arXiv:1910.10512v2 [stat.ME] 11 Mar 2020

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

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 organizations. The levels are linked by an affiliation relationship, each individual belonging to a unique organization. We design a Stochastic block model (SBM) suited to multilevel networks. The SBM is a latent variable model for networks, where the connections between nodes depend on a latent clustering (blocks) introducing 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 individual and organizational levels 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 programs trade fair. The inter-organizational level is the economic network between companies and the inter-individual level is the informal network between their representatives.

Keywords: Latent variable model; Hierarchical modeling; Social network; Variational inference

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 the multilevel networks. Multilevel networks arise in sociology of organizations and collective action when willing to study jointly the social network of individuals and the interaction network of organizations the individuals belong to. Indeed, the individuals not only interact with each others but are also members of interacting organizations. Following Lazega and Snijders

*UMR MIA-Paris, AgroParisTech, INRAe, Université Paris-Saclay, 75005, Paris, France
†Institut d'Études Politiques de Paris, France

(2015), one might think that these two types of interactions (between individuals and between organizations) are interdependent, the individuals shaping their organizations and the organizations having an influence on the individuals. We aim to propose a statistical model for multilevel networks in order to understand how the two levels are intertwined and how one level impacts the other.

In what follows, a multilevel network is defined as the collection of an inter-individual network, an inter-organizational network and the affiliation of the individuals to the organizations. Besides, we assume that the individuals belong to a unique organization. Such a dataset is studied by Lazega et al. (2008), some researchers in cancerology being the individuals and their laboratories the organizations. Brailly et al. (2016) deals with another dataset concerned with the economic network of audiovisual firms and the informal network of their sales representatives during a trade fair. This latter dataset will be analyzed in this paper.

In the last years, the Stochastic Block Model (SBM developed by 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 (clusters) and that the members of a same block share a similar profile of connectivity. Compared to other graph clustering methods such as modularity maximization, hierarchical clustering or spectral clustering (see Kolaczyk, 2009, and references therein), the SBM is a generative model and can fit to a wide range of topologies since it gathers into blocks the nodes that are structurally equivalent. This includes but is not restricted to the detection of assortative communities where the probability of connection within a block is higher than the probability of connection between the blocks. The 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 (MLVSBM).

Our contribution In a few words, we model the heterogeneity in the inter-individual and inter-organizational connections by introducing blocks of individuals and blocks of organizations, the blocks containing homogeneous groups of actors (individuals or organizations) with respect to their connectivity. The two levels are assumed to be interdependent through their latent blocks. More specifically, the latent blocks of the inter-individual level depend on the latent blocks of the inter-organizational level and the affiliation. This bi-clustering approach allows us to determine how group of organizations influence the connectivity patterns of their individuals.

Due to the latent variables, the estimation of the parameters is a complex task. We resort to a variational version of the Expectation-Maximization (EM) algorithm. For the 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; Côme and Latouche, 2015). 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. We propose an adapted version of the Integrated Complete Likelihood (ICL) criterion. First developed by Biernacki et al. (2000) for mixture models as an alternative to the Bayesian Information Criterion (BIC), it was then adapted by Daudin et al. (2008) to the SBM . The ICL has since illustrated its efficiency and relevance for various SBMs and their extensions such as multiplex network (Barbillon et al., 2017), dynamic SBM (Matias and Miele, 2017; Bartolucci et al., 2018) or degree corrected SBM (Yan, 2016). Besides, a critical issue in sociology is to verify the multilevel hypothesis of a given dataset. We propose a criterion to decide whether the two levels (inter-individual and inter-organizational) 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 model 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 organizations 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-organizational network into an inter-individual network thus adopting a multiplex network approach (the individuals interconnect directly or through the organizations 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 optimized 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 tolopogy 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 organizations are not too different. Indeed it requires to duplicate the data of the inter-organizational 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 organizations. 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 modeled 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 organizations are considered on the same level as the affiliations, and the clustering might be strongly influenced by the number of individuals in each organization.

Finally, our work is also different from the SBM with edges covariates (Mariadassou et al., 2010) with the individuals as nodes and the inter-organizational 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 organizations.

Outline of the paper The paper is organized as follows. The SBM 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 the variational EM and the ICL criterion are postponed to the Appendix sections. In Section 4, we present an extensive simulation study illustrating the relevance of our inference method, model selection criterion and procedure. Section 5 is dedicated to the analysis of a sociological dataset by our MLVSBM. 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 organizations. 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 : $\forall (i,i') \in \{1,\ldots,n_I\}^2$:

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

X^O is the binary $n_O \times n_O$ matrix representing the inter-organizational network, $\forall (j,j') \in \{1,\ldots,n_O\}^2$:

$$X^O_{jj'} = \begin{cases} 1 & \text{if there is an interaction from organization } j \text{ to organization } j', \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Remark. *In general, no self-loop are considered in the network, thus the interactions are defined for $i \neq i'$ and $j \neq j'$. Moreover, if the interactions are undirected then*

$$X^I_{ii'} = X^I_{i'i} \quad \forall (i,i') \qquad \textit{or/and} \qquad X^O_{jj'} = X^O_{j'j} \quad \forall (j,j').$$

In what follows, we present the methodology for undirected networks. However, all the results can be adapted to directed networks without any difficulty.

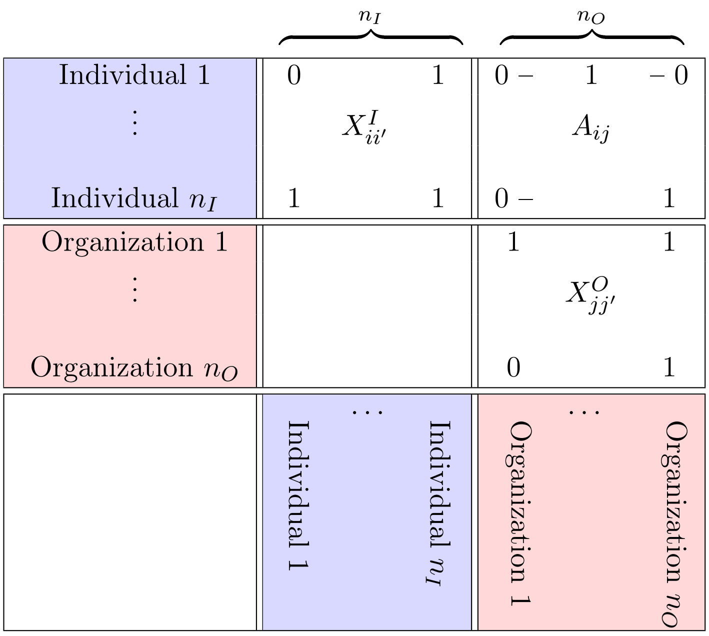

Figure 1: Matrix representation of a multilevel network

Let A be the affiliation matrix. A is a $n_I \times n_O$ matrix such that:

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

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 organization. A synthetic view of a generic dataset is provided in Figure 1.

We propose a joint modeling of the inter-individual and inter-organizational networks based on an extension of the SBM. More precisely, assume that the n_O organizations 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 organization j belongs to block l ($l \in \{1, \ldots, Q_O\}$) and $Z^I_i = k$ if individual i belongs to block k ($k \in \{1, \ldots, Q_I\}$).

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

$$\begin{array}{rcl} \mathbb{P}(X^O_{jj'} = 1 | Z^O_j, Z^O_{j'}) & = & \alpha^O_{Z^O_j Z^O_{j'}} \\ \mathbb{P}(X^I_{ii'} = 1 | Z^I_i, Z^I_{i'}) & = & \alpha^I_{Z^I_i Z^I_{i'}}. \end{array} \tag{3}$$

As a consequence, the blocks gather nodes (blocks of individuals on the one hand and blocks of organizations on the other hand) sharing the same profiles of connectivity.
In order to take into account the fact that organizations may shape the individual behaviors, we assume that the memberships of the individuals (Z^I) depend on the blocks

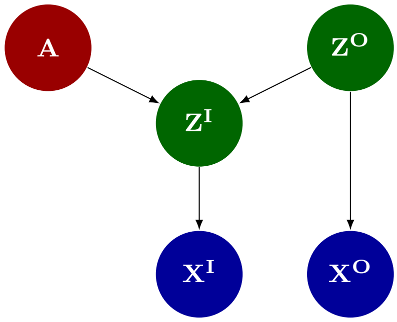

Figure 2: DAG of the stochastic block model for multilevel network (MLVSBM)

of the organizations (Z^O) they are affiliated to. More precisely, we set:

$$\mathbb{P}(Z_i^I = k | Z_j^O, A_{ij} = 1) = \gamma_{k Z_j^O} \quad \forall i \in \{1, \dots, n_I\} \quad \forall k \in \{1, \dots, Q_I\}, \tag{4}$$

where γ is a $Q_I \times Q_O$ matrix such that $\sum_{k=1}^{Q_I} \gamma_{kl} = 1 \ \forall l \in \{1, \dots, Q_O\}$. The (Z_j^O) are assumed to be independent random variables distributed as

$$\mathbb{P}(Z_j^O = l) = \pi_l^O, \qquad \forall j \in \{1, \dots, n_O\} \quad \forall l \in \{1, \dots, Q_O\}, \tag{5}$$

with $\sum_{l=1}^{Q_O} \pi_l^O = 1$.
Equations (4) and (5) state that the clustering of an individual is not completely driven by his/her behavior but is also shaped by the clustering of the organization he/she belongs to. In particular, if $Q_O = Q_I$ and γ 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 organizations. At the opposite, if all the columns of γ are equal, then the clustering of the individuals is independent on the clustering of the organizations. This point will be developed hereafter.

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

Likelihood From Equations (3), (4) and (5), we derive the complete log-likelihood for a directed MLVSBM:

$$\begin{aligned}\log \ell_\theta \left(X^I, X^O, \mathbf{Z} | A\right) &= \log \ell_{\pi^O}(Z^O) + \log \ell_\gamma(Z^I | Z^O, A) + \log \ell_{\alpha^I}(X^I | Z^I) + \log \ell_{\alpha^O}(X^O | Z^O) \\ &= \sum_{j,l} \mathbb{1}_{Z_j^O = l} \log \pi_l^O + \sum_{i,k} \mathbb{1}_{Z_i^I = k} \sum_{j,l} A_{ij} \mathbb{1}_{Z_j^O = l} \log \gamma_{kl} \qquad (6) \\ &+ \frac{1}{2} \sum_{i' \neq i} \sum_{k,k'} \mathbb{1}_{Z_i^I = k} \mathbb{1}_{Z_{i'}^I = k'} \log \phi(X_{ii'}^I, \alpha_{kk'}^I) + \frac{1}{2} \sum_{j' \neq j} \sum_{l,l'} \mathbb{1}_{Z_j^O = l} \mathbb{1}_{Z_{j'}^O = l'} \log \phi(X_{jj'}^O, \alpha_{ll'}^O),\end{aligned}$$

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 $\ell_\theta(\mathbf{X}|A)$ is obtained by integrating out the latent variables $\mathbf{Z}$ in Equation 6. As soon as n_O, n_I, Q_O, 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 variational EM algorithm to maximize this likelihood (see Section 3).

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

Proposition 1. *In the MLVSBM, the two following properties are equivalent:*

1. *Z^I is independent on Z^O,*
2. *$\gamma_{kl} = \gamma_{kl'} \quad \forall l, l' \in \{1, \ldots, Q_O\}$*

and imply that:

3. *X^I and X^O are independent.*

This proposition is proved in A. Proposition 1 can be interpreted as follows: in the case where the clustering of the individuals does not depend on the clustering of the organizations, all column vectors of γ are identical. Hence, under this restriction on γ, the model for multilevel network can be rewritten as the product of two independent SBMs, one for each level. Conversely, in the case of a strong dependence between the levels, each column of γ will have one coefficient close to one, the others being close to 0. Therefore, the individuals affiliated to organizations belonging to the same block of organizations will be affiliated to one block of individuals. Even if the γ'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 constraint on the corresponding connection parameters α^O and α^I.

Identifiability The identifiability conditions for the MLVSBM are given in the following proposition.

Proposition 2. *The MLVSBM 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.

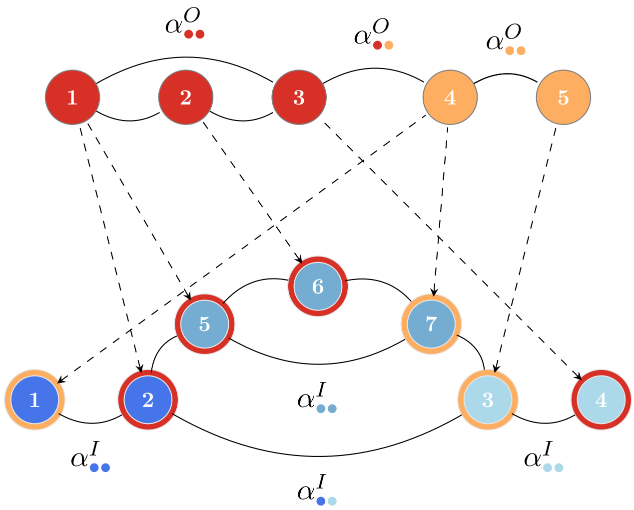

Figure 3: MLVSBM with inter-organizational level on the top and inter-individual level on the bottom. The various shades of blue depict the clustering of the individuals and the various shades of red depict the clustering of the organizations. The parameters α over the plain links between nodes are the probabilities of connections given the nodes colors (clustering/blocks). The outer circles around the nodes of the individuals represent the blocks of the organizations they are affiliated to. The dashed links stand for the affiliations.

The set of parameters that does not verify assumption $\mathcal{A}1$ has null Lebesgue measure. Assumption $\mathcal{A}2$ is very weak in practice. Assumption $\mathcal{A}3$, on the affiliation, means that at least some organizations must not be empty and enough individuals belong to different organizations. The proof of this proposition is provided in B and results from an extension of the proof given in Celisse et al. (2012).

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, $\ell_\theta(\mathbf{X}|A)$ is obtained by integrating out the latent variables $\mathbf{Z}$ in the complete data likelihood (6). However, this calculus becomes not computationally tractable as the numbers of nodes and blocks increase.

The Expectation-Maximization algorithm (EM) (Dempster et al., 1977) is a popular solution to maximize the likelihood of models with latent variables. However it requires the computation of $\mathbb{P}_\theta(\mathbf{Z}|\mathbf{X}, A)$ 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 the SBM by Daudin et al. (2008).

In a few words, the variational EM algorithm maximizes the so-called variational bound i.e. a lower bound of the log-likelihood denoted $\mathcal{I}_\theta(\mathcal{R}(\mathbf{Z}|\mathbf{X}))$ and defined as follows:

$$\begin{aligned} \mathcal{I}_\theta(\mathcal{R}(\mathbf{Z}|A)) &:= \mathbb{E}_\mathcal{R}\left[\ell_\theta\left(\mathbf{Z}, \mathbf{X}|A\right)\right] + \mathcal{H}\left(\mathcal{R}(\mathbf{Z}|A)\right) \\ &= \ell_\theta\left(\mathbf{X}|A\right) - \mathrm{KL}\left(\mathcal{R}(\mathbf{Z}|A)\|\mathbb{P}_\theta(\mathbf{Z}|\mathbf{X}, A)\right) \leq \ell_\theta(\mathbf{X}|A), \end{aligned} \tag{7}$$

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

$$\mathcal{R}(\mathbf{Z}|A) = \prod_{i=1}^{n_I}\prod_{k=1}^{Q_I}(\tau_i^I)^{\mathbb{1}_{Z_i^I = k}}\prod_{j=1}^{n_O}\prod_{l=1}^{Q_O}(\tau_j^O)^{\mathbb{1}_{Z_j^O = l}}, \tag{8}$$

where $\tau_{ik}^I = \mathbb{P}_\mathcal{R}(Z_i^I = k)$ and $\tau_{jl}^O = \mathbb{P}_\mathcal{R}(Z_j^O = l)$.
Inputting Equations (6) and (8) into Equation (7), the variational bound for the MLVSBM can be written as follows:

$$\begin{aligned} \mathcal{I}_\theta(\mathcal{R}(\mathbf{Z}|A)) &= \sum_{j,l}\tau_{jl}^O \log \pi_l^O + \sum_{i,k}\tau_{ik}^I \sum_{j,l} A_{ij}\tau_{jl}^O \log \gamma_{kl} \\ &+ \frac{1}{2}\sum_{i'\neq i}\sum_{k,k'}\tau_{ik}^I\tau_{i'k'}^I \log\phi\left(X_{ii'}^I, \alpha_{kk'}^I\right) + \frac{1}{2}\sum_{j'\neq j}\sum_{l,l'}\tau_{jl}^O\tau_{j'l'}^O \log\phi\left(X_{jj'}^O \alpha_{ll'}^O\right) \\ &- \sum_{i,k}\tau_{ik}^I \log \tau_{ik}^I - \sum_{j,l}\tau_{jl}^O \log \tau_{jl}^O. \end{aligned}$$

The variational EM algorithm consists in iterating two steps. Step VE maximizes the variational bound with respect to the parameters of the approximate distribution defined in Equation (8). This is equivalent to minimizing the Kullbach-Leibler divergence term. Step M maximizes the variational bound with respect to the model parameters θ. The procedure is given in Algorithm 1 and details of the calculus and algorithm are developed in C. Algorithm 1 can be slightly modified to handle missing data (dyads which are not observed on either levels) by summing up on observed dyads only. An interesting feature of the MLVSBM is to make use of one level to help the prediction of missing dyads of the other level.

Algorithm 1: Variational EM algorithm

Data: $\{X^I, X^O, Z^I, Z^O, A\}$, a multilevel network with an initial clustering of size (Q_I, Q_O).

Procedure:

• Set $\{\tau^I, \tau^O\}$ from the initial clustering.

while $\mathcal{I}_\theta(\mathcal{R}(\mathbf{Z}|A))$ *is increasing* **do**

- **M step** compute

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

by updating the model parameters as follows:

$$\widehat{\pi_l^O} = \frac{1}{n_O}\sum_j \widehat{\tau_{jl}^O} \qquad \widehat{\alpha_{ll'}^O} = \frac{\sum_{j'\neq j} \widehat{\tau_{jl}^O} X_{jj'}^O \widehat{\tau_{j'l'}^O}}{\sum_{j'\neq j} \widehat{\tau_{jl}^I}\widehat{\tau_{j'l'}^I}}$$

$$\widehat{\gamma}_{kl} = \frac{\sum_{i,j} \widehat{\tau_{ik}^I} A_{ij} \widehat{\tau_{jl}^O}}{\sum_{i,j} A_{ij}\widehat{\tau_{jl}^O}} \qquad \widehat{\alpha_{kk'}^I} = \frac{\sum_{i'\neq i} \widehat{\tau_{ik}^I} X_{ii'}^I \widehat{\tau_{i'k'}^I}}{\sum_{i'\neq i} \widehat{\tau_{ik}^I}\widehat{\tau_{i'k'}^I}}.$$

- **VE step** compute

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

by updating the variational parameters with the following fixed points relationships:

$$\widehat{\tau_{jl}^O} \propto \pi_l^O \prod_{i,k} \gamma_{kl}^{A_{il}\widehat{\tau_{ik}^I}} \prod_{j'\neq j}\prod_{l'} \phi(X_{jj'}^O, \alpha_{ll'}^O)^{\widehat{\tau_{j'l'}^O}}$$

$$\widehat{\tau_{ik}^I} \propto \prod_{j,l} \gamma_{kl}^{A_{il}\widehat{\tau_{jl}^O}} \prod_{i'\neq i}\prod_{k'} \phi(X_{ii'}^I, \alpha_{kk'}^I)^{\widehat{\tau_{i'k'}^I}}.$$

return $\mathcal{I}_\theta(\mathcal{R}(\mathbf{Z}|A))$, $\widehat{\theta}$ *and* $\{\widehat{\tau^I}, \widehat{\tau^O}\}$

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 values both goodness of fit and classification sharpness (Mariadassou et al., 2010).

Our criterion is equal to:

$$\mathrm{ICL}_{\mathrm{MLVSBM}}(Q_I, Q_O) = \log \ell_{\hat{\theta}}(X^I, X^O, \widehat{Z^I}, \widehat{Z^O} | A, Q_I, Q_O) - \mathrm{pen}_{\mathrm{MLVSBM}}(Q_I, Q_O), \quad (9)$$

where

$$\begin{aligned}\mathrm{pen}_{\mathrm{MLVSBM}}(Q_I, Q_O) =& \frac{1}{2}\frac{Q_I(Q_I+1)}{2} \log \frac{n_I(n_I-1)}{2} + \frac{Q_O(Q_I-1)}{2} \log n_I + \\ & \frac{1}{2}\frac{Q_O(Q_O+1)}{2} \log \frac{n_O(n_O-1)}{2} + \frac{Q_O-1}{2} \log n_O \end{aligned} \quad (10)$$

where $\widehat{Z^O}$ and $\widehat{Z^I}$ are the imputed latent variables using the maximum a posteriori (MAP) of $\mathbb{P}_{\hat{\theta}}(\mathbf{Z}|\mathbf{X}, A; Q_I, Q_O)$. The calculus is provided in D. As for the variational inference, $\mathbb{P}_{\hat{\theta}}(\mathbf{Z}|\mathbf{X}, A; Q_I, Q_O)$ is unknown and, in practice, we replace it by its mean-field approximation $\mathcal{R}_{\hat{\theta}}(\mathbf{Z}; Q_I, Q_O)$.

Remark. *Once again, note that the penalty* (10) *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*

$$\mathrm{pen}_{\mathrm{SBM}}(Q) = \frac{1}{2}\frac{Q(Q+1)}{2} \log \frac{n(n-1)}{2} + \frac{Q-1}{2} \log n. \quad (11)$$

The penalty term in Equation (10) for the inter-organizational level is the same as the one given in Equation (11). For the inter-individual network, the factor in front of $\log n_I$ is $Q_O(Q_I-1)$ instead of Q_I-1 for the SBM as in Equation (11), that is the penalty term which corresponds to the degree of freedom of γ.

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 γ is forced to have all its columns identical, then the penalty term on γ becomes $\frac{1}{2}(Q_I-1)\log n_I$ and, as a consequence:

$$\mathrm{ICL}_{\mathrm{Ind}}(Q_I, Q_O) = \mathrm{ICL}^I_{\mathrm{SBM}}(Q_I) + \mathrm{ICL}^O_{\mathrm{SBM}}(Q_O). \quad (12)$$

The ICL criterion favors independence if

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

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.

Remark. *If $Q_I = 1$ or $Q_O = 1$, the MLVSBM is the product of two independents SBM, as such* $\mathrm{ICL}_{\mathrm{Ind}}(Q_I,Q_O) = \mathrm{ICL}_{\mathrm{MLVSBM}}(Q_I,Q_O)$.

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 computationally intensive tasks. Indeed, we should compare all the possible models – one model corresponding to a given (Q_I,Q_O) – through the ICL criterion. Furthermore, for each model, the variational EM algorithm should be initialized 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 variational EM algorithm. The procedure we suggest is given in Algorithm 2.

Each step of the algorithm requires $O(\max\{Q_I,Q_O\}^2)$ variational EM algorithms which converge in a few iterations as a result of the local initialization. Inferring an independent SBM on each level beforehand is a fast way to start with good initialization 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_{Ind} and $\mathrm{ICL}_{\mathrm{MLVSBM}}(\widehat{Q_I},\widehat{Q_O})$.

Package All the codes are available as an `R` package 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. It also handles missing at random data (Rubin, 1976) on the adjacency matrices and link prediction.

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).*

Algorithm 2: Model selection algorithm

Data: $\{X^I, X^O, A\}$, a multilevel network

.

Procedure:

• Infer independent SBMs on X^I and X^O for a respective range of Q_I and Q_O. Deduce

$$\widehat{Q_I}^{\text{Ind}} = \arg\max_{Q_I} \text{ICL}^I_{\text{SBM}}(Q_I) \quad \text{and} \quad \widehat{Q_O}^{\text{Ind}} = \arg\max_{Q_O} \text{ICL}^O_{\text{SBM}}(Q_O).$$

Compute $\text{ICL}_{\text{Ind}} = \text{ICL}^I_{\text{SBM}}(\widehat{Q_I}^{\text{Ind}}) + \text{ICL}^O_{\text{SBM}}(\widehat{Q_O}^{\text{Ind}})$.

• Start at $Q_I = \widehat{Q_I}^{\text{Ind}}$ and $Q_O = \widehat{Q_O}^{\text{Ind}}$.

while *ICL is increasing* **do**

- Fit a MLVSBM on every model of size $(Q_I \pm 1, Q_O \pm 1)$ initialized by merging 2 blocks or splitting a block with hierarchical clustering.
- Among all estimated models, keep the one with the highest ICL.

return $(\widehat{Q_I}, \widehat{Q_O}) = \arg\max \text{ICL}(Q_I, Q_O)$, $\widehat{\theta}_{(\widehat{Q_I}, \widehat{Q_O})}$ *and* $\widehat{\mathbf{Z}}$.

4.1 Experimental design

In what follows, we set $Q_O = Q_I = 3$. The networks are of sizes: $n_O = 20,\ Q_O = 60$ and $n_I = 60,\ Q_I = 180$.

Let d be a density parameter: the lower d, the sparser the network and the harder the inference. ϵ (≥ 1) is a parameter tuning the strength of the communities; when ϵ 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: $\alpha^I = d * \begin{bmatrix} 1 & \epsilon & \epsilon \\ \epsilon & 1 & \epsilon \\ \epsilon & \epsilon & 1 \end{bmatrix}$.
- *Core-periphery.* A core block is highly connected to the whole network while the probability of connection in the periphery is low: $\alpha^I = d * \begin{bmatrix} \epsilon & \epsilon & 1 \\ \epsilon & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$.

We fix the topology of the inter-organizational 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 to 0.01, 0.05 or 0.1 while ϵ ranges from 1 to 10 by stepize of 0.5. $\epsilon = 1$ corresponds to an Erdős-Rényi graph and the communities should be indistinguishable.
The affiliation matrix A is generated from a power-law distribution in order to get different sizes of organizations. Other distributions were tried but the results (not reported here) show that their impact on the inference is weak.

Finally, δ is a parameter for the strength of the dependence between levels, ranging from 0 to 1. More precisely, we set:

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

where γ has been defined in Equation (4). $\delta = 1/Q_I$ corresponds to the case of independence between levels. The further δ is from $1/Q_I$, the stronger the dependence between levels. $\delta = 1$ implies a deterministic link between the clustering of the two levels, ie. the block of an individual is fully determined by the block of his/her organization.

4.2 Simulation results

First, we fix $\delta = 0.8$ and make ϵ vary. Each situation is simulated 50 times. We test the ability of our model to recover the true clustering of Z^I from (X^I, X^O). We compare our performances to the ones obtained by applying a standard (unilevel) SBM on X^I. Because (Q_I, Q_O) are assumed to be unknown, two types of error may occur: one for not selecting the right Q_I 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 the 3 topologies (assortative, core-periphery and disassortative) – the ARI when using MLVSBM (plain line) and SBM (dashed line) as ε varies. We observe that, for any topology, the MLVSBM starts to recover perfectly the clustering for a lower value of ϵ than the SBM. The difficulty of the inference increases as ϵ decreases: as can be seen in Figure 4 A, MLVSBM still performs well ($ARI > 0$) for small values of ϵ while the SBM is unable to recover the clustering.

In Figures 4 B and C, we plot the number of blocks chosen by the MLVSBM (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 using the MLVSBM allows to recover more precisely Q_I than using the SBM. $\widehat{Q_I}$ varies from 1, when no structure is detected to 3 which is the true number of blocks. The procedure never selects more blocks than expected, which is coherent with prior knowledge that the ICL for the SBM tends to select models of smaller size (Hayashi et al., 2016; Brault, 2014).

On the three topologies with $\epsilon = 3$, depending on the density d, MLVSBM and SBM supply Z^I either a perfect recovery of the clustering or a random clustering or something

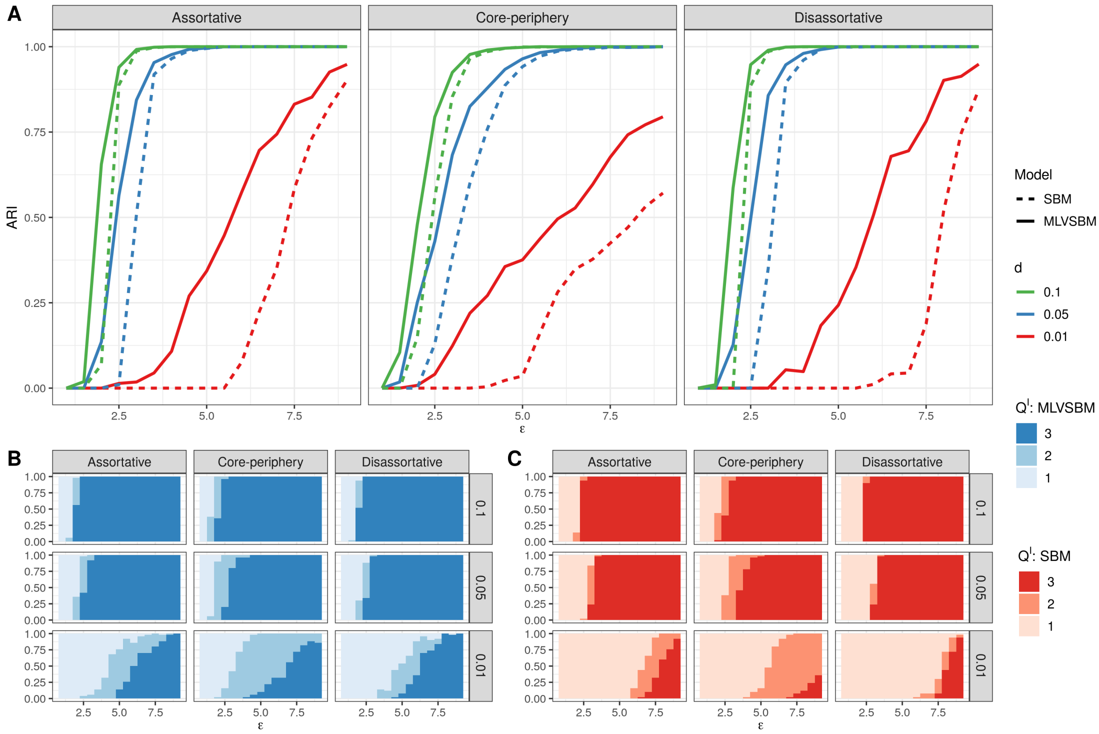

Figure 4: Clustering and model selection for 3 different topologies on the inter-individual level, varying ϵ and density d. Each situation is simulated 50 times. **A**: ARI for the inter-individual level, comparing the model used for inference. **B**: Stacked frequency barplot of the selected number of blocks for the inter-individual level in the MLVSBM (in blue). **C**: Stacked frequency barplot of the selected number of blocks for the inter-individual level chosen in the SBM (in red).

in between. In order to understand better this phenomenon, we fix ϵ to 3 and make δ – which quantifies the dependency between the two levels – vary. The results are reported in Figure 5 for 50 simulations of each situation.

When $\delta = 1/3$ (yellow vertical line in Figure 5 A), the two levels are independent and the results in terms of clustering are the same for the MLVSBM and the SBM on X^I (see ARI in Figure 5 A). As soon as δ departs from this value, the MLVSBM is able to recover some of the structure of the inter-individual level thanks to the inter-organizational level and this ability is observed even for very low density when δ 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 $d = 0.01$, X^I is very sparse, $\widehat{Q_I} = 1$ (no structure is detected on the inter-individual level) leading to $\mathrm{ICL}_{ind} = \mathrm{ICL}_{\mathrm{MLVSBM}}$ and enabling us from detecting any dependency. For higher densities, we see as expected, that if $\delta \approx 1/3$, the independent SBM will be preferred. On the contrary the further δ departs

from 1/3 the more the MLVSBM will be selected, even-though the MLVSBM and the independent SBM may provide the same clusterings. This phenomenon occurs faster for higher density d. In our simulation, the MLVSBM is never selected when $\delta = 1/3$. This is a consequence of the conservative nature of ICL, requiring strong evidence from the likelihood to select a more complex model.

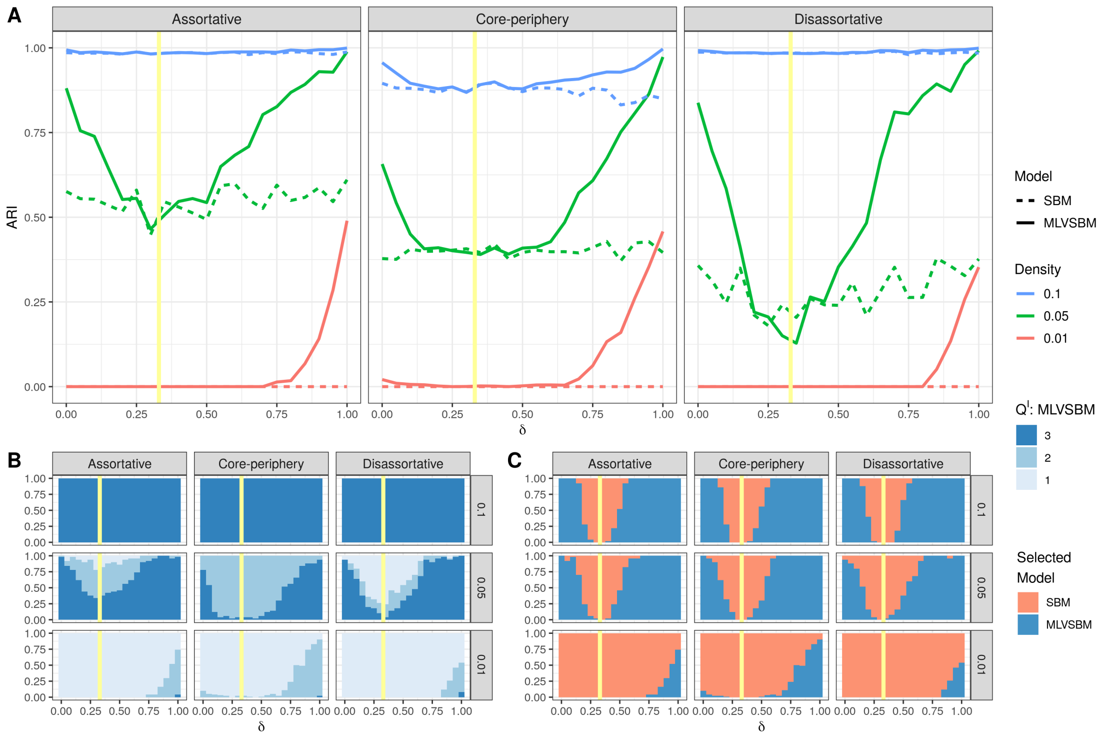

Figure 5: Clustering and model selection for 3 different topologies on the inter-individual level, as function of δ and density d. Each situation is simulated 50 times. The yellow vertical lines represent a $\delta = 1/3$ (i.e. a γ with uniform coefficients, resulting into independence between the two levels). **A**: ARI for the inter-individual level, comparing the model used for inference. **B**: Stacked frequency barplot of the selected number of blocks for the inter-individual level in the MLVSBM. **C**: Stacked frequency barplot of the selected model with respect to inter-level dependence.

Simulations gave similar results (not reported here) when we inverse the topologies on X^I and X^O, 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 recovering of 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 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 data gather observations on one particular audiovisual product, namely animation and cartoons. From a sociological perspective, reconstituting and analyzing multilevel (inter-individual and inter-organizational) networks in this industry is important. In economic sociology, it helps redefine the nature of markets (Brailly et al., 2016, 2017; Lazega and Mounier, 2002). In the sociology of culture, it helps understand, from a structural perspective, the mechanisms underlying contemporary globalization and standardization of culture (Brailly et al., 2016; Favre et al., 2016). In the sociology of organizations 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.

The data were collected by face-to-face interviews. 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). The individuals were affiliated to 109 organizations, each one containing from one to six individuals. At the inter-organizational level, two kinds of interactions were collected: a deal network (deals signed since the last trade fair) and a meeting network (derived from the aggregation at the inter-organizational level of the meetings planned by individuals on the trade fair's website). Both networks are symmetric with respective densities 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-organizational level, the other dataset to the meeting network at the inter-organizational level). In both cases the ICL criterion favors dependence between the two levels and chooses $\widehat{Q_I} = 4$ blocks of individuals. $\widehat{Q_O}$ is equal to 3 for the deal network and 4 for the meeting network.

In order to determine which is the most relevant inter-organizational 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 set $X^I_{ii'} = \mathrm{NA}$ for a certain percentage of (i, i') (this percentage ranging from 5% to 40% by step-size of 5%). We also propose to remove existing links (ie. forcing $X^I_{ii'} = 0$ when $X^I_{ii'} = 1$ was observed, for some randomly chosen (i, i')). The percentage of removed existing links varies from 5% to 95% (with step-size of 5%). We repeat the following procedure 100 times:

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 $\mathbb{P}(X^I_{ii'} = 1; \widehat{\theta})$ for each missing dyad or for each dyad such that $X^I_{ii'} = 0$

3. Predict link among all missing dyads or among all dyads such that $X^I_{ii'} = 0$.

Missing data are handled as Missing At Random (Tabouy et al., 2019) and the probability of existence of an edge is given by: $\mathbb{P}(X^I_{ii'} = 1; \widehat{\theta}) = \sum_{k,k'} \widehat{\tau^I_{ik}} \widehat{\alpha^I_{kl}} \widehat{\tau^I_{i'k'}}$. Since the result of our procedure is equivalent to a binary classification problem, we assess the performance through the area under the ROC curve (AUC) (a random classification corresponding to AUC $= 0.5$).

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. Moreover, 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-organizational level by making the union of links of the deal and the meeting network. The improvement in terms of prediction over the deal network is not very significant and this composite network is much harder to analyze sociologically.

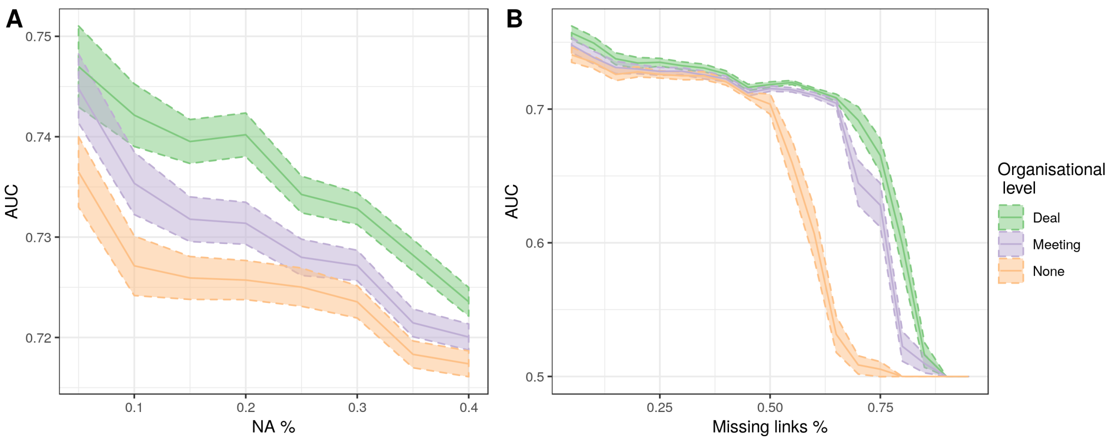

Figure 6: AUC of the prediction for **A**: missing dyads, **B**: missing links, in function of the missing proportion for the inter-individual level. Colors represent different network at the inter-organizational level. None (beige) is equivalent to a single layer SBM on the individuals. The confidence interval is given by $mean \pm stderror$.

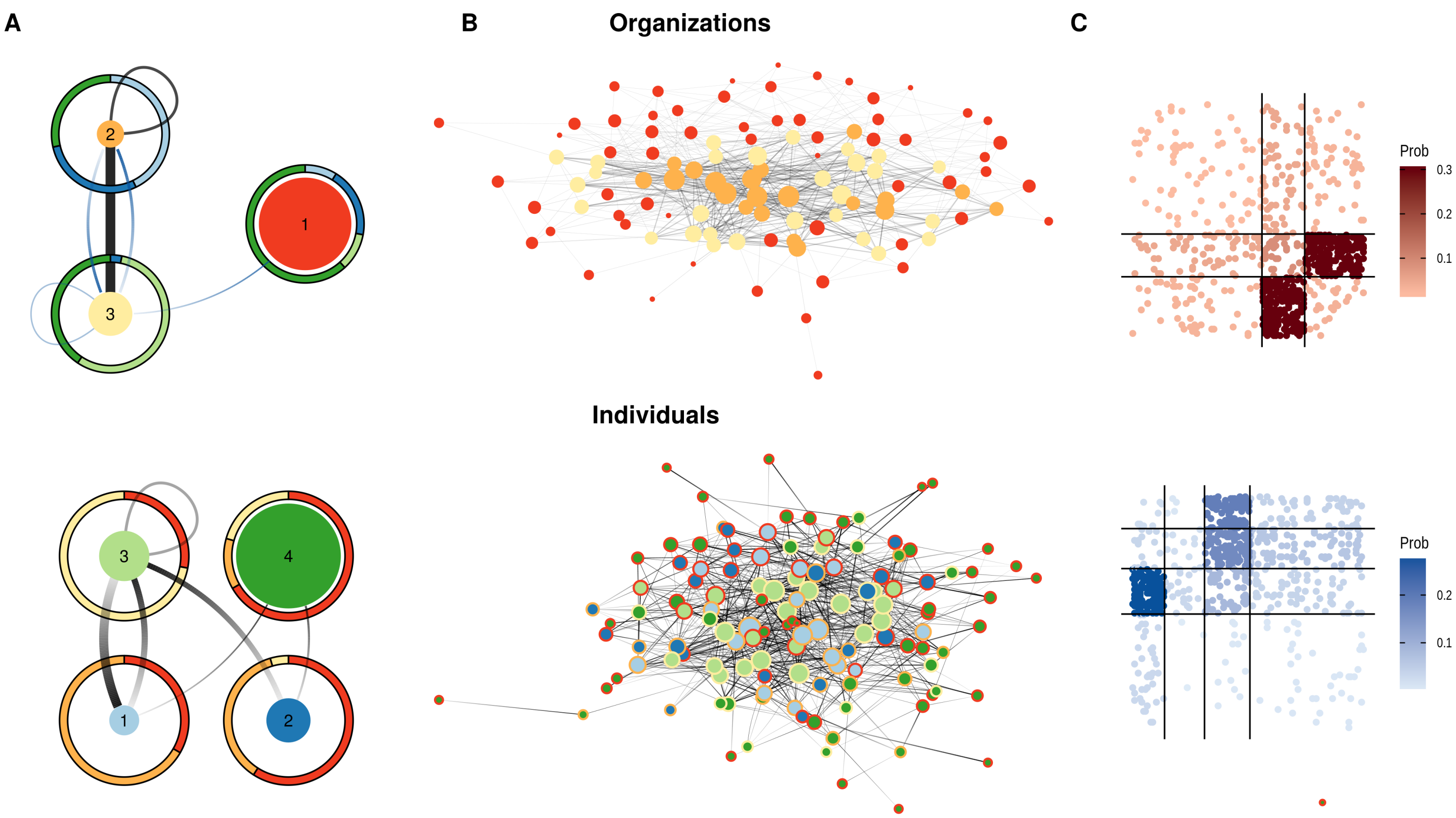

Figure 7: Multilevel network of the Promoshow East trade fair 2011. Above: the deal network for the organizations and below: the advice network for the individuals. **A**: Mesoscopic view of the multilevel network. Nodes stand for the blocks, donut charts show the relation between Z^O and Z^I. Black edges are the probabilities of connection α^I and α^O, blue edges stand for $\mathbb{P}(X^I_{ii'} = 1 | Z^O_{A_i}, Z^O_{A_{i'}})$. For sake of clarity only edges with probabilities above the density are shown. **B**: View of the network. The size of a node is proportional to its centrality degree. Colors represent the clustering obtained with the multilevel SBM. **C**: Adjacency matrices of the advice network between individuals and the deal network between organizations. Entries are reordered by block from left to right and top to bottom.

5.3 Analysis and comments

For the analysis, we use the MLVSBM inferred from the deal network. We select $\widehat{Q_O} = 3$ and $\widehat{Q_I} = 4$ blocks and the ICL is in favor of a dependence between the two networks. This network is plotted in Figure 7 B and we reordered the adjacency matrices of both levels by blocks in Figure 7 C. In Figure 7 A, we plot a synthetic view of the blocks of this multilevel network. The size of each node is proportional to the cardinal of each block. For the inter-organizational level, we link blocks of organizations by α^O (plain black edges) and by the probability of interactions of their individuals $\mathbb{P}(X^I_{ii'} = 1 | Z^O_{A_i}, Z^O_{A_{i'}})$ (gradual blue edges). The donut charts around the nodes is the parameter γ. For the inter-individual level, blocks of individuals are linked by α^I and the donut chart for a given block is the apportionment of each block of organizations in the individuals' affiliation.

We can now interpret the block with respect to the actors' covariates shown in Table 1. At the inter-organizational level, block 1 (in red) is a residual group composed of 61 organizations that are weakly connected to the rest of the organizations. Block 2 (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-block connection, but deals are mainly done between organizations of the blocks 2 and 3 (block 3 in yellow), the latter mostly containing distributors.

At the inter-individual level, blocks 1 and 2 consist of buyers (exclusively for block 1). They differ in their affiliations, both are affiliated to the second block of organizations but a larger proportion of the individuals of block 2 are affiliated to the residual block of organizations. 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 block 3 of organizations. It is also the block that has the strongest intra-block connections.

The blue edges in Figure 7 A show that the organizations of blocks 2 and 3 and their respective individuals follow the same pattern for their inter-block connections but differ in their intra-block connections. Individuals affiliated to organizations of block 3 have above average intra-block connections while few contracts are signed between their organizations (mainly distributors).

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 organizations help manage the tensions between these directly competing organizations (Lazega et al., 2016; Lazega, 2009). 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*, i.e. 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 organizations (broadcasting companies and distributors) exchange advice from sellers of block 3 representing production and distribution companies: this is the normal, stabilized, overlapping, commercial ties between companies embedded in social ties between representatives.

As seen above, block 3 has strong intra-block connections which may signal discreet coordination efforts between sellers as shown by Brailly (2016); Brailly et al. (2016). When a seller has closed a deal with a buyer, he/she 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. This is the most unexpected phenomenon from an orthodox economic perspective and should lead to new perspectives in neo-structural economic sociology (Lazega and Mounier, 2002).

This cross-level interdependence between inter-organizational 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 realized 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 SBM 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 the levels at the same time. There are clear advantages at considering a joint modeling of the two levels over an independent model for each level. Indeed, we show on some 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. On the trade fair dataset, this joint modeling brought us a synthetic representation of the two networks unraveling their intertwined structure and provide new insights on the social organization.

In lieu of a Bernoulli distribution, the edge distribution of any level may be extended to a valued distribution and/or to include edge covariates in a similar way as for the SBM (Mariadassou et al., 2010). One way to account for the degree distibution would be to use nodes degrees as covariates, another would be to rewrite the edge distribution as the Degree Corrected SBM (Karrer and Newman, 2011). Our choice to model the

Organizations		Covariates				
Block	Size	Producer	Distributor	Media group	Independent buyer	Broadcaster
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

Table 1: Contingency table of covariates and clustering for the organizations (top) and the individuals (bottom)

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 consider a unique affiliation per indicidual since this was the case on the datasets available to us, but this approach could be extended to a less restricted number of affiliations (this model is 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 modeling 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 organizations. But, doing so, we must make an assumption on how individuals are affiliated within a block of organizations. In a work not presented in this paper, we built such a model assuming that the affiliations were uniform within the blocks. Some simulation studies showed that the obtained clustering is very similar to the one obtained with the model presented in this paper, when the organizations belonging to the same block are of comparable sizes. In the contrary, if the sizes of the organizations in each block vary a lot, the clustering of the organizations 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 organization'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 a clustering of layer by aggregating the most similar. Our work considers multilevel networks where each level has nodes of different natures and Figure 6 shows that the 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.

Bartolucci, F., M. F. Marino, and S. Pandolfi (2018). Dealing with reciprocity in dynamic stochastic block models. *Computational Statistics & Data Analysis 123*, 86–100.

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.

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.

Mariadassou, M., S. Robin, and C. Vacher (2010). Uncovering latent structure in valued graphs: a variational approach. *The Annals of Applied Statistics 4*(2), 715–742.

Matias, C. and V. Miele (2017). Statistical clustering of temporal networks through a dynamic stochastic block model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79*(4), 1119–1141.

Rubin, D. B. (1976). Inference and missing data. *Biometrika 63*(3), 581–592.

Snijders, T. A. (2001). The statistical evaluation of social network dynamics. *Sociological methodology 31*(1), 361–395.

Snijders, T. A. (2017). Stochastic actor-oriented models for network dynamics.

Stanley, N., S. Shai, D. Taylor, and P. J. Mucha (2016). Clustering network layers with the strata multilayer stochastic block model. *IEEE transactions on network science and engineering 3*(2), 95–105.

Sweet, T. M., A. C. Thomas, and B. W. Junker (2013). Hierarchical network models for education research: Hierarchical latent space models. *Journal of Educational and Behavioral Statistics 38*(3), 295–318.

Sweet, T. M., A. C. Thomas, and B. W. Junker (2014). Hierarchical mixed membership stochastic blockmodels for multiple networks and experimental interventions. *Handbook on mixed membership models and their applications*, 463–488.

Tabouy, T., P. Barbillon, and J. Chiquet (2019). Variational inference for stochastic block models from sampled data. *Journal of the American Statistical Association*, 1–23.

Wang, P., G. Robins, P. Pattison, and E. Lazega (2013). Exponential random graph models for multilevel networks. *Social Networks 35*(1), 96–115.

Yan, X. (2016). Bayesian model selection of stochastic block models. In *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 323–328. IEEE Press.

Žiberna, A. (2014). Blockmodeling of multilevel networks. *Social networks 39*, 46–61.

Zijlstra, B. J., M. A. Van Duijn, and T. A. Snijders (2006). The multilevel p2 model. *Methodology 2*(1), 42–47.

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

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

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 organization j, we can write, for any k:

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

However, this quantity does not depend on Z_j^O so $\gamma_{k Z_j^O} = \gamma_{k\cdot}$ 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.:

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

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:

$$\tau_l = \mathbb{P}(X^O_{ij} = 1|Z^O_i = l) = \sum_{l'} \alpha^O_{ll'}\pi^O_{l'} = (\alpha^O \cdot \pi^O)_l, \quad \forall(i,j). \tag{B.13}$$

Moreover, a quick computation proves that

$$\mathbb{P}(X^O_{i,j:(j+k)} = 1|Z^O_i = l) = \tau_l^{k+1} \tag{B.14}$$

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

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

is invertible. We define u^O_i as follows :

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

The existence of $(u^O_i)_{i=0,\dots,2Q_O-1}$ comes from Assumption $\mathcal{A}2$ ($n_O \ge 2Q_O$). 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 $1 \le i \le 2Q_O - 1$:

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

using Equation (B.14). Now, let us define M^O a $(Q_O + 1) \times Q_O$ matrix such that:

$$M^O_{ij} = u^O_{i+j-2} = \sum_{l=1}^{Q_O} \tau^{i-1}_l \pi^O_l \tau^{j-1}_l, \quad 1 \le i \le Q_O + 1, \quad 1 \le j \le Q_O. \tag{B.15}$$

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:

$$B^O(x) = \sum_{k=0}^{Q_O} (-1)^{k+Q_O} \delta_{k+1} x^k. \tag{B.16}$$

- B^O is of degree Q_O. Indeed, $\delta_{Q_O+1} = \det(M^O_{-(Q_O+1)})$ and $M_{-(Q_O+1)} = R^O D_{\pi^O} R^{O\prime}$ 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, $\forall l = 1, \dots, Q_O$, $B^O(\tau_l) = 0$. Indeed, $B^O(\tau_l) = \det(N^O_l)$ where N^O_l is the concatenated matrix $N^O_l = \big(M^O \,|\, V_l\big)$ with $V_l = [1, \tau_l, \dots, \tau^{Q_O}_l]'$ (computation of the determinant development against the last column). However, from Equation (B.15), we have $M^O_{\bullet j} = \sum_l \tau^{j-1}_l \pi^O_l V_l$, i.e. each column vector of M^O is a linear combination of $V_1, \dots, V_{Q_O}$. As a consequence, $\forall l = 1, \dots, Q_O$, N^O_l is of rank $< Q_O + 1$, and so $B^O(\tau_l) = 0$.

The $(\tau_l)_{l=1,\dots,Q_O}$ being the roots of B, they can be expressed in a unique way (up to label switching) as functions of $(\delta_k)_{k=0,\dots,Q_O}$, which themselves are derived from $\mathbb{P}_{\mathbf{X},\theta}$. As a consequence, the identifiability of R^O is derived from the identifiability of $(\tau_l)_{l=1,\dots,Q_O}$. Using the fact that $D_{\pi^O} = R^{O^{-1}} M^O_{-Q_O} R^{O'^{-1}}$, 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 $k = 1\dots, Q_I$ and for $i = 2, \dots, 2Q_I$, let σ_k be the probability $\mathbb{P}(X_{1i}^I = 1 | Z_1^I = k, A)$, A being such that $A_{jj} = 1, \forall j = 1, \dots, 2Q_I$.

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

Moreover,

$$\begin{aligned}\mathbb{P}(Z_i^I = k' | Z_1^I = k, A) &= \sum_l \mathbb{P}(Z_i^I = k' | Z_i^O = l, Z_1^I = k, A) \mathbb{P}(Z_i^O = l | Z_1^I = k, A) \\ &= \sum_l \gamma_{kl} \mathbb{P}(Z_i^O = l | Z_1^I = k, A)\,. \end{aligned} \tag{B.17}$$

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 organization 1 and i to organization 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.17), we have:

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

and so:

$$\begin{aligned}\sigma_k &= \sum_{k'} \mathbb{P}(X^I_{1i} = 1 | Z^I_1 = k, Z^I_i = k') \sum_l \gamma_{k'l}\pi^O_k \\ &= \sum_{k'l} \alpha^I_{kk'}\gamma_{k'l}\pi^O_l = (\alpha^I \cdot \gamma \cdot \pi^O)_k \\ &= (\alpha^I \cdot \pi^I)_k, \qquad \text{where } \pi^I = \gamma \cdot \pi^O.\end{aligned}$$

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

$$\mathbb{P}(X^I_{1,2:(i+1)} = 1 | Z^I_1 = k, A) = \sigma^i_k. \tag{B.18}$$

Indeed,

$$\begin{aligned}&\mathbb{P}(X^I_{1,2:(i+1)} = 1 | Z^I_1 = k, A) \\ &= \sum_{k_{2:(i+1)}} \mathbb{P}(X^I_{1,2:(i+1)} = 1 | Z^I_{1:(i+1)} = (k, k_{2:(i+1)}), Z^I_1 = k)\mathbb{P}(Z^I_{2:(i+1)} = k_{2:i+1} | Z^I_1 = k, A) \\ &= \sum_{k_{2:(i+1)}} \mathbb{P}(X^I_{1,2:(i+1)} = 1 | Z^I_{1:(i+1)} = (k, k_{2:(i+1)}))\mathbb{P}(Z^I_{2:(i+1)} = k_{2:i+1} | A) \\ &= \sum_{k_{2:(i+1)}} \mathbb{P}(X^I_{1,2:(i+1)} = 1 | Z^I_{1:(i+1)} = (k, k_{2:(i+1)})) \sum_{l_{2:(i+1)}} \mathbb{P}(Z^I_{2:(i+1)} = k_{2:(i+1)}, Z^O_{2:(i+1)} = l_{2:(i+1)}, A).\end{aligned}$$

Note that, to go from line 2 to line 3, we used the fact that $\mathbb{P}(Z^I_{2:(i+1)} = k_{2:i+1} | Z^I_1 = k, A) = \mathbb{P}(Z^I_{2:(i+1)} = k_{2:i+1} | A)$, which is due the the particular structure of A (left diagonal block of size at least $2Q_I$, i.e. for any $i' = 1, \ldots, 2Q_I$, individual i' belongs to organization i'). Moreover, we can write:

$$\begin{aligned}&\mathbb{P}(Z^I_{2:(i+1)} = k_{2:(i+1)}, Z^O_{2:(i+1)} = l_{2:i+1} | A) \\ = &\left[\prod_{\lambda=2,\ldots i+1} \mathbb{P}(Z^I_\lambda = k_\lambda | Z^O_\lambda = l_\lambda)\mathbb{P}(Z^O_\lambda = l_\lambda)\right] \\ = &\left[\prod_{\lambda=2,\ldots i+1} \gamma_{k_\lambda l_\lambda}\pi^O_\lambda\right].\end{aligned}$$

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

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

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 $\forall 1 \le i \le 2Q_I - 1$:

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

  Note that the (u^I)'s can be defined because $n_I \ge 2Q_I$ (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.16)). 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 $D_{\pi^I} = \operatorname{diag}(\pi^I) = \operatorname{diag}(\gamma \cdot \pi^O)$, 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 $2 \le i \le Q_I$ and $2 \le j \le Q_O$, 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 $n_O \ge Q_I + Q_O - 1$ and $n_I \ge Q_I$ (assumption $\mathcal{A}2$).

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

$$
\begin{aligned}
U_{ij}^{IO} &= \mathbb{P}(X_{1,2:i}^I = 1, X_{1,(i+1):(i+j-1)}^O = 1|A) \\
&= \sum_{k,l} \mathbb{P}(X_{1,2:i}^I = 1, X_{1,(i+1):(i+j-1)}^O = 1|A, Z_1^I = k, Z_1^O = l) \\
& \quad \times \mathbb{P}(Z_1^I = k, Z_1^O = l, A)\,. \qquad \text{(B.19)}
\end{aligned}
$$

- We first prove that :

$$
\mathbb{P}(X_{1,2:i}^I = 1, X_{1,i+1:i+j-1}^O = 1|A, Z_1^I = k, Z_1^O = l) = \sigma_k^{i-1}\tau_l^{j-1}\,. \qquad \text{(B.20)}
$$

Indeed,

$$
\begin{aligned}
&\mathbb{P}(X_{1,2:i}^I = 1, X_{1,(i+1):(i+j-1)}^O = 1|A, Z_1^I = k, Z_1^O = l) = \\
&= \sum_{k_{2:i}, l_{2:n_O}} \mathbb{P}(X_{1,2:i}^I = 1, X_{1,(i+1):(i+j-1)}^O = 1|Z_{1:i}^I = (k, k_{2:i}), Z^O = (l, l_{2:n_O}), A) \\
&\quad \times \ \mathbb{P}(Z_{2:i}^I = k_{2:i}, Z_{2:n_O}^O = l_{2:n_O}|Z_1^I = k, Z_1^O = l, A) \\
&= \sum_{k_{2:i}, l_{2:n_O}} \mathbb{P}(X_{1,2:i}^I = 1|Z_{1:i}^I = (k, k_{2:i})) \\
&\quad \times \ \mathbb{P}(X_{1,(i+1):(i+j-1)}^O = 1|Z_1^O = l, Z_{(i+1):(i+j-1)}^O = l_{(i+1):(i+j-1)}) \\
&\quad \times \ \mathbb{P}(Z_{2:i}^I = k_{2:i}, Z_{2:n_O}^O = l_{2:n_O}|Z_1^I = k, Z_1^O = l, A)\,. \qquad \text{(B.21)}
\end{aligned}
$$

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.21) and decomposing the summation we obtain:

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

Finally, we have :

$$
\begin{aligned}
\mathbb{P}(X^I_{1,2:i} = 1 | Z^I_1 = k, A) &= \sigma_k^{i-1}, \quad \text{from Equation (B.18)} \\
\mathbb{P}(X^O_{1,(i+1):(i+j-1)} = 1 | Z^O_1 = l) &= \tau_l^{j-1},
\end{aligned}
$$

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

- 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.20) and (B.19) 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 $\gamma = (R^I)^{-1} U^{IO} (R^{O'})^{-1} D^{-1}_{\pi^O}$ and γ is identified.

$\square$

C Details of the Variational EM

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

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

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

VE step compute

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

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:

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

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 emphasizes the dependency structure of this multilevel model and the role of γ as the dependency parameter of the model. Notice also that when $\gamma_{kl} = \gamma_{kl'} = \pi_k^I$ for all l, l', that is the case of independence between the two levels then we can rewrite the fixed point relationships as follows:

$$\widehat{\tau_{jl}^O} \propto \pi_l^O \prod_{j' \neq j} \prod_{l'} \phi(X_{jj'}^O, \alpha_{ll'}^O)^{\widehat{\tau_{j'l'}^O}} \quad \text{and} \quad \widehat{\tau_{ik}^I} \propto \pi_k^I \prod_{i' \neq i} \prod_{k'} \phi(X_{ii'}^I, \alpha_{kk'}^I)^{\widehat{\tau_{i'k'}^I}},$$

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:

$$\widehat{\pi_l^O} = \frac{1}{n_O} \sum_j \widehat{\tau_{jl}^O} \qquad \widehat{\alpha_{ll'}^O} = \frac{\sum_{j' \neq j} \widehat{\tau_{jl}^O} X_{jj'}^O \widehat{\tau_{j'l'}^O}}{\sum_{j' \neq j} \widehat{\tau_{jl}^I} \widehat{\tau_{j'l'}^I}}$$

$$\widehat{\gamma}_{kl} = \frac{\sum_{i,j} \widehat{\tau_{ik}^I} A_{ij} \widehat{\tau_{jl}^O}}{\sum_{i,j} A_{ij} \widehat{\tau_{jl}^O}} \qquad \widehat{\alpha_{kk'}^I} = \frac{\sum_{i' \neq i} \widehat{\tau_{ik}^I} X_{ii'}^I \widehat{\tau_{i'k'}^I}}{\sum_{i' \neq i} \widehat{\tau_{ik}^I} \widehat{\tau_{i'k'}^I}}$$

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_l^O is the mean of the posterior probabilities for the organizations to belong to block l. $\alpha_{kk'}^I$ (resp. $\alpha_{ll'}^O$) is the 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 block k that are affiliated to any organization of block l on the number of individuals that are affiliated to any organization of block l. If γ is such that the levels are independent, then any column of γ represents the proportion of individuals in the different blocks:

$$\pi_k^I = \gamma_{k1} = \frac{1}{n_I} \sum_i \widehat{\tau_{ik}^I}.$$

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 (9)) 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:

$$
\begin{aligned}
\log \ell_\theta(\mathbf{X}, \mathbf{Z}|A, Q_I, Q_O) &= \log\left(\int_\Theta \ell_\theta(X^I, X^O, Z^I, Z^O|\theta, A, Q_I, Q_O)p(\theta|Q_I, Q_O)\mathrm{d}\theta\right) \\
&= \log \ell_{\alpha^I}(X^I|Z^I, Q_I) \qquad \text{(D.22)} \\
&\quad + \log \ell_\gamma(Z^I|A, Z^O, Q_I, Q_O) \qquad \text{(D.23)} \\
&\quad + \log \ell_{\alpha^O, \pi^O}(X^O, Z^O|Q_O)\,. \qquad \text{(D.24)}
\end{aligned}
$$

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

$$
\begin{array}{rcl}
\log \ell_{\alpha^O}(X^O, \widehat{Z^O}, Q_O) & \approx & {}_{n_O\to\infty} \log \ell_{\widehat{\alpha^O}, \widehat{\pi^O}}(X^O, \widehat{Z^O}|Q_O) - \mathrm{pen}(\pi^O, \alpha^O, Q_O) \\
\mathrm{pen}(\pi^O, \alpha^O, Q_O) & = & \frac{Q_O-1}{2}\log n_O + \frac{1}{2}\frac{Q_I(Q_I+1)}{2}\log\frac{n_I(n_I-1)}{2}
\end{array} . \qquad \text{(D.25)}
$$

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.22)) leads to:

$$
\begin{array}{rcl}
\log \ell_{\alpha^I}(X^I|\widehat{Z^I}, Q_I) & = & {}_{n_I\to\infty} \log \ell_{\widehat{\alpha^I}}(X^I|\widehat{Z^I}, Q_I) + \mathrm{pen}(\alpha^I, Q_I) \\
\text{with } \mathrm{pen}(\alpha^I, Q_I) & = & \frac{1}{2}\frac{Q_I(Q_I+1)}{2}\log\frac{n_I(n_I-1)}{2}
\end{array} . \qquad \text{(D.26)}
$$

For quantity (D.23) 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:

$$\begin{aligned}
\log \ell_\gamma(Z^I|A, Z^O, Q_I, Q_O) \quad \approx \quad & {}_{(n_O,n_I)\to\infty} \textstyle\sum_{k,l} (N_{kl} \log N_{kl} + N_{kl}) \\
& - \textstyle\sum_l \left(\frac{Q_I-1}{2} + \sum_k N_{kl}\right) \log\left(\sum_k N_{kl}\right) - \sum_{k,l} N_{kl}.
\end{aligned} \tag{D.27}$$

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

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

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

$$\log \ell_\gamma(\widehat{Z^I}|A, \widehat{Z^O}, Q_I, Q_O) \approx_{(n_O,n_I)\to\infty} \log \ell_{\hat\gamma}(\widehat{Z^I}|A, \widehat{Z^O}, Q_I, Q_O) - \frac{Q_I-1}{2} Q_O \log n_I . \tag{D.28}$$

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