. They are mainly affiliated to producing and distributing companies of the third block of organisations. It is also the block that has the strongest intra-cluster connections.

These results confirm neo-structural insights into the functioning of markets. Competition between producers/distributors is strong: they all need to find broadcasting companies and distributors on the buying side. However, most of them arrive to the trade fair without updated information about the products in which buyers are interested in that year, their available budgets for each category of product, their willingness to negotiate, etc. The value of multilevel network analysis that is used here is to show that inter-individual personal relationships between individuals affiliated with competing organisations help manage the tensions between these directly competing organisations.
Figure 7: Multilevel network of the Promoshow East trade fair 2011. Above: the deal network for the organisations on the top and individuals on the bottom. The size of the nodes is proportional to the size of the blocks. For individuals the gradient on the edge represents directed arrows from light to dark. The donut charts around each node give the proportion of the block membership of their respective organisations. B: View of the network. The size of a node is proportional to its centrality degree. Colours represent the clustering obtained with the multilevel SBM. C: Adjacency matrices of the advice network for individuals and the deal network for the organisations. Entries are reordered by block from left to right and top to bottom.
This is where personal ties between individuals affiliated in these companies, especially among sellers and buyers, but also less visibly among sellers, are important: they help manage the strong tensions between companies by creating coopetition, cooperation among their competing firms. Here social/advice ties between buyers (blocks 1 and 2 of individuals) affiliated to buying companies in block 2 of organisations (broadcasting companies and distributors) exchange advice from sellers of block 3 representing production and distribution companies: this is the normal, stabilised, overlapping, commercial ties between companies embedded in social ties between representatives.

As seen above, block 3 has strong intra-cluster connections which signals discreet coordination efforts between sellers as shown by Brailly (2016); Brailly et al. (2016). When a seller has closed a deal with a buyer, s/he can advise and update another seller – i.e. a coopetitor in terms of affiliation to a competing company – about other products in which this buyer is interested, what budget is left in his/her pocket, i.e. precious information for the next sellers. This kind of personal service is expected to be reciprocated over the years; otherwise the relationship decays. It is the density of block 3 of individuals that represents the most unexpected phenomenon from an orthodox economic perspective and the most crucial phenomenon from the perspective of neo-structural economic sociology (Lazega and Mounier, 2002).

This cross-level interdependence between inter-organisational ties and inter-individual ties is strong enough for companies to be unable to lay off its sales representatives. Having long tried to replace costly trade fairs with online websites and catalogues, companies realised that they still need the service that real persons and their personal relational capital provide in terms of multilevel management of coopetition (Lazega, 2019).

6 Discussion

In this paper, we propose a stochastic block model for multilevel networks. We develop variational methods for the inference of the model and a criterion that allows us to choose the number of blocks and to state on the independence between levels at the same time. There are clear advantages at considering a joint modelling of the two levels over an independent model for each level. Indeed, we show on simulation studies that when we detect dependence between levels, it helps us to recover the block structure of a level with low signal thanks to the structure of the other level and also to improve the prediction of missing links or dyads.

Instead of Bernoulli, the edge distribution of any level may be extended to valued distribution in a similar way as SBM (Mariadassou et al., 2010). This includes extension to the Degree Corrected SBM (Karrer and Newman, 2011) for example by using nodes degree as edges covariates.

Our choice to model the interaction levels given the affiliations (A being fixed) is driven by the fact that, in a lot of applications, these affiliations are known and the object of the analysis is the interactions. We choose to restrict the number of affiliations to one as this was the case on the data sets available to us, but this approach could be
Table 1: Contingency table of covariates and clustering for the organisations (top) and the individuals (bottom)

| Organisations | Covariates          | Block | Size | Producer | Distributor | Media group | Independent | Broadcaster | Buyer |
|---------------|---------------------|-------|------|----------|-------------|-------------|-------------|-------------|-------|
|               |                     | 1     | 61   | 14       | 16          | 9           | 14          | 8           |
|               |                     | 2     | 20   | 1        | 0           | 2           | 7           | 10          |
|               |                     | 3     | 28   | 3        | 19          | 5           | 1           | 0           |

| Individuals   | Covariates          | Affiliation |
|---------------|---------------------|-------------|
|               |                     | Block | Size | Buyer | Seller | 1 | 2 | 3 |
|               |                     | 1     | 18   | 18    | 0      | 6 | 12| 0 |
|               |                     | 2     | 22   | 16    | 6      | 13| 8 | 1 |
|               |                     | 3     | 25   | 2     | 23     | 7 | 0 | 18|
|               |                     | 4     | 63   | 15    | 48     | 42| 8 | 13|

extended to a less restricted number of affiliations (implemented in our R package). We could even consider any hierarchical structure such as multi-scale networks to model the levels given the hierarchy or more generally multi-layer networks by modelling the layers given the inter-layers.

One way to construct a multilevel SBM that models jointly the affiliations and the interactions would be to make the affiliation depends on both the block memberships of the individuals and the organisations. But, doing so, we must make an assumption on how individuals are affiliated within a block of organisations. In a work not presented in this paper, we built such a model assuming that within block the affiliations were uniform. Simulation studies showed that the obtained clustering is very similar to the one obtained with the model presented in this paper, when the organisations belonging to the same block are of comparable sizes. In the contrary, if the sizes of the organisations in each block vary a lot, the clustering of the organisations depends more of their sizes than of their interactions. As a consequence, we think that, unless we want to specifically study the affiliations, our model is less restrictive.

Furthermore, our model is able to decide about the independence of the structure of connections of the two levels. This is done by a model selection criterion. It would be interesting to test (in a statistical meaning) this independence but we know that the variance of our estimators is underestimated because of the variational approach (see [Blei et al. 2017](#) for a review). Besides, sociological studies stated that some individuals benefit more than others from their organisation’s interactions ([Lazega and Snijders 2015](#)), which could lead us to consider more local independence between levels.

For multiplex networks, [De Bacco et al. 2017](#) use dyad predictions as a way to
define interdependence between layers while [Stanley et al.] (2016) make clusters layer by aggregating the most similar. Our work considers multilevel networks where each level have nodes of different natures and Figure 3 shows that dependence between levels leads to a better recovery of missing information. This can be used to help data collection or to correct spurious information on existing data as suggested in [Clauset et al.] (2008) or [Guimerà and Sales-Pardo] (2009). Indeed, the data of one level can be easier to collect or to verify than the other one as it is public, already exists or is just cheaper to collect. Thus, we think that this approach could help scientists from the social science or ecology in their sampling efforts.

Acknowledgements

The authors would like to thank Julien Brailly for providing the dataset. This work was supported by a public grant as part of the Investissement d’avenir project, reference ANR-11-LABX-0056-LMH, LabEx LMH. This work was partially supported by the grant ANR-18-CE02-0010-01 of the French National Research Agency ANR (project EcoNet). This project received financial support from INRA and CIRAD as part of the SEARS project funded by the GloFoods metaprogram. This work was presented and discussed within the framework of working days organized by the MIRES group (with the financial support of INRA) and the GDR RESODIV (with the financial support of CNRS).

References

Airoldi, E. M., D. M. Blei, S. E. Fienberg, and E. P. Xing (2008). Mixed membership stochastic blockmodels. *Journal of machine learning research* 9(Sep), 1981–2014.

Bar-Hen, A., P. Barbillon, and S. Donnet (2018). Block models for multipartite networks. Applications in ecology and ethnobiology. *arXiv preprint arXiv:1807.10138*.

Barbillon, P., S. Donnet, E. Lazega, and A. Bar-Hen (2017). Stochastic block models for multiplex networks: an application to a multilevel network of researchers. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 180(1), 295–314.

Bianconi, G. (2018). *Multilayer Networks: Structure and Function*. Oxford University Press.

Biernacki, C., G. Celeux, and G. Govaert (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE transactions on pattern analysis and machine intelligence* 22(7), 719–725.

Blei, D. M., A. Kucukelbir, and J. D. McAuliffe (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association* 112(518), 859–877.

Brailly, J. (2016). Dynamics of networks in trade fairs—a multilevel relational approach to the cooperation among competitors. *Journal of Economic Geography* 16(6), 1279–1301.
Brailly, J., C. Comet, S. Delarre, F. Eloire, G. Favre, E. Lazega, L. Mounier, J. Montes-Lihn, M. Oubenal, E. Penalva-Icher, and A. Piña-Stranger (2017). Neo-structural economic sociology beyond embeddedness. economic sociology_the european electronic newsletter 19(3), 36–49.

Brailly, J., G. Favre, J. Chatellet, and E. Lazega (2016). Embeddedness as a multilevel problem: A case study in economic sociology. Social Networks 44, 319–333.

Brault, V. (2014). Estimation et sélection de modèle pour le modèle des blocs latents. Ph. D. thesis, Université Paris Sud-Paris XI.

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.

Choi, D. S., P. J. Wolfe, and E. M. Airoldi (2012). Stochastic blockmodels with a growing number of classes. Biometrika 99(2), 273–284.

Clauset, A., C. Moore, and M. E. Newman (2008). Hierarchical structure and the prediction of missing links in networks. Nature 453(7191), 98.

Côme, E. and P. Latouche (2015). Model selection and clustering in stochastic block models based on the exact integrated complete data likelihood. Statistical Modelling 15(6), 564–589.

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

De Bacco, C., E. A. Power, D. B. Larremore, and C. Moore (2017). Community detection, link prediction, and layer interdependence in multilayer networks. Physical Review E 95(4), 042317.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the em algorithm. Journal of the Royal Statistical Society: Series B (Methodological) 39(1), 1–22.

Favre, G., J. Brailly, J. Chatellet, and E. Lazega (2016). Inter-organizational network influence on long-term and short-term inter-individual relationships: The case of a trade fair for tv programs distribution in sub-saharan africa. In Multilevel network analysis for the social sciences, pp. 295–314. Springer.

Guimerà, R. and M. Sales-Pardo (2009). Missing and spurious interactions and the reconstruction of complex networks. Proceedings of the National Academy of Sciences 106(52), 22073–22078.

Hayashi, K., T. Konishi, and T. Kawamoto (2016). A tractable fully bayesian method for the stochastic block model. arXiv preprint arXiv:1602.02256.
Hileman, J. and M. Lubell (2018). The network structure of multilevel water resources governance in central america. *Ecology and Society* 23(2).

Holland, P. W., K. B. Laskey, and S. Leinhardt (1983). Stochastic blockmodels: First steps. *Social networks* 5(2), 109–137.

Hubert, L. and P. Arabie (1985). Comparing partitions. *Journal of classification* 2(1), 193–218.

Jordan, M. I., Z. Ghahramani, T. S. Jaakkola, and L. K. Saul (1999). An introduction to variational methods for graphical models. *Machine learning* 37(2), 183–233.

Karrer, B. and M. E. Newman (2011). Stochastic blockmodels and community structure in networks. *Physical review E* 83(1), 016107.

Kivelä, M., A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter (2014). Multilayer networks. *Journal of complex networks* 2(3), 203–271.

Kolaczyk, E. D. (2009). *Statistical Analysis of Network Data: Methods and Models* (1st ed.). Springer Publishing Company, Incorporated.

Latouche, P., E. Birmele, and C. Ambroise (2012). Variational bayesian inference and complexity control for stochastic block models. *Statistical Modelling* 12(1), 93–115.

Lazega, E. (2009). Theory of cooperation among competitors: A neo-structural approach. *Sociologica* 1, 1–34.

Lazega, E. (2019+). Bureaucracy, collegiality and social change: Redefining organization with multilevel relational infrastructures. Forthcoming paper.

Lazega, E., A. Bar-Hen, P. Barbillon, and S. Donnet (2016). Effects of competition on collective learning in advice networks. *Social Networks* 47, 1–14.

Lazega, E., M.-T. Jourda, L. Mounier, and R. Stofer (2008). Catching up with big fish in the big pond? multi-level network analysis through linked design. *Social Networks* 30(2), 159–176.

Lazega, E. and L. Mounier (2002). Interdependent entrepreneurs and the social discipline of their cooperation: a research programme for structural economic sociology in a society of organizations. *Conventions and structures in economic organization: markets, networks, and hierarchies*, Cheltenham, Edward Elgar Publishing.

Lazega, E. and T. A. Snijders (2015). *Multilevel network analysis for the social sciences: Theory, methods and applications*, Volume 12. Springer.

Leger, J.-B. (2016). Blockmodels: A r-package for estimating in latent block model and stochastic block model, with various probability functions, with or without covariates. *arXiv preprint arXiv:1602.07587*. 

24
A Proof of Proposition 1

Proposition 1. In the MLVSMB, the two following properties are equivalent: [1.]: $Z^I$ is independent on $Z^O$, [2.]: $\gamma_{kl} = \gamma_{kl}'$ $\forall l, l' \in \{1, \ldots, Q_O\}$ and imply that: [3.]: $X^I$ and $X^O$ are independent.
Proof. We first derive an expression for $\ell_\gamma(Z^I) = \ell_\gamma(Z^I|A)$:

$$\ell_\gamma(Z^I|A) = \int_{Z^O} \ell_\gamma(Z^I|A, Z^O) dP(Z^O)$$

$$= \sum_{l_1, \ldots, l_{n_O}} \ell_\gamma(Z^I|A, Z^O = l_1, \ldots, Z^O = l_{n_O}) P(Z^O = l_1, \ldots, Z^O = l_{n_O})$$

$$= \sum_{l_1, \ldots, l_{n_O}} \prod_j \left( \prod_i \ell_\gamma(Z^I_i|A, Z^O_A = l_A) \right) P(Z^O = l_j)$$

$$= \sum_{l_1, \ldots, l_{n_O}} \prod_j \left( \prod_{i,k} \gamma_1 Z^I_{i,k} A_{ij} \right) \prod_j \sum_l \prod_{i,k} A_{ij} Z^I_{i,k} \gamma_1 Z^O_{l_j} \approx \prod_j \sum_l \prod_{i,k} A_{ij} Z^I_{i,k} \gamma_1 Z^O_{l_j}$$

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

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

$$\ell_\gamma(Z^I|Z^O, A) = \prod_{k,l} \sum_{i,j} A_{ij} Z^I_{i,k} Z^O_{j,l} = \prod_{k} \sum_{i,j} A_{ij} Z^I_{i,k} Z^O_{j,l}$$

$$= \prod_{i,k} \sum_{l} A_{ij} Z^I_{i,k} Z^O_{l_j}$$

and

$$\ell_\gamma(Z^I|A) = \prod_{j} \sum_{l} \prod_{i,k} A_{ij} Z^I_{i,k} Z^O_{l_j}$$

$$= \prod_{j} \sum_{l} \prod_{i,k} A_{ij} Z^I_{i,k} Z^O_{l_j}$$

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 $\ell_\gamma(Z^I_1|Z^O, A) = \ell_\gamma(Z^I_1|A)$. Assuming that individual 1 belongs to organisation $j$, we can write, for any $k$:

$$P(Z^I_1 = k|Z^O, A_{ij} = 1) = \gamma_k Z^O_j.$$ 

However, this quantity does not depend on $Z^O_j$ so $\gamma_k Z^O_j = \gamma_k$ for any value of $k$ and $Z^O_j$. And so we have $\gamma_{kl} = \gamma_{kl'}$ for any $(\ell, \ell')$.  

26
1. ⇒ 3.:

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

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

$$= \int_{z^I} \ell_{\alpha^I}(X^I|Z^I = z^I) \mathbb{P}(Z^I = z^I) dz^I \int_{z^O} \ell_{\alpha^O}(X^O|Z^O = z^O) \mathbb{P}(Z^O = z^O) dz^O$$

$$= \ell_{\alpha^I}(X^I) \ell_{\alpha^O}(X^O)$$

which is the definition of the independence.

\[\square\]

B Proof of Proposition 2

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

A1. All coefficients of $\alpha^I \cdot \gamma \cdot \pi^O$ are distinct and all coefficients of $\alpha^O \cdot \pi^O$ are distinct.

A2. $n_I \geq 2Q_I$ and $n_O \geq \max(2Q_O, Q_O + Q_I - 1)$.

A3. 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 $\theta$ 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 $\pi^O$ and $\alpha^O$—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:

$$x_{ik} := (x_i, \ldots, x_k), \quad X_{j,ik} = (X_{ji}, \ldots, X_{jk}).$$

Moreover, $\{X_{j,ik} = 1\}$ stands for $\{X_{ji} = 1, \ldots, X_{jk} = 1\}$.

Identifiability of $\pi^O$. For any $l = 1, \ldots, Q_O$, let $\tau_l$ be the following probability:

$$\tau_l = \mathbb{P}(X_{ij}^O = 1|Z_i^O = l) = \sum_{l'} \alpha_{il}^O \pi_{l'}^O = (\alpha^O \cdot \pi^O)_l, \quad \forall (i,j).$$  \hspace{1cm} \text{(B.12)}$$

Moreover, a quick computation proves that

$$\mathbb{P}(X_{i,j(k+l)}^O = 1|Z_i^O = l) = \tau_{l}^{k+1} \hspace{1cm} \text{(B.13)}$$
According to Assumption $\mathcal{A}1$, the coordinates of vector $(\tau_1, \ldots, \tau_{Q_O})$ are all different. Hence, the Vandermonde matrix $R^O$ of size $Q_O \times Q_O$ such that

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

is invertible. We define $u^O_i$ as follows:

$$u^O_i = \mathbb{P}_{X, \theta}(X_{1,2:(i+1)}^O = 1) \quad \text{for } 1 \leq i \leq 2Q_O - 1$$

$$u^O_0 = 1.$$ 

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

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

$$u^O_i = \sum_{l=1}^{Q_O} \mathbb{P}(X_{1,2:(i+1)}^O = 1) \mathbb{P}(Z^O_1 = l) \mathbb{P}(Z^O_i = l) = \sum_{l=1}^{Q_O} \tau_i^l \pi_i^O,$$

using equation (B.13). Now, let us define $M^O$ a $(Q_O + 1) \times Q_O$ matrix such that:

$$M^O_{ij} = u^O_{i+j-2} = \sum_{l=1}^{Q_O} \tau_i^{j-1} \pi_i^O \tau_l^j, \quad 1 \leq i \leq Q_O + 1, \quad 1 \leq j \leq Q_O. \quad (B.14)$$

For $k \in \{1, \ldots, Q_O + 1\}$, we define $\delta_k$ as $\delta_k = \text{Det}(M^O_k)$ 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:

$$B^O(x) = \sum_{k=0}^{Q_O} (-1)^{k+Q_O} \delta_{k+1} x^k. \quad (B.15)$$

- $B^O$ is of degree $Q_O$. Indeed, $\delta_{Q_O+1} = \text{Det}(M^O_{-Q_O})$ and $M_{-(Q_O+1)} = R^O D_{\pi^O} R^O'$ where $D_{\pi^O} = \text{diag}(\pi^O)$. As a consequence, $M_{-(Q_O+1)}$ 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, $B^O(\tau_l) = \text{det}(N^O_l)$ where $N^O_l$ is the concatenated matrix $N^O_l = (M^O | V_l)$ with $V_l = [1, \tau_1, \ldots, \tau^Q_O]$ (computation of the determinant development against the last column). However, from Equation (B.14), we have $M^O_j = \sum_l \tau_i^{j-1} \pi_i^O V_l$, i.e. each column vector of $M^O$ is a linear combination of $V_1, \ldots, V_{Q_O}$. As a consequence, $\forall l = 1, \ldots, Q_O, \ N^O_l$ 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 $\mathbb{P}_{X, \theta}$. 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 $D_{\pi^O} = R^O^{-1} M_{-Q_O} R^O'$, we can identify $\pi^O$ in a unique way.
Identifiability of $\alpha^O$. For $1 \leq i, j \leq Q_O$, we define $U_{ij}$ as follows:

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

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

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

and as consequence $U^O = R^O D_{\pi_O} \alpha^O D_{\pi_O} R^O'$. $D_{\pi_O}$ and $R^O$ being invertible, we get:

$$\alpha^O = D_{\pi_O}^{-1} R^O^{-1} U^O R^O' D_{\pi_O}^{-1} \pi_O.$$

And so $U^O$ is uniquely derived from $P_X$, so $\alpha^O$ is identified.

Identifiability of $\alpha^I$. To identify $\alpha^I$, 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 $A3$).

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

  $$\sigma_k = P(X_{1i}^I = 1|Z_1^I = k, A) = \sum_{k'} P(X_{1i}^I = 1|Z_1^I = k, Z_1^I = k') P(Z_1^I = k'|Z_1^I = k, A).$$

Moreover,

$$P(Z_1^I = k'|Z_1^I = k, A) = \sum_l P(Z_1^I = k'|Z_1^O = l, Z_1^I = k, A) P(Z_1^O = l|Z_1^I = k, A)$$

$$= \sum_l \gamma_{kl} P(Z_1^O = l|Z_1^I = k, A). \quad (B.16)$$

However, by Bayes’ formula

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

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: $P(Z_1^I = k|Z_1^O = l, A) = P(Z_1^I = k|A)$. And so

$$P(Z_1^O = l|Z_1^I = k, A) = P(Z_1^O = l|A) = \pi_1^O.$$

Consequently, from equation $(B.16)$, we have:

$$P(Z_1^I = k'|Z_1^I = k, A) = \sum_l \gamma_{k'l} \pi_1^O.$$
and so:

\[ \sigma_k = \sum_{k'} \mathbb{P}(X_{1i}^l = 1 | Z_{1i}^l = k, Z_{1i}^l = k') \sum_{l} \gamma_{k'l} \pi_{k}^{O} \]

\[ = \sum_{k'l} \alpha_{kk'} \gamma_{k'l} \pi_{k}^{O} = (\alpha^{I} \cdot \gamma \cdot \pi^{O})_{k} \]

\[ = (\alpha^{I} \cdot \pi^{I})_{k}, \quad \text{where} \; \pi^{I} = \gamma \cdot \pi^{O}. \]

• Now, we prove that \( \forall i = 1, \ldots, 2Q_{I} - 1, \)

\[ \mathbb{P}(X_{1;2:(i+1)}^l = 1 | Z_{1}^l = k, A) = \sigma_{k}^{i}. \quad \text{(B.17)} \]

Indeed,

\[ \mathbb{P}(X_{1;2:(i+1)}^l = 1 | Z_{1}^l = k, A) = \sum_{k_{2:(i+1)}} \mathbb{P}(X_{1;2:(i+1)}^l = 1 | Z_{1:(i+1)}^l = (k, k_{2:(i+1)}), Z_{1}^l = k) \mathbb{P}(Z_{1:(i+1)}^l = k_{2:(i+1)} | Z_{1}^l = k, A) \]

\[ = \sum_{k_{2:(i+1)}} \mathbb{P}(X_{1;2:(i+1)}^l = 1 | Z_{1:(i+1)}^l = (k, k_{2:(i+1)}) ) \mathbb{P}(Z_{2:(i+1)}^l = k_{2:(i+1)} | A) \]

\[ = \sum_{k_{2:(i+1)}} \mathbb{P}(X_{1;2:(i+1)}^l = 1 | Z_{1:(i+1)}^l = (k, k_{2:(i+1)}) ) \sum_{l_{2:(i+1)}} \mathbb{P}(Z_{2:(i+1)}^l = k_{2:(i+1)}, Z_{2:(i+1)}^O = l_{2:(i+1)}, A). \]

Note that, to go from line 2 to line 3, we used the fact that \( \mathbb{P}(Z_{1:(i+1)}^l = k_{2:(i+1)} | Z_{1}^l = k, A) = \mathbb{P}(Z_{2:(i+1)}^l = k_{2:(i+1)} | A), \)

which is due to the particular structure of \( A \) (left diagonal block of size at least \( 2Q_{I} \), i.e. for any \( i' = 1, \ldots, 2Q_{I} \), individual \( i' \) belongs to organisation \( i' \)). Moreover, we can write:

\[ \mathbb{P}(Z_{2:(i+1)}^l = k_{2:(i+1)}, Z_{2:(i+1)}^O = l_{2:(i+1)} | A) \]

\[ = \prod_{\lambda=2,\ldots,i+1} \mathbb{P}(Z_{\lambda}^l = k_{\lambda} | Z_{\lambda}^O = l_{\lambda}) \mathbb{P}(Z_{\lambda}^O = l_{\lambda}) \]

\[ = \prod_{\lambda=2,\ldots,i+1} \gamma_{k_{\lambda}l_{\lambda}} \pi_{k_{\lambda}}^{O}. \]

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

\[ \mathbb{P}(X_{1;2:(i+1)}^l = 1 | Z_{1}^l = k, Z_{2:(i+1)}^l = k_{2:(i+1)}) = \prod_{\lambda=2,\ldots,i+1} \alpha_{kk_{\lambda}}^{I}. \]
As a consequence,
\[
\mathbb{P}(X_{1,2}^{I}(i+1) = 1|Z_{1}^{l} = k, A) = \sum_{k_{2},\ldots,i+1} \prod_{\lambda = 2}^{i+1} \alpha_{k_{\lambda}}^{l} \gamma_{k_{\lambda}} \pi_{\lambda}^{O} \\
= \prod_{\lambda = 2}^{i+1} \sum_{k_{\lambda}} \alpha_{k_{\lambda}}^{l} \gamma_{k_{\lambda}} \pi_{\lambda}^{O} = \sigma_{k}^{l}
\]

- Then we define \((u_{i}^{l})_{i=0, \ldots, 2Q_{I} - 1}\), such that \(u_{0}^{l} = 1\) and \(\forall 1 \leq i \leq 2Q_{I} - 1:\)
  \[
u_{i}^{l} = \mathbb{P}(X_{1,2}^{I}(i+1) = 1|A) = \sum_{k, l} \mathbb{P}(X_{1,2}^{I}(i+1) = 1|Z_{1}^{l} = k) \mathbb{P}(Z_{1}^{l} = k, A) \mathbb{P}(Z_{1}^{O} = l) 
  = \sum_{k} \sigma_{k}^{l} \sum_{\lambda} \gamma_{kl} \pi_{\lambda}^{O} = \pi_{k}^{l}
  = \sum_{k} \sigma_{k}^{l} \pi_{k}^{l}.
\]

Note that the \((u_{i}^{l})\)'s can be defined because \(n_{I} \geq 2Q_{I}\) (assumption A2).

- To conclude we use the same arguments as the ones used for the identifiability of \(O\), i.e. we define \(M^{l}\) a \((Q_{I} + 1) \times Q_{I}\) matrix such that \(M_{ij}^{l} = u_{i+j-2}^{l}\) together with the matrices \(M_{-k}^{l}\) and the polynomial function \(B^{l}\) (see equation (B.15)). Let \(R^{l}\) be a \(Q_{I} \times Q_{I}\) matrix such that \(R_{ik}^{l} = \sigma_{k}^{l-1}\). \(R^{l}\) is an invertible Vandermonde matrix because of assumption A1 on \(\alpha^{l} \cdot \gamma \cdot \pi^{O}\). As before, \(R^{l}\) can be identified in unique way from \(B^{l}\). Then, noting that \(M_{-Q_{I}+1}^{l} = R^{l} D_{\pi^{l}} R^{l}\) where \(D_{\pi^{l}} = \text{diag}(\pi^{l}) = \text{diag}(\gamma \cdot \pi^{O})\), we obtain: \(D_{\pi^{l}} = (R^{l})^{-1} M_{-Q_{I}}^{l} (R^{l})^{-1}\), which is uniquely defined by \(\mathbb{P}_{X}\). Now, let us introduce
\[
U_{ij}^{l} = \mathbb{P}(X_{1,2}^{I}(i+1) = 1, X_{3,2}^{I}(n_{j} = j+2) = 1)
\]
with \(U_{i1}^{l} = \mathbb{P}(X_{1,2}^{I}(i+1) = 1)\). Then we have \(U^{l} = R^{l} D_{\pi^{l}} \alpha^{l} D_{\pi^{l}} R^{l}\) and so \(\alpha^{l} = D_{\pi^{l}}^{-1}(R^{l})^{-1} U^{l}(R^{l})^{-1} D_{\pi^{l}}^{-1}\). As a consequence, \(\alpha^{l}\) is uniquely identified from \(\mathbb{P}_{X}\).

**Identifiability of \(\gamma\).** For any \(2 \leq i \leq Q_{I}\) and \(2 \leq j \leq Q_{O}\), let \(U_{i,j}^{IO}\) be the probability that \(X_{1,2}^{I}(i+1) = 1\) and \(X_{1,2}^{O}(i+j-1) = 1\). Note that the \(U_{i,j}^{IO}\) can be defined because \(n_{O} \geq Q_{I} + Q_{O} - 1\) and \(n_{I} \geq Q_{I}\) (assumption A2).

- Then, for all \(2 \leq i \leq Q_{I}\) and \(2 \leq j \leq Q_{O}\),
\[
U_{i,j}^{IO} = \mathbb{P}(X_{1,2}^{I}(i+1) = 1, X_{1,2}^{O}(i+j-1) = 1|A) = \sum_{k, l} \mathbb{P}(X_{1,2}^{I}(i+1) = 1|A, Z_{1}^{l} = k, Z_{1}^{O} = l) \times \mathbb{P}(Z_{1}^{l} = k, Z_{1}^{O} = l, A).
\]
• We first prove that:

\[
\mathbb{P}(X_{1,2;i}^l = 1, X_{1,i+1:j-1}^O = 1 | A, Z_1^I = k, Z_1^O = l) = \sigma_{k}^{j-1} \eta_{l}^{j-1}.
\] (B.20)

Indeed,

\[
\mathbb{P}(X_{1,2;i}^l = 1, X_{1,(i+1):(i+j-1)}^O = 1 | A, Z_1^I = k, Z_1^O = l) = \\
= \sum_{k_2;i,l_2,n_0} \mathbb{P}(X_{1,2;i}^l = 1, X_{1,(i+1):(i+j-1)}^O = 1 | Z_1^I = (k, k_2;i), Z_1^O = (l, l_2,n_0), A) \\
\times \mathbb{P}(Z_{2;i} = k_2;i, Z_{2,n_0}^O = l_2,n_0 | Z_1^I = k, Z_1^O = l, A) \\
= \sum_{k_2;i,l_2,n_0} \mathbb{P}(X_{1,2;i}^l = 1 | Z_{1,i}^I = (k, k_2;i)) \\
\times \mathbb{P}(X_{1,(i+1):(i+j-1)}^O = 1 | Z_1^O = l, Z_1^O = (l,(i+1):(i+j-1)) \\
\times \mathbb{P}(Z_{2;i} = k_2;i, Z_{2,n_0}^O = l_2,n_0 | Z_1^I = k, Z_1^O = l, A).
\] (B.21)

Moreover, let us have a look at \(\mathbb{P}(Z_{2;i}^l = k_2;i, Z_1^O = l_2,n_0 | Z_1^I = k, Z_1^O = l, A)\):

\[
\mathbb{P}(Z_{2;i} = k_2;i, Z_{2,n_0}^O = l_2,n_0 | Z_1^I = k, Z_1^O = l, A) = \\
\mathbb{P}(Z_{2;i}^l = k_2;i, Z_{2,n_0}^O = l_2,n_0, Z_1^I = k, Z_1^O = l, A) \times \mathbb{P}(Z_{2,n_0}^O = l_2,n_0 | Z_1^I = k, Z_1^O = l, A).
\]

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

- \(\mathbb{P}(Z_{2;i}^l = k_2;i, Z_{2,n_0}^O = l_2,n_0, Z_1^I = k, Z_1^O = l, A) = \mathbb{P}(Z_{2;i}^l = k_2;i, Z_{2,i}^O = l_2;i),\)
- \(\mathbb{P}(Z_{2,n_0}^O = l_2,n_0 | Z_1^I = k, Z_1^O = l, A) = \mathbb{P}(Z_{2,n_0}^O = l_2,n_0).\)

As a consequence,

\[
\begin{align*}
\mathbb{P}(Z_{2;i}^l = k_2;i, Z_{2,n_0}^O = l_2,n_0 | Z_1^I = k, Z_1^O = l, A) = \\
\mathbb{P}(Z_{2;i}^l = k_2;i, Z_{2,i}^O = l_2;i) \mathbb{P}(Z_{2,i}^O = l_2;i) \mathbb{P}(Z_{(i+1):(i+j-1)}^O = l_{(i+1):(i+j-1)}) \\
\times \mathbb{P}(Z_{(i+j):n_0}^O = l_{(i+j):n_0}).
\end{align*}
\]
Going back to equation (B.21) and decomposing the summation we obtain:

\[
\mathbb{P}(X_{1,2i}^I = X_{1,(i+1):(i+j-1)}^O = 1 | A, Z_1^I = k, Z_1^O = l) = \sum_{k_{2:i}, l_{2:i} : nO} \mathbb{P}(X_{1,2i}^I = 1 | Z_{1,i}^I = (k, k_{2:i}))
\]

\[
\quad \times \mathbb{P}(X_{1,(i+1):(i+j-1)}^O = 1 | Z_1^O = l, Z_{(i+1):(i+j-1)}^O = l_{(i+1):(i+j-1)})
\]

\[
\quad \times \mathbb{P}(Z_{2:2i}^2 = k_{2:i}, Z_{2:1}^O = l_{2:i}) \mathbb{P}(Z_{2:1}^O = l_{(i+1):(i+j-1)})
\]

\[
= \sum_{k_{2:i}} \mathbb{P}(X_{1,2i}^I = 1 | Z_1^I = k, Z_{2:1}^I = k_{2:i}) \mathbb{P}(Z_{2:2i}^2 = k_{2:i} | A) \times \mathbb{P}(X_{1,(i+1):(i+j-1)}^O = 1 | Z_1^O = l) \]

\[
= \sum_{k_{2:i}} \mathbb{P}(X_{1,2i}^I = 1 | Z_1^I = k, Z_{2:1}^I = k_{2:i}) \mathbb{P}(Z_{2:2i}^2 = k_{2:i} | Z_1^I = k, A) \]

\[
\times \mathbb{P}(X_{1,(i+1):(i+j-1)}^O = 1 | Z_1^O = l) \]

\[
= \mathbb{P}(X_{1,2i}^I = 1 | Z_1^I = k, A) \mathbb{P}(X_{1,(i+1):(i+j-1)}^O = 1 | Z_1^O = l).
\]

Finally, we have:

\[
\mathbb{P}(X_{1,2i}^I = 1 | Z_1^I = k, A) = \sigma_i^{-1}, \quad \text{from equation (B.17)}
\]

\[
\mathbb{P}(X_{1,(i+1):(i+j-1)}^O = 1 | Z_1^O = l) = \tau_l^{-1},
\]

and so, we have proved equality (B.20).

- Now, \( A_{11} = 1 \) implies \( \mathbb{P}(Z_1^I = k, Z_1^O = l | A) = \gamma_{kl} \pi_l^I \) and combining this result with equations (B.20) and (B.18) leads to: \( U^{IO}_{1,j} = \sum_k \sigma_k \gamma_{kl} \pi_l^O \tau_l^{-1} \). Setting

\[
U^{IO}_{1,j} = \mathbb{P}(X_{1,j+1}^O, \ldots, X_{1,i+1}^O = 1 | A) = \sum_k \gamma_{kl} \pi_l^O \tau_l^{-1}, \quad \text{for } j > 1
\]

\[
U^{IO}_{i,1} = \mathbb{P}(X_{1,2}^I, \ldots, X_{i,i}^I = 1 | A) = \sum_k \gamma_{kl} \pi_l^I, \quad \text{for } i > 1
\]

\[
U^{IO}_{i,i} = 1
\]

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 \( \gamma = (R^I)^{-1} U^{IO} (R^{O'})^{-1} D^{-1}_{\pi^O} \) and \( \gamma \) is identified.
C Details of the Variational EM

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

\[
\mathcal{I}_\theta(\mathcal{R}(Z^I, Z^O|A)) = \sum_{j,l} \tau^O_{jl} \log \pi^O_l + \sum_{i,k} \sum_{j,l} A_{ij} \tau^O_{jl} \log \gamma_{kl} + \frac{1}{2} \sum_{i,k} \sum_{j,l} A_{ij} \tau^O_{jl} \log \phi(X^I_{ii}, \alpha^I_{kk}) + \frac{1}{2} \sum_{j,l} \tau^O_{jl} \log \gamma_{kl} \tag{1}
\]

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

**VE step** compute

\[\{\tau^I, \tau^O\}^{(t+1)} = \arg \max_{\tau^I, \tau^O} \mathcal{I}_{\theta^{(t)}}(\mathcal{R}(Z^I, Z^O|A)) = \arg \min_{\tau^I, \tau^O} KL(\mathcal{R}(Z^I, Z^O|A) \| P_{\theta^{(t)}}(Z^I, Z^O|X^I, X^O, A)) .\]

**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:

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

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

\[
\begin{align*}
\hat{\tau}^O_{jl} &\propto n^O_l \prod_{i,k} A_{ik} \prod_{j,l} \phi(X^O_{jl}, \alpha^O_{kl}). \\
\hat{\tau}^I_{jl} &\propto \prod_{j,l} A_{kl} \prod_{i,k'} \phi(X^I_{ji}, \alpha^I_{kk'}).
\end{align*}
\]

which are used in the VE step to update the \(\tau^I_i\)'s and \(\tau^O_j\)'s.

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

$$\hat{\tau}_{jl}^O \propto \pi_l^O \prod_{j' \neq j} \phi(X_{jl}^O, \alpha_{kl}^O) \hat{\tau}_{j'l'}^O \quad \text{and} \quad \hat{\tau}_{jl}^I \propto \pi_k^I \prod_{l' \neq l} \phi(X_{jl}^I, \alpha_{kk'}^I) \hat{\tau}_{k'l'}^I,$$

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:

$$\hat{\pi}_l^O = \frac{1}{n_O} \sum_j \hat{\tau}_{jl}^O \quad \text{and} \quad \hat{\alpha}_{kl}^O = \frac{\sum_{j' \neq j} \hat{\tau}_{jl}^O X_{jl}^O \hat{\tau}_{j'l'}^O}{\sum_{j' \neq j} \hat{\tau}_{j'l'}^O},$$

$$\hat{\gamma}_{kl} = \frac{\sum_{i,j} \hat{\tau}_{ik}^O A_{ij} \hat{\tau}_{jl}^O}{\sum_{i,j} A_{ij} \hat{\tau}_{jl}^O},$$

for which the gradient

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

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

Model parameters have natural interpretations. $\pi_l^O$ is the mean of the posterior probabilities for the organisations to belong to cluster $l$. $\alpha_{kk'}^O$ (resp. $\alpha_{ll'}^I$) is ratio of existing links over possible links between blocks $k$ and $k'$ (resp. $l$ and $l'$). $\gamma_{kl}$ 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 $\gamma$ is such that the levels are independent, then any column of $\gamma$ represents the proportion of individuals in the different blocks:

$$\pi_k^I = \gamma_{k1} = \frac{1}{n_I} \sum_i \hat{\tau}_{ik}. $$

### 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 $\hat{Z}^O$ and $\hat{Z}^I$ the imputed 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:

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

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 \( \theta \):

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

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:

\[
\log \ell_\theta(X, Z|A, Q_I, Q_O) = \log \left( \int \ell_\theta(X^I, X^O, Z^I, Z^O|\theta, A, Q_I, Q_O)p(\theta|Q_I, Q_O)d\theta \right)
\]

\[
= \log \ell_{\alpha^I}(X^I|Z^I, Q_I) \\
+ \log \ell_{\gamma}(Z^I|A, Z^O, Q_I, Q_O) \\
+ \log \ell_{\alpha^O}(X^O, Z^O|Q_O).
\]  

(D.22)

The quantity defined in (D.24) evaluated at \( Z^O := \widehat{Z}^O \) is approximated as in Daudin et al. (2008) by

\[
\log \ell_{\alpha^O}(X^O, \widehat{Z}^O, Q_O) \approx n_O \rightarrow \log \ell_{\alpha^O_{\hat{\pi}O}}(X^O, \widehat{Z}^O|Q_O) - \text{pen}(\pi^O, \alpha^O, Q_O) \\
\text{pen}(\pi^O, \alpha^O, Q_O) = \frac{Q_{O-1}}{2} \log n_O + \frac{1}{2} \frac{Q_I(Q_I+1)}{2} \log \frac{n_I(n_I-1)}{2}
\]

(D.25)

This approximation results from a BIC-type approximation of \( \log \ell_{\alpha^O}(X^O|\widehat{Z}^O, Q_O) \) and a Stirling approximation of \( \log \ell_{\alpha^O}(\widehat{Z}^O, Q_O) \).

The same BIC-type approximation on \( \log \ell_{\alpha^I}(X^I|\widehat{Z}^I, Q_I) \) (Equation (D.22)) leads to:

\[
\log \ell_{\alpha^I}(X^I|\widehat{Z}^I, Q_I) = n_I \rightarrow \log \ell_{\alpha^I_{\hat{\gamma}I}}(X^I|\widehat{Z}^I, Q_I) + \text{pen}(\alpha^I, Q_I) \\
\text{with pen}(\alpha^I, Q_I) = \frac{1}{2} \frac{Q_I(Q_I+1)}{2} \log \frac{n_I(n_I-1)}{2}
\]

(D.26)

For quantity (D.23) depending on \( \gamma \) and \( Z^I \) given \( (Q_I, Q_O) \), we have to adapt the calculus. Let us set independent Dirichlet prior distributions of order \( Q_I \mathcal{D}(1/2, \cdots, 1/2) \) on the
columns $\gamma_l$. We are able to derive an exact expression of $\log \ell_\gamma(Z^I|A, Z^O, Q_I, Q_O)$:

$$
\ell_\gamma(Z^I|A, Z^O, Q_I, Q_O) = \int \ell(Z^I|A, Z^O, \gamma, Q_I, Q_O)p(\gamma, Q_I, Q_O)d\gamma
$$

$$
= \prod_i \int \prod_{j,k,l} A_{ij}Z_{ik}^l Z_{jl}^O p(\gamma_{kl})d\gamma_{kl}
$$

$$
= \prod_i \int \prod_k \gamma_{kl}^{N_{kl}}p(\gamma_{kl})d\gamma_{kl}, \quad \text{where } N_{kl} = \sum_{ij} A_{ij}Z_{ik}^l Z_{jl}^O
$$

$$
= \prod_i \int \prod_k \gamma_{kl}^{N_{kl} + a - 1} \frac{\Gamma(1/2 \cdot Q_I)}{\Gamma(1/2)^{Q_I}} d\gamma_{kl}
$$

$$
= \frac{\Gamma(1/2Q_I)^{Q_O}}{\Gamma(1/2)^{Q_O+Q_I}} \prod_i \prod_k \Gamma(N_{kl} + 1/2)
$$

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

$$
\log \ell_\gamma(Z^I|A, Z^O, Q_I, Q_O) \approx (n_O, n_I) \rightarrow \infty \sum_{kl} (N_{kl} \log N_{kl} + N_{kl}) - \sum_i \left(\frac{Q_I - 1}{2} + \sum_k N_{kl}\right) \log (\sum_k N_{kl}) - \sum_{k,l} N_{kl}.
$$

The quantity (D.27) evaluated at $(Z^I, Z^O) := (\hat{Z}^I, \hat{Z}^O)$ can be reformulated in the following way:

$$
\log \ell_\gamma(\hat{Z}^I|A, \hat{Z}^O, Q_I, Q_O) \approx (n_O, n_I) \rightarrow \infty \log \ell_\gamma(\hat{Z}^I|A, \hat{Z}^O, Q_I, Q_O) - \frac{Q_I - 1}{2} \sum_i \log \sum_{j,l} A_{ij} \hat{Z}_{jl}^O
$$

with $\hat{\gamma}_{kl} = \frac{\sum_{ij} Z_{ik}^l A_{ij} \hat{Z}_{jl}^O}{\sum_{ij} A_{ij} \hat{Z}_{jl}^O}$

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

$$
\log \ell_\gamma(\hat{Z}^I|A, \hat{Z}^O, Q_I, Q_O) \approx (n_O, n_I) \rightarrow \infty \log \ell_\gamma(\hat{Z}^I|A, \hat{Z}^O, Q_I, Q_O) - \frac{Q_I - 1}{2} Q_O \log n_I.
$$

Combining Equations (D.25), (D.26) and (D.28) we obtain the given expression.