Stochastic block models for multiplex networks: an application to a multilevel network of researchers

Pierre Barbillon and Sophie Donnet,

AgroParisTech, Paris, Institut National de la Recherche Agronomique, Paris, and Université Paris-Saclay, France

Emmanuel Lazega

Institut d'Études Politiques de Paris, France

and Avner Bar-Hen

Université Paris Descartes, France

Summary. Modelling relationships between individuals is a classical question in social sciences and clustering individuals according to the observed patterns of interactions allows us to uncover a latent structure in the data. The stochastic block model is a popular approach for grouping individuals with respect to their social comportment. When several relationships of various types can occur jointly between individuals, the data are represented by multiplex networks where more than one edge can exist between the nodes. We extend stochastic block models to multiplex networks to obtain a clustering based on more than one kind of relationship. We propose to estimate the parameters—such as the marginal probabilities of assignment to groups (blocks) and the matrix of probabilities of connections between groups—through a variational expectation–maximization procedure. Consistency of the estimates is studied. The number of groups is chosen by using the integrated completed likelihood criterion, which is a penalized likelihood criterion. Multiplex stochastic block models arise in many situations but our applied example is motivated by a network of French cancer researchers. The two possible links (edges) between researchers are a direct connection or a connection through their laboratories. Our results show strong interactions between these two kinds of connection and the groups that are obtained are discussed to emphasize the common features of researchers grouped together.

Keywords: Bivariate stochastic block model; Multilevel or multiplex networks; Social network

1. Introduction

Network analysis has emerged as a key technique for understanding and for investigating social interactions through the properties of relationships between and within units. From a statistical point of view, a network is a realization of a random graph formed by a set of nodes V representing the units (e.g. individuals, actors or companies) and a set of edges E representing relationships between pairs of nodes.

The system in which the same nodes belong to multiple networks is typically referred to as a multiplex network or multigraph (see Wasserman (1994), for example). In recent literature, there

has been an upsurge of interest in multiplex networks (see, for example, Cozzo *et al.* (2012), Loe and Jeldtoft Jensen (2014), Rank *et al.* (2010), Szell *et al.* (2010), Mucha *et al.* (2010), Maggioni *et al.* (2013), Brummitt *et al.* (2012), Saumell-Mendiola *et al.* (2012), Bianconi (2013) and Nicosia *et al.* (2013)). In these multiplex networks, different kinds of link (or connection) are possible for each pair of nodes. This induced link multiplexity is a fundamental aspect of social relations (Snijders and Baerveldt, 2003) since these multiple links are frequently interdependent: links in one network may have an influence on the formation or dissolution of links in other networks.

The simultaneous analysis of several networks also arises when one is interested in the social comportment of individuals belonging to organized entities (such as companies, laboratories or political groups), with some individuals possibly belonging to the same institution. Whereas the actors will exchange resources (such as advice) at the individual level, their respective organizations of affiliation will also share resources at the institutional level (financial resources for instance). Each level (individuals and organizations) constitutes a system of exchange of different resources that has its own logic and could be studied separately. However, studying the two networks jointly (and hence embedding the individuals in the multilevel relational and organizational structures constituting the interorganizational context of their actions) would allow us to identify the individuals who benefit from relatively easy access to the resources circulating in each level, which is of much more interest. In other words, studying the two levels jointly could help us to understand how an individual can benefit from the position of its organization in the institutional network.

In this paper, we are interested in studying the advice relationships between researchers and the exchanges of resources between laboratories. We adopt the following individual-oriented strategy (this point is discussed in the paper): the institutional network is used to define a new network on the individual level, i.e. the set of nodes consists in the set of individuals and, for a pair of individuals, two kinds of link are possible: a direct connection given by the individual network and a connection through their organizations given by the organizational network. As a consequence, the individual and institutional levels are fused into a multiplex.

We then develop a statistical model that can detect in multiplex substantial non-trivial topological features, with patterns of connection between their elements that are not purely regular. Several models such as scale-free networks and small world networks have been proposed to describe and understand the heterogeneity that is observed in networks. These models allow us to derive properties of the network at the macroscale and to understand the outcomes of interactions. To explore heterogeneity at others scales (such as microscale or mesoscale) in social networks, specific models such as stochastic block models (SBMs) (Snijders and Nowicki, 1997) have been developed for uniplex networks. In this paper, we propose an original extension of SBMs to the multiplex case. Our model is efficient in modelling not only the main effects (that correspond to a classical uniplex) but also the pairwise interactions between the nodes. We estimate the parameters of the multiplex SBM by using an extension of the variational expectation–maximization (EM) algorithm. Consistency of the estimation of the parameters is proved. As for uniplex SBMs, a key issue is to choose the number of blocks. We use a penalized likelihood criterion, namely the integrated completed likelihood (ICL). The inference procedure is performed on the cancer researchers and laboratories data set.

Since original work of Holland *et al.* (1983), probabilistic models for networks have been studied. It is out of the scope of this paper to review this very large literature but the most commonly used models are discussed in the survey by Goldenberg *et al.* (2010) as well as in Matias and Robin (2014a). From a theoretical point of view, the Szemerédi regularity lemma allows us to include SBMs in the theory of graph limits or inhomogeneous random graphs

(see Komlós and Simonovits (1996)). These results were recently extended to multigraphs (see Kunszenti-Kovács *et al.* (2014)). We shall not focus on limit theorems in the paper and we restrict our goals to estimation of the parameters (as well as interpretation) of the model proposed. Computational complexity of most of the proposed extensions leads to intractable models and we focus here on a multiplex graph with (unknown) block structure.

Another classical question when recovering block or communities in graphs is sparsity. Real world large-scale networks are usually sparse, in the sense that the node degrees are very small compared with the total number of nodes. Since the data are generally a sample it is important to understand what happens when the size of the sample increases. The extension of results on sparsity to the case of multiplex SBMs is an important open question. The first step is to extend the definition of sparsity to multiplex networks. In this work, we sampled the whole population and the aim of this study is to understand the block structure of this given population of researchers. Therefore we are not concerned with sparsity in our framework and we shall not study this question in this paper.

The paper is organized as follows. The extension of SBMs to multiplex networks is presented in Section 2, and the proofs of model identifiability and the consistency of the variational EM procedure are postponed until Appendices A and B. In Section 3, we describe the data set of Lazega *et al.* (2008) apply the new modelling and discuss the results. Finally, the contribution of multiplex SBMs to the analysis of multiplex networks is highlighted in Section 4.

The data that are analysed in the paper and the programs that were used to analyse them can be obtained from

```
http://wileyonlinelibrary.com/journal/rss-datasets
```

2. Multiplex stochastic block model

2.1. Description of the model

The main objective is to cluster the individuals (or nodes) into blocks sharing connection properties with the other individuals of the multiplex network. SBMs (Nowicki and Snijders, 2001) for random graphs have emerged as a natural tool to perform such a clustering based on uniplex networks (directed or not; valued or not). In what follows, we propose an extension of the SBM to multiplex networks.

Let $\mathbf{X}^1, \ldots, \mathbf{X}^K$ be K directed graphs relying on the same set of nodes $E = \{1, \ldots, n\}$. We assume that $\forall (i, j)$, $i \neq j$, and $\forall k \in \{1, \ldots, K\}$, $X_{ij}^k \in \{0, 1\}$. We define a joint distribution on $\mathbf{X}^{1:K} = (\mathbf{X}^1, \ldots, \mathbf{X}^K)$.

The simplest distribution would be $\forall (i, j) \in \{1, \ldots, n\}^2$, $i \neq j$, and $\forall w \in \{0, 1\}^K$,

$$P(X_{ij}^{1:K} = w) = \pi^{(w)} \qquad \sum_{w \in \{0,1\}^K} \pi^{(w)} = 1,$$

where the $(X_{ij}^{1:K})_{i,j}$ are mutually independent. This model is quite simple since any relationship between two individuals (a relationship being a collection of edges) is independent of the relationships between the other individuals. However, note that the different kinds of relationships ($k = 1, \ldots, K$) between two individuals (i, j) are not assumed to be independent.

To consider more complex patterns, we use the SBM approach, considering a mixture distribution on the $\{X_{ij}^{1:K}, i = 1, \ldots, n, j = 1, \ldots, n, i \neq j\}$.

Let Q be the number of blocks (groups or clusters). We introduce the latent variables $(Z_i)_{i=1,\ldots,n}$ such that $Z_i = q$ if individual i belongs to block q (note that an individual can belong to only one block in this version). The multiplex version of SBMs is written as follows: $\forall (i, j) \in \{1, \ldots, n\}^2, i \neq j, \forall w \in \{0, 1\}^K$ and $\forall (q, l) \in \{1, \ldots, Q\}^2$,

$$P(X_{ij}^{1:K} = w | Z_i = q, Z_j = l) = \pi_{ql}^{(w)},$$

$$P(Z_i = q) = \alpha_q. \tag{1}$$

Conditionally on the affectations $\mathbf{Z} = (Z_1, \dots, Z_n)$, the $(X_{ij}^{1:K})_{i \neq j}$ are independent. We stress again the fact that the independence is between ties and not between the types of tie ($k = 1, \dots, K$). Some particular types of dependence (for instance Markov dependence) between the various types or levels of ties can be introduced by constraints on the set of parameters $(\pi_{ql}^{(w)})_{w \in \{0,1\}^K, (q,l) \in \{1,\dots,Q\}^2}$.

The heterogeneity in the network topology is introduced through the latent variables.

Such a model includes $(2^K - 1)Q^2 + Q - 1$ parameters. Introducing the notation

$$\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_Q),$$

$$\boldsymbol{\pi} = (\pi_{ql}^{(w)})_{w \in \{0,1\}^K, (q,l) \in \{1,\dots,Q\}^2},$$

$$\theta = (\boldsymbol{\alpha}, \boldsymbol{\pi}),$$

the likelihood function is written as:

$$\begin{aligned} l(\mathbf{X}^{1:K}; \theta) &= \int_{\mathbf{z} \in \{1,\dots,Q\}^n} l_c(\mathbf{X}^{1:K}, \mathbf{z}; \theta) \, d\mathbf{z} \\ &= \sum_{\mathbf{Z} \in \{1,\dots,Q\}^n} p(\mathbf{X}^{1:K} | \mathbf{z}; \boldsymbol{\pi}) \, p(\mathbf{z}; \boldsymbol{\alpha}), \\ &= \sum_{\mathbf{Z} \in \{1,\dots,Q\}^n} \prod_{i,j, i \neq j} \pi_{Z_i Z_j}^{(X_{ij}^{1:K})} \prod_{i=1}^{n} \alpha_{Z_i}, \end{aligned} \tag{2}$$

where $l_c(\mathbf{X}^{1:K}, \mathbf{Z}; \theta)$ is the complete likelihood and $p(\mathbf{X}^{1:K} | \mathbf{z}; \boldsymbol{\pi})$ is the conditional probability. The latent variables $\mathbf{z}$ (the block affectations) are integrated out. The identifiability of the model can be proved (see Appendix A, theorem 1) and some consistency results can be proved on the maximum likelihood estimators (theorem 2).

Remark 1. In the multiplex SBM, the modelling is actor based, which means that the individuals are the same for all the networks $X^1, \dots, X^K$ and we model conjointly all the connections.

Remark 2. Some data sets can be enhanced by covariates $\mathbf{y}_{ij}$ characterizing the couples of nodes (i, j); the covariates $\mathbf{y}_{ij}$ depend on only the couple (i, j) and are not linked to the network under consideration. Introducing the covariates in the model can be judicious if one wants to highlight (and infer) heterogeneous patterns of connections once the effect of the covariates has been removed. In this particular case, the probability connections would be written as

$$P(X_{ij}^{1:K} = w | Z_i = q, Z_j = l) = \frac{\exp(\mu_{ql}^{(w)} + \beta^{(w)\mathrm{T}} \mathbf{y}_{ij})}{1 + \sum_{v \neq (0,\dots,0)} \exp(\mu_{ql}^{(v)} + \beta^{(v)\mathrm{T}} \mathbf{y}_{ij})},$$

$$P\{X_{ij}^{1:K} = (0,\dots,0) | Z_i = q, Z_j = l\} = \frac{1}{1 + \sum_{v \neq (0,\dots,0)} \exp(\mu_{ql}^{(v)} + \beta^{(v)\mathrm{T}} \mathbf{y}_{ij})}, \tag{3}$$

where x^{T} denotes the transposed vector of x.

2.2. Maximum likelihood estimation

As soon as n or Q is large, the observed likelihood (2) is not tractable (because of the sum

on $\mathbf{Z} \in \{1, \dots Q\}^n)$ and its maximization is a challenging task. Several approaches have been developed in the literature (for a review, see Matias and Robin (2014b)), both in the frequentist and in the Bayesian frameworks, starting from Snijders and Nowicki (1997) and Nowicki and Snijders (2001). However, when the latent data space is really large, these techniques can be burdensome. Some other strategies have been proposed, such as Bickel and Chen (2009), relying on a profile likelihood optimization or moment estimation proposed by Ambroise and Matias (2012) to name just a couple.

The variational EM algorithm in the context of the SBM proposed by Daudin *et al.* (2008) is a flexible tool to tackle the computational challenge in many types of graph. In a few words, the standard EM algorithm is a popular tool for estimation in models with latent variables, but it cannot be implemented easily in this context, owing to the absence of an explicit expression for the conditional distribution $p(\mathbf{Z}|\mathbf{X}^{1:K}; \theta)$. The variational EM algorithm aims at optimizing a lower bound of the log-likelihood, namely

$$\mathcal{I}_{\theta}(\mathcal{R}_{\mathbf{X}}^{1:K}) = \log\{l(\mathbf{X}^{1:K}; \theta)\} - \mathrm{KL}\{\mathcal{R}_{\mathbf{X}^{1:K}}, p(\cdot|\mathbf{X}^{1:K}; \theta)\}$$

where KL is the Kullback–Leibler divergence, $p(\cdot|\mathbf{X}^{1:K}; \theta)$ is the true posterior distribution of the latent variables $\mathbf{Z}$ given the observed data $\mathbf{X}^{1:K}$ and $\mathcal{R}_{\mathbf{X}^{1:K}}$ is an approximation of this conditional distribution. $\mathcal{I}_{\theta}(\mathcal{R}_{\mathbf{X}}^{1:K}) = \log\{l(\mathbf{X}^{1:K}; \theta)\}$ if and only if the $\mathcal{R}_{\mathbf{X}^{1:K}} = p(\cdot|\mathbf{X}^{1:K}; \theta)$. Dealing with the exact distribution is impossible, and so we approximate it by $\mathcal{R}_{\mathbf{X}}^{1:K}$, where $\mathcal{R}_{\mathbf{X}}^{1:K}$ belongs to a certain class of distributions. The variational EM algorithm alternates between the maximization of $\mathcal{I}_{\theta}(\mathcal{R}_{\mathbf{X}}^{1:K})$ with respect to $\mathcal{R}_{\mathbf{X}}^{1:K}$ and its maximization with respect to θ.

Simulation studies showed its practical efficiency (Mariadassou *et al.*, 2010; Gazal *et al.*, 2012). Moreover, its theoretical convergence towards the maximum likelihood estimates has been studied by Celisse *et al.* (2012) and more recently by Bickel *et al.* (2013) for standard SBMs. In Appendix B, we give the details of the implementation of the variational EM algorithm in multiplex SBMs. Some first results about its convergence towards the true parameter are proved (theorem 3).

2.3. Selection of the number of clusters Q

We chose to perform the selection of the most adequate number of blocks Q by using the ICL criterion whose principle is recalled later.

The Bayesian information criterion BIC is a penalized likelihood criterion, which is widely used for model choice. It is defined as

$$\mathrm{BIC} = \log\{l(\mathbf{X}^{1:K}; \hat{\theta}, \mathcal{M}_Q)\} - \mathrm{Pen}_{\mathrm{BIC}}(\mathcal{M}_Q), \tag{4}$$

$\mathrm{Pen}_{\mathrm{BIC}}(\mathcal{M}_Q)$ (penalizing the complexity of the model) derives from a Laplace approximation of the marginal likelihood $\int l(\mathbf{X}^{1:K}; \theta, \mathcal{M}_Q) \, p(\theta) \, \mathrm{d}\theta$. The BIC-criterion provides, under regularity conditions, a reliable approximation to this integrated likelihood. However, these regularity conditions on the likelihood function do not hold for mixture models or SBMs. Even if BIC is used in practice for model selection in mixture models, it suffers from major drawbacks (see the introduction in Biernacki *et al.* (2000)). Moreover, in the SBM context, the quantity $\log\{l(\mathbf{X}^{1:K}; \hat{\theta}, \mathcal{M}_Q)\}$ has no explicit expression (owing to the integration over the latent variables $\mathbf{Z}$).

As a consequence, the ICL criterion is an alternative to BIC in the mixture model context. ICL relies on the integrated complete log-likelihood $\log\{\int_{\theta} l_{\mathrm{c}}(\mathbf{X}^{1:K}, \mathbf{Z}; \theta, \mathcal{M}_Q) \, p(\theta) \, \mathrm{d}\theta\}$.

A direct and easy extension of the demonstration in Daudin *et al.* (2008) gives

$$\log\left\{\int_\theta l_c(\mathbf{X}^{1:K}, \mathbf{Z}; \theta, \mathcal{M}_Q)\, p(\theta)\,\mathrm{d}\theta\right\} = \log\{l_c(\mathbf{X}^{1:K}, \mathbf{Z}; \hat{\theta}, \mathcal{M}_Q)\}$$
$$-\tfrac{1}{2}[Q^2(2^K-1)\log\{Kn(n-1)\} + (Q-1)\log(n)]$$

when the prior distribution on θ, $p(\theta)$, is a well-chosen non-informative prior. This alternative penalty

$$\mathrm{Pen}_{\mathrm{ICL}}(\mathcal{M}_Q) = \tfrac{1}{2}\{Q^2(2^K-1)\log\{Kn(n-1)\} + (Q-1)\log(n)\}$$

is then used in the ICL criterion, defined as

$$\mathrm{ICL} = E_{\mathbf{Z}|\mathbf{X}^{1:K}, \hat{\theta}, \mathcal{M}_Q}[\log\{l_c(\mathbf{X}^{1:K}, \mathbf{Z}; \hat{\theta}, \mathcal{M}_Q)\}] - \mathrm{Pen}_{\mathrm{ICL}}(\mathcal{M}_Q). \tag{5}$$

Remark 3. Note that the ICL is a penalized conditional complete likelihood. Besides, using the following decomposition of $\log\{l(\mathbf{X}^{1:K}; \hat{\theta}, \mathcal{M}_Q)\}$, namely

$$\log\{l(\mathbf{X}^{1:K}; \hat{\theta}, \mathcal{M}_Q)\} = E_{\mathbf{Z}|\mathbf{X}^{1:K}, \hat{\theta}, \mathcal{M}_Q}[\log\{l_c(\mathbf{X}^{1:K}, \mathbf{Z}; \hat{\theta}, \mathcal{M}_Q)\}] + H(\mathbf{Z}|\mathbf{X}^{1:K}; \hat{\theta}, \mathcal{M}_Q) \tag{6}$$

where $H(\mathbf{Z}|\mathbf{X}^{1:K}, \hat{\theta}, \mathcal{M}_Q) = -\int \log\{p(\mathbf{Z}|\mathbf{X}^{1:K}; \hat{\theta}, \mathcal{M}_Q)\} p(\mathbf{Z}|\mathbf{X}^{1:K}; \hat{\theta}, \mathcal{M}_Q)\,\mathrm{d}\mathbf{Z}$ is the entropy of the conditional distribution $p(\mathbf{Z}|\mathbf{X}^{1:K}, \hat{\theta}, \mathcal{M}_Q)$, ICL can be seen as a penalized maximum likelihood, where the penalty includes the complexity of the model Pen(ICL) and the entropy of $p(\mathbf{Z}|\mathbf{X}^{1:K}, \hat{\theta}, \mathcal{M}_Q)$. Using the ICL will automatically encourage clustering configurations with well-separated groups. Its capacity to outline the clustering structure in the data has been tested in the literature, either in mixture models (Baudry *et al.*, 2008), latent block models (Keribin *et al.*, 2014) or SBMs (Mariadassou *et al.*, 2010).

In the SBM context, under standard asymptotic assumptions, the posterior distribution $p(\mathbf{Z}|\mathbf{X}^{1:K}, \hat{\theta}, \mathcal{M}_Q)$ concentrates on the true affectations (see Mariadassou and Matias (2015)). As a consequence, for n large, the entropy vanishes and ICL becomes a standard penalized maximum likelihood. We expect the same type of results in the multiplex SBM, but additional theoretical work is required.

2.3.1 Computing the integrated completed likelihood

Now, computing $E_{\mathbf{Z}|\mathbf{X}^{1:K}, \hat{\theta}, \mathcal{M}_Q}[\log\{l_c(\mathbf{X}^{1:K}, \mathbf{Z}; \hat{\theta}, \mathcal{M}_Q)\}]$ is the difficult part of the procedure in the SBMs or multiplex SBMs. Indeed, using the expression of the complete log-likelihood, the conditional expectation can be decomposed into

$$E_{\mathbf{Z}|\mathbf{X}^{1:K}; \hat{\theta}, \mathcal{M}_Q}[\log\{l_c(\mathbf{X}^{1:K}, \mathbf{Z}; \hat{\theta}, \mathcal{M}_Q)\}] = \sum_{i,j=1, i\neq j}^{n} \sum_{q,l=1}^{Q} E_{\mathbf{Z}|\mathbf{X}^{1:K}; \hat{\theta}, \mathcal{M}_Q}[\mathbb{1}_{Z_i=q, Z_j=l}]\log(\hat{\pi}_{ql}^{(X_{ij}^{1:K})})$$
$$+ \sum_{i=1}^{n} \sum_{q=1}^{Q} E_{\mathbf{Z}|\mathbf{X}^{1:K}; \hat{\theta}, \mathcal{M}_Q}[\mathbb{1}_{Z_i=q}]\log(\hat{\alpha}_q).$$

The posterior expectations of the $E_{\mathbf{Z}|\mathbf{X}^{1:K}; \hat{\theta}, \mathcal{M}_Q}[\mathbb{1}_{Z_i=q, Z_j=l}]$ have no explicit expression. For that reason, the EM algorithm could not be applied directly and we had to resort to the variational version of the EM algorithm (which is described in Appendix B) to maximize the likelihood. The variational EM algorithm proposed an approximation to the conditional distribution. This approximation can be used once again here: $p(\mathbf{Z}|\mathbf{X}^{1:K}; \hat{\theta}, \mathcal{M}_Q)$ is replaced by $\mathcal{R}_{\mathbf{X}^{1:K}}(\mathbf{Z})$. Using the notation that is introduced in Appendix B, we obtain

$$E_{\mathbf{Z}|\mathbf{X}^{1:K};\hat{\theta},\,\mathcal{M}_Q}[\mathbb{1}_{Z_i=q}] = \hat{\tau}_{iq},$$

$$E_{\mathbf{Z}|\mathbf{X}^{1:K};\hat{\theta},\,\mathcal{M}_Q}[\mathbb{1}_{Z_i=q,Z_j=l}] = \hat{\tau}_{iq}\hat{\tau}_{jl}$$

resulting in a tractable computation of the ICL criterion.

In Appendix C, we propose a simulation study to assess the performance of the ICL in our context.

Remark 4. Recently, Côme and Latouche (2014) and Wyse *et al.* (2014) considered a Bayesian context and proposed an exact version of the ICL criterion. In their context, the BIC- or asymptotic approximations are replaced by an exact computation of the integrated likelihood and they proposed an adequate algorithm to calculate the ICL value.

Remark 5. Exactly as before, if we consider covariates as in equation (3), then the ICL penalty should be adapted as follows:

$$\text{Pen}_{\text{ICL}}(\mathcal{M}_Q) = \tfrac{1}{2}\{P_Q\log\{Kn(n-1)\} + (Q-1)\log(n)\},$$

where P_Q is the dimension of $(\boldsymbol{\mu},\boldsymbol{\beta}) = (\mu_{ql}^{(w)}, \beta_{ql}^{(w)})_{q,l\in\{1,\dots,Q\}^2,w\in\{0,1\}^K}$.

3. Analysis for the laboratory–researcher data

3.1. The data

French scandals during the 1990s involving the voluntary sector around cancer research dried up large donations that funded research laboratories. In the 2000s, cancer research became politicized, with the launch of the 'Cancer plan' and the creation of a dedicated institution. The aim of this public agency is to co-ordinate cancer research and to promote collaboration around top researchers. In this context, Lazega *et al.* (2008) studied the relationships of advice between French cancer researchers who were identified as 'elite' conjointly with the relationships of their corresponding laboratories.

At the interindividual level, the actors (researchers) were sent a list of cancer researchers and asked in interviews whom they sought advice from. Two actors are linked if there is at least one kind of advice between them. The different types of advice are advice to deal with choices about the direction of projects, advice to find institutional support, advice to handle financial resources, advice for recruitment and finally advice about manuscripts before submitting them to journals. Note that the links are directed. At the laboratory level (concerning only laboratories with elite researchers), the laboratory directors were asked to specify what type of resources they exchanged with the other laboratories on the list. The resources examined were the recruitment of postdoctoral students and researchers, the development of programmes of joint research, joint responses to tender offers, sharing of technical equipment, sharing of experimental material, mobility of administrative personnel and invitations to conferences and seminars. Two laboratories are said to be linked if there is at least one link between them. From this network of laboratories, we can derive indirect links between the researchers, i.e. two researchers are connected if their laboratories exchange resources. We finally have two adjacency matrices on the same set of nodes (researchers).

This corresponds to transforming the multilevel network (individual and organization) into a multiplex network (several kinds of relationships between individuals). This is reasonable since the majority of laboratories contains a unique elite researcher. Thus, there is no big difference in the number of nodes between the institutional and individual levels. In addition, auxiliary

Table 1. (a) Size of blocks obtained by multiplex SBM, (b) cross-frequencies of blocks *versus* the laboratories' location (in Île-de-France (IDF) or not), (c) cross-frequencies of blocks *versus* the researcher's status (laboratory director or not) and (d) cross-frequencies of blocks *versus* the researcher's specialities†

	Block			Not IDF	IDF		Not director	Director
(a)		*(b)*				*(c)*		
1	2		1	1	1	1	1	1
2	48		2	26	22	2	21	27
3	19		3	10	9	3	12	7
4	26		4	11	15	4	12	14

	PH	Su	He	ST	Fph	FMo	FMoG
(d)							
1	0	0	0	0	0	1	1
2	12	5	6	10	7	7	1
3	0	1	3	1	1	10	3
4	6	1	7	4	2	2	4

†PH, public health; Su, surgery; He, haematology; ST, solid tumours; FPh, fundamental pharmacology; FM0, fundamental molecular research; FMoG, fundamental molecular genetic research.

covariates are available to describe the researchers: their age, their speciality, two publication performance scores based on two periods of 5 years, their status (director of the laboratory or not). Auxiliary covariates are also available for the laboratories: their size (number of researchers) and their location.

Remark 6. We do not aim at looking for heterogeneity in the multiplex network once the effect of the covariates has been taken into account—model (3)—but rather uncovering blocks based only on the observed connections—model (1)—and, then, understanding their composition through the covariates.

Remark 7. Another approach could be to consider a uniplex SBM network of researchers with the laboratories relationships as covariates. More precisely, on each edge, it could be possible to introduce a covariate y_{ij}^{L} such that $y_{ij}^{L} = 1$ if a relationship exists between the laboratory of researcher i and the laboratory of researcher j. However, it is unsatisfactory since it creates a disequilibrium between the two levels: analysing laboratories with researchers as covariate is not the same as analysing the researchers with laboratories as covariate.

Complete data for 95 researchers identified as the elite of French cancer research working in 76 laboratories are available.

3.2. Group inference for the multiplex researchers network

To estimate the parameters of the multiplex SBM model on this data set, we use a modified version of the variational EM algorithm (Daudin *et al.*, 2008) which is described in Appendix B. The optimization of the ICL criterion derived from likelihood (2) leads to four blocks (indistinctly denominated clusters or groups). For clarity, we index by R and L (rather than $\mathbf{X}^1$, $\mathbf{X}^2$) the two adjacency matrices (respectively the direct and indirect matrices).
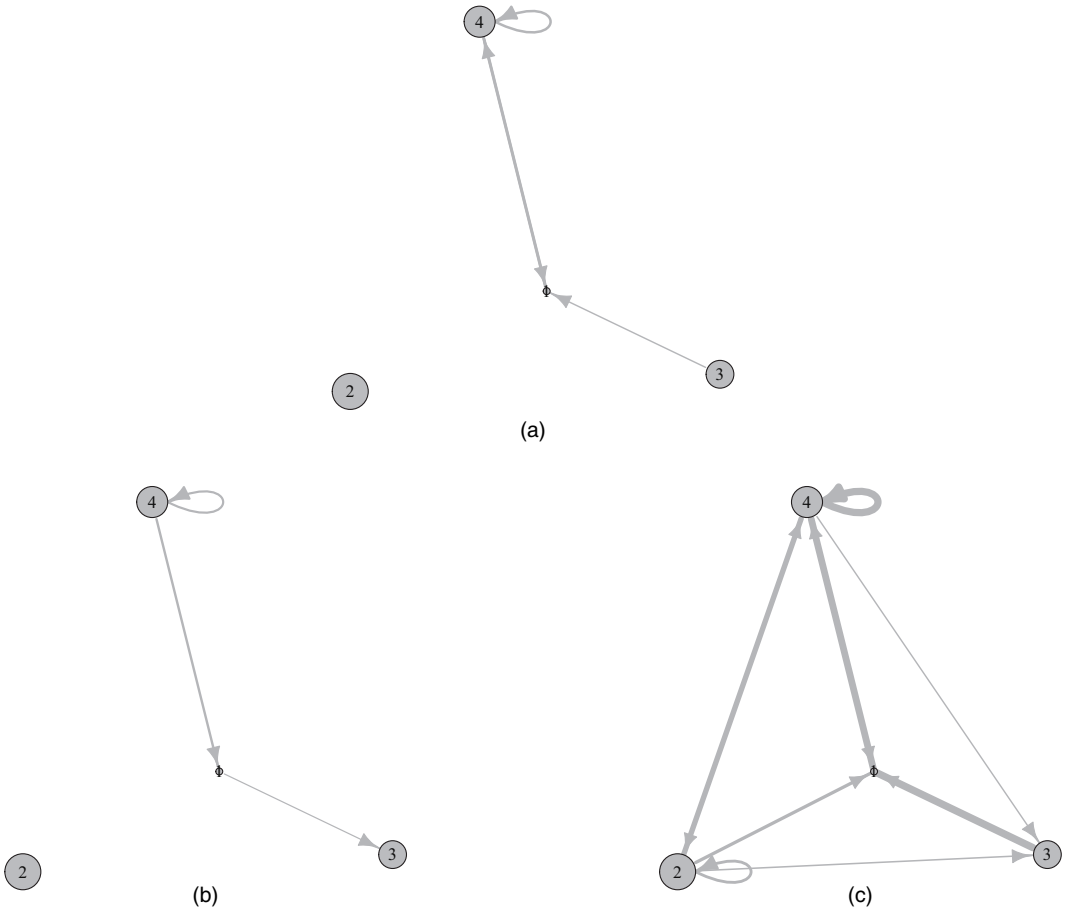
Fig. 1.   Marginal probabilities of researcher connections (a) between and within blocks ($P(R = 1)$), and probabilities of researcher connections between and within blocks conditionally on (b) absence ($P(R = 1|L = 0)$) or (c) presence ($P(R = 1|L = 1)$) of a laboratory connection: the size of the vertex is proportional to the size of the block; the width of the edge is proportional to the probabilities of connection; if this probability is smaller than 0.1, edges are not displayed

In Figs 1 and 2 we plot the marginal and conditional probabilities of the connections of researchers and laboratories respectively between and within blocks.

Note that the study of the estimated marginal distributions allows us to have results on the researchers without considering the laboratories. This gives a clear interpretation of the importance of the laboratory for the researcher network structure. The blocks obtained are described in Table 1: block (a) gives the sizes of the four blocks; blocks (b), (c) and (d) describe the blocks with respect to the covariates 'location', 'director or not' and 'speciality'. The estimations by the variational EM procedure were conducted with wmixnet (Leger, 2014) with our specific implementation of the bivariate Bernoulli model.

We now discuss the results. The multiplex SBM reveals interesting structural features of the multiplex network. More precisely, collaboration takes place in a clustered manner for both researchers and laboratories; collaborating laboratories tend to have affiliated researchers seeking advice from one another. Indeed, Fig. 1 shows that the existence of a connection (an exchange of resources) between laboratories clearly increases the probability of connection (sharing advice)
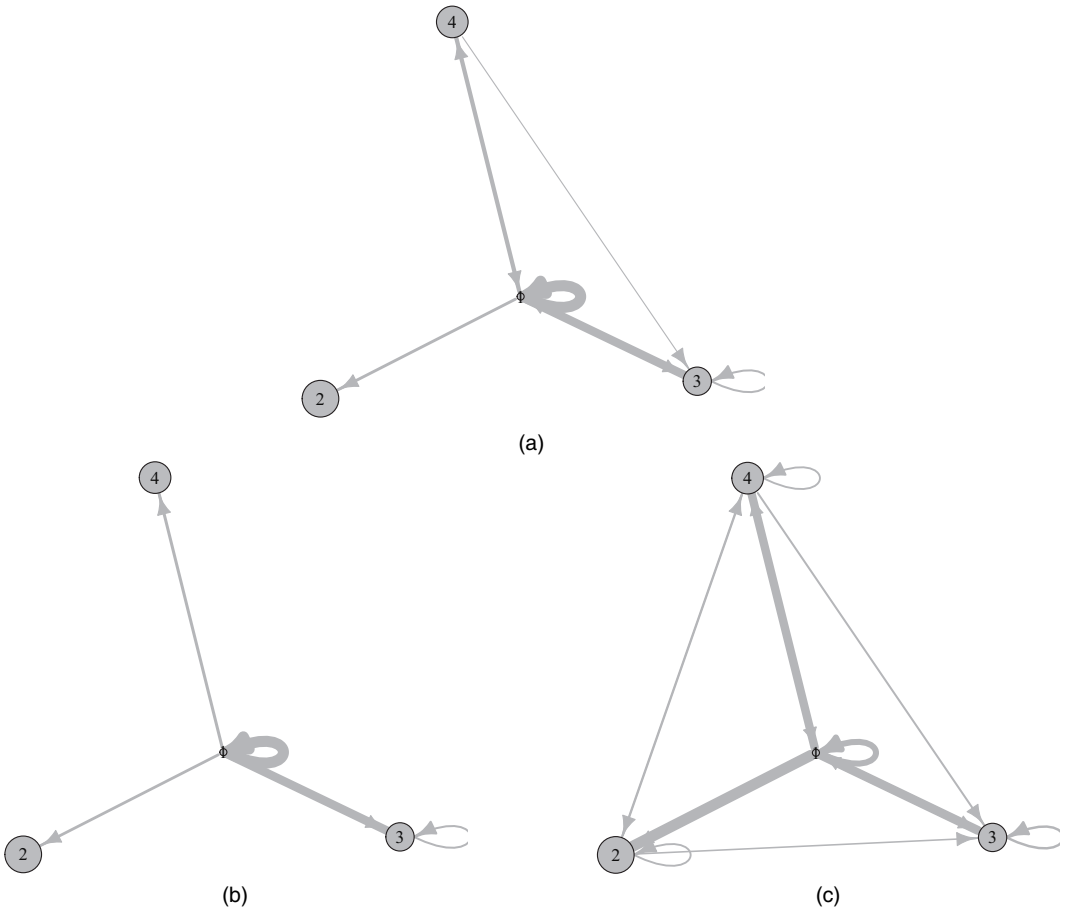
Fig. 2.   Marginal probabilities of laboratory connections (a) between and within blocks ($P(L = 1)$), and probabilities of laboratory connections between and within blocks conditionally on (b) absence ($P(L = 1|R = 0)$) or (c) presence ($P(L = 1|R = 1)$) of researcher connections: the size of the vertex is proportional to the size of the block; the width of the edge is proportional to the probabilities of connection; if this probability is smaller than 0.1, edges are not displayed

between researchers. The reinforcement of this probability of connection is clearly outstanding in block 2. In this block, the researcher connections are quite unlikely within the block or with other blocks. However, conditionally on the existence of a laboratory connection, the researcher connections become more important especially with block 4. In block 4, the links between researchers are strengthened given a connection between their laboratories. Researchers in block 3 seem to be the least affected by the connections that are provided by their laboratories. The case of block 1 is quite peculiar since it contains two researchers only. This clustering demonstrates that not all researchers benefit on equal terms from the institutional level. Some researchers are more dependent on their laboratories in terms of connections. Furthermore, Fig. 2 shows that the likeliest connections between laboratories are mainly with block 1. The fact that researchers are sharing advice inflates the probability of exchanging resources between laboratories.

More importantly, some complex features of the within-level network structure are explained mainly by cross-level interactions. In this empirical case, units of each level are clustered in

blocks that make sense from the perspective of the categories of people and organizations, at each level separately and together.

Block 1 members work in the biggest laboratories in terms of size (Fig. 3(b)). The SBM clusters them because they have many more relations than the other members of the network. They have among the highest indegrees, and average outdegrees in laboratories that have highest average indegrees and outdegrees (Figs 3(e), 3(f), 3(g) and 3(h)). They have the highest performance in terms of publication performance scores in both periods (Figs 3(c) and 3(d)). In fact, they have a similar relational profile since they are providers of transgenic mice for the experiments of many similar colleagues.

Block 2 members are among the lowest indegrees and outdegrees in laboratories that have the lowest average indegrees and outdegrees (Figs 3(e), 3(f), 3(g) and 3(h)), are slightly older than the others (Fig. 3(a)) and mainly in the smallest laboratories in terms of size. This block is heterogeneous in terms of specialities (especially 40% clinicians and 25% diagnostic, or pre-vention or epidemiology specialists) except fundamental research (Table 1, part (d)). They also have among the lowest performance levels for both periods (Figs 3(c) and 3(d)), although this is increasing. This is the biggest block. Their behaviour may be described as fusional as pro-posed in Lazega *et al.* (2008) since it corresponds to individuals for whom the probabilities of connection are the most affected by the connections of their laboratories.

In block 3, younger fundamental researchers in laboratories carrying out fundamental research (mostly molecular research) prevail (Fig. 3(a) and Table 1, part (d)). They have relatively low indegrees and outdegrees and are in laboratories that have the highest indegrees (after block 1) and average outdegrees (Figs 3(e), 3(f), 3(g) and 3(h)). 70% are among the top performers of this population, i.e. the highest performance levels after block 1 members, for both periods (Figs 3(c) and 3(d)). Their dominant relational strategies are individualist or in-dependentists (as shown in Fig. 1) since the probabilities of connection between researchers remain quite unchanged, no matter whether their laboratories are connected.

Block 4 is also heterogeneous in terms of specialities but its largest subgroup is composed of haematologists (clinical and fundamentalists) (Table 1, part (d)). Researchers have average indegrees and outdegrees in laboratories that have relatively low indegrees and outdegrees and that are also of average size (Figs 3(e), 3(f), 3(g) and 3(h)). There are proportionally more direc-tors of laboratories in this block than in the others (Table 1, part (c)). Their dominant strategy can be called fusional or collectivist. The performance levels of the majority for both periods are somewhat mixed and average, but decreasing (Figs 3(c) and 3(d)). In this block, researchers can take advantage of their laboratories to be connected with colleagues but they can also hold relationships apart from their laboratories. In this case, SBMs highlight in the data a specific kind of block structure that provides an interesting understanding of the effect of dual positioning. The SBM mixes people and laboratories that previous analyses used to separate. It is interesting, for example, that the SBM partitions the population regardless of geographic location (Table 1, part (b))—a criterion that was previously shown to be meaningful to understand collective ac-tion in this research *milieu*. In fact this partition may highlight the links between laboratories and between individuals that cut across the boundaries and remoteness that is created by geography, showing that certain categories of actors tend to reach across these separations when it suits them, either as investments to prepare future collaborations, or as a follow-up for past investments.

4. Discussion

The essence of 'networks' is to help actors to cut across organizational boundaries to create new relationships (Baker, 1992), to identify new opportunities and, eventually, to create new
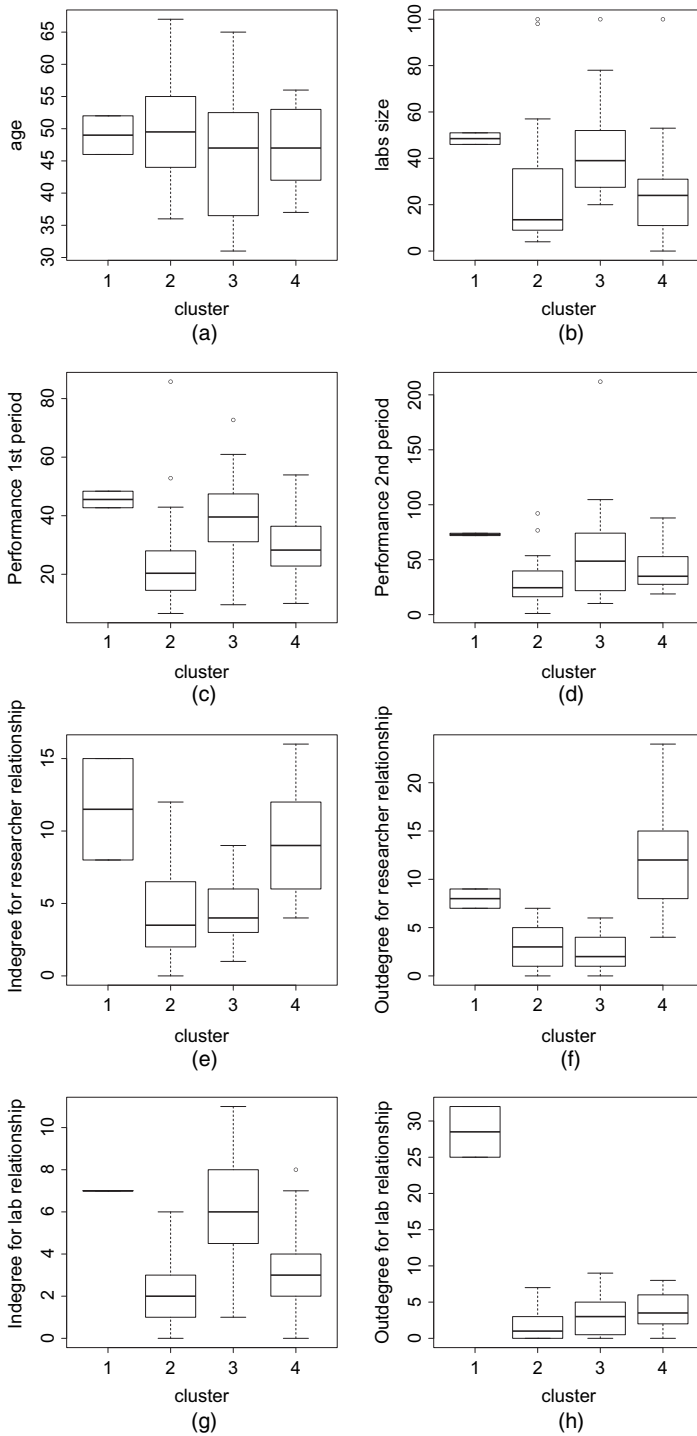
Fig. 3. Boxplots of (a) age, (b) size of laboratory, (c) performance in the first period, (d) performance in the second period, (e) indegree for researcher relationship, (f) outdegree for researcher relationship, (g) indegree for laboratory relationship and (h) outdegree for laboratory relationship, grouped by blocks

organizations to use or hoard these new opportunities (Lazega, 2012; Lazega *et al.*, 2013; Tilly, 1998). This work is a step forward to a more precise comprehension of such mechanisms.

In this paper, we proposed an SBM for multiplex networks. This way of partitioning the graph and calculating the probabilities that two individuals are connected takes into account more parameters in the graph than just centrality scores and size as in Lazega *et al.* (2007). Besides, compared with multilevel exponential random graph models (Wang *et al.*, 2013), this method takes inhomogeneities of actors more into account because the separate blocks tend to make sense as blocks with a specific identity. Multiplex SBM statistical analysis designed for the analysis of multilevel networks given as 'superposed' (Lazega *et al.*, 2007) network data aims at identifying the rules that govern the formation of links at the interindividual level conditioned by the characteristics of the interorganizational network. Our analysis found a general structure that shows the ways in which these rules can be derived in this case and help to assess actor inhomogeneities in terms of their influence on parameter estimates. This assessment is new in the sense that it shows that formation of connections in the interindividual network does not apply to all actors in the network identically.

In this paper, the model is applied to $K = 2$ networks. Applying with K much larger would imply difficulties in estimation since the number of parameters exponentially depends on K. However, particular assumptions (such as Markovian dependences) can reduce the difficulty and lead to feasible cases. Besides, in our example, going from multilevel to multiplex was quite natural since few laboratories contained more than one researcher. If the number of organizations were far smaller then the number of individuals then maybe it would be interesting to introduce another kind of relationship, specifying whether the individuals are in the same laboratory or not. In this case, the probabilities vector would have a special structure, taking this specificity into account (if the individuals belong to the same organization then they are automatically linked through their organization).

The consistency results that were obtained for the variational approach require an additional assumption for the parameters of the marginal distribution on the blocks memberships. We think that it could be possible to extend the consistency result for uniplex binary networks that was proved in Bickel *et al.* (2013) to the case of multiplex networks including asymptotic regimes with sparsity. However, this extension is far from trivial and would constitute a contribution *per se* mainly focused on theoretical aspects.

In this version of the analysis, the covariates were studied *a posteriori*, the classification being done purely on the network. From a practical point of view, including the covariates in the model would imply difficulty in estimation (and even more if the effect of the covariates on the connection probabilities depends on the blocks). However, even if attractive at first sight, the choice of including the covariates in the modelling must be questioned. Including them will cluster the data beyond the effect of these covariates, whereas not including them will provide a description of blocks on the basis of the relationship, and this relationship may be influenced by these covariates. This second strategy may lead to more interpretable results.

The first aim of Lazega *et al.* (2008) was cancer research management. Thus, the covariate 'performance in terms of publication' attracts more attention. Indeed, this modelization can generate other questions such as the influence of the networks on the performance of the actors (publications in our case). In this work, the performance is treated as a factor to explain the relationships. However, one could dream of a model which would help the actors to build the ideal network to optimize the performance. If it is true that contemporary society is an 'organizational society' (Coleman, 1982; Perrow, 1991; Presthus, 1962)—in the sense that action and performance measured at the individual level strongly depend on the capacity of the actor to construct and to use organizations as instruments, and thus to manage his or her interdepen-

dences at different levels in a strategic manner—then the study of interdependences jointly at the interindividual and the interorganizational level is important for numerous sets of problems. Proposing a hierarchical model in this direction is out of the scope of the paper but is clearly a possible extension.

To conclude, application to sociological network analysis will clearly benefit from this methodological contribution. In spite of several limitations that were listed before, this analysis of multiplex networks seems therefore adapted to certain types of question that sociologists ask when they try to combine both individual and contextual factors to estimate the likelihood that an individual or a group will adopt a given behaviour or to reach a given level of performance. More generally, this approach explores a complex mesosocial level of accumulation, of appropriation and of sharing of multiple resources. This level, which is still poorly known, is difficult to observe without a structural approach.

Appendix A: Identifiability of the multiplex stochastic block model and convergence of the maximum likelihood estimates

A.1. Identifiability

Celisse *et al.* (2012) have proved the identifiability of the parameters for uniplex Bernoulli SBMs. We extend their results to the multiplex SBM. First we recall the notation $\forall (i, j) \in \{1, \ldots, n\}^2$, $i \neq j$, $\forall (q, l) \in \{1, \ldots, Q\}^2$ and $\forall w \in \{0, 1\}^K$; we set

$$\pi_{ql}^{(w)} = P(X_{ij}^{1:K} = w | Z_i = q, Z_j = l),$$
$$\alpha_q = P(Z_i = q).$$

Let us introduce $\boldsymbol{\pi}^{(w)} = (\pi_{ql}^{(w)})_{(q, l) \in \{1, \ldots, Q\}}$ and $\boldsymbol{\pi} = (\boldsymbol{\pi}^{(w)})_{w \in \{0,1\}^K}$. $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_Q)$. Theorem 1 sets the identifiability of $\theta = (\boldsymbol{\alpha}, \boldsymbol{\pi})$ in the multiplex SBM.

Theorem 1.   Let $n \geqslant 2Q$. Assume that, for any $q \in \{1, \ldots, Q\}$, $\alpha_q > 0$ and, for every $w \in \{0, 1\}^K$, the co-ordinates of $\mathbf{r}^{(w)}(\theta) = \boldsymbol{\pi}^{(w)} \cdot \boldsymbol{\alpha}$ are distinct. Then, the multiplex SBM parameter $\theta = (\boldsymbol{\alpha}, \boldsymbol{\pi})$ is identifiable.

Proof.   The proof given by Celisse *et al.* (2012) can be directly extended to our case, leading to the expected result. As explained in Celisse *et al.* (2012), the identifiability can be proved by using algebraic tools. The only difference in the multiplex context is that the proof must be applied for any type of tie $w \in \{0, 1\}^K$.

For any $w \in \{0, 1\}^Q$, we set $r_q^{(w)}(\theta)$ the probability that a member of block q has a tie of type w with another individual: $r_q^{(w)}(\theta) = \Sigma_{l=1}^{Q_q} \pi_{ql}^{(w)} \alpha_l$. Let $R^{(w)}(\theta)$ be the Q-square matrix such that $R_{iq}^{(w)}(\theta) = \{r_q^{(w)}(\theta)\}^i$ for $i = 0, \ldots, Q - 1$ and $q = 1, \ldots, Q$. $R^{(w)}$ is a Vandermonde matrix, which is invertible by assumptions on the co-ordinates of $\mathbf{r}^{(w)}(\theta)$. Now, for any w and any $i = 0, \ldots, 2Q - 1$, we set

$$u_i^{(w)}(\theta) = P(X_{11}^{1:K} = \ldots = X_{1i}^{1:K} = w; \theta) = \sum_{q=1}^{Q} \alpha_q \{r_q^{(w)}(\theta)\}^i, \tag{7}$$

and $M^{(w)}(\theta)$ is a $(Q + 1) \times Q$ matrix such that

$$M_{ij}^{(w)}(\theta) = u_{i+j}^{(w)}(\theta), \qquad i = 0, \ldots, Q, \quad j = 0, \ldots, Q - 1. \tag{8}$$

For any $i = 0, \ldots, Q$ we define the Q-square matrix $M^{i(w)}(\theta)$ by removing line i from this matrix. In particular,

$$M^{Q(w)}(\theta) = R^{(w)}(\theta) A(\theta) R^{(w)}(\theta)^{\mathrm{T}}, \tag{9}$$

where $A(\theta)$ is the $\boldsymbol{\alpha}$-diagonal matrix. As all the α_q are non-null and $R^{(w)}(\theta)$ is invertible, then $\det(M^{Q(w)}) > 0$. So, if we define

$$B(X; \theta) = \sum_{i=0}^{Q} (-1)^{i+Q} \det\{M^{i(w)}(\theta)\} X^i.$$

B is of degree Q. Let us define $V^{i(w)}(\theta) = (1, r_i^{(w)}(\theta), \ldots, \{r_i^{(w)}(\theta)\}^Q)$; then

$$B\{r_i^{(w)}(\theta); \theta\} = \det(M^{(w)}(\theta), V_i^{(w)}(\theta)),$$

where $(M^{(w)}(\theta), V_i^{(w)}(\theta))$ is a $Q+1$ square matrix. As the columns of M are linear combinations of the $V_i^{(w)}$, we obtain $B\{r_i^{(w)}(\theta); \theta\} = 0$ for any $i = 0, \ldots, Q-1$. So, we can factorize B as

$$B(x; \theta) = \det\{M^{Q(w)}(\theta)\} \prod_{i=0}^{Q-1} \{x - r_i^{(w)}(\theta)\}. \tag{10}$$

Now, assume that $\theta = (\pi, \alpha)$ and $\theta' = (\pi', \alpha')$ are two sets or parameters such that, for any multiplex graph $\mathbf{X}^{1:K}$, $l(\mathbf{X}^{1:K}; \theta) = l(\mathbf{X}^{1:K}; \theta')$. Consequently, from equation (7), we obtain $u_i^{(w)}(\theta) = u_i^{(w)}(\theta')$ and, from equation (8), we deduce that $M^{i(w)}(\theta) = M^{i(w)}(\theta')$ for any $i = 0, \ldots, Q-1$. As the polynomial function B depends on the determinant of the $M^{i(w)}(\theta)$s, we have $B(\cdot; \theta) = B(\cdot; \theta')$, leading to $r_i(w)(\theta) = r_i(w)(\theta')$ for all $i = 0, \ldots, Q-1$, by equation (10). Thus $R^{(w)}(\theta) = R^{(w)}(\theta')$ and

$$A(\theta) = (R^{(w)}(\theta)^{\mathrm{T}})^{-1} M^{Q(w)}(\theta) R^{(w)}(\theta) = A(\theta').$$

As a consequence, $\alpha = \alpha'$.

Finally, let $U_{i,j}^{(w)}$ $(0 \geqslant i, j \geqslant Q)$ denote

$$U_{i,j}^{(w)} = P(X_{1,k}^{1:K} = w, k = 1, \ldots, i+1 \text{ and } k = n - j + 1, \ldots, n).$$

We can write the $Q \times Q$ matrix $U^{(w)}(\theta)$ as $U^{(w)}(\theta) = R^{(w)}(\theta) A(\theta) \pi^{(w)} A(\theta) R^{(w)}(\theta)^{\mathrm{T}}$. Using the fact that $R^{(w)}(\theta) = R^{(w)}(\theta')$ and $A(\theta) = A(\theta')$, we obtain

$$U^{(w)}(\theta) = U^{(w)}(\theta') \Rightarrow \pi^{(w)} = (\pi'^{(w)}), \qquad \forall w \in \{0, 1\}^K.$$

And the theorem is demonstrated.

A.2. Consistency of the maximum likelihood estimator in multiplex stochastic block models

We study the asymptotic properties of the maximum likelihood estimator in multiplex SBMs.

Assumption 1. For every $q \neq q'$, there exists $l \in \{1, \ldots, Q\}$ such that $\pi_{ql} \neq \pi_{q'l}$, or $\pi_{lq} \neq \pi_{lq'}$.

Assumption 2. There exists $\zeta > 0$ such that, $\forall (q, l) \in 1, \ldots, Q, \pi_{ql}^{(w)} \in]0, 1[\Rightarrow \pi_{ql}^{(w)} \in [\zeta, 1 - \zeta]$.

Assumption 3. There exists $\gamma \in (0, 1/Q)$ such that, $\forall q \in 1, \ldots, Q, \alpha_q \in [\gamma, 1 - \gamma]$.

Let $\mathbf{X}^{1:K}$ be a realization from the multiplex SBM: $\mathbf{X}^{1:K} \sim p(\mathbf{X}^{1:K} | \mathbf{Z}^*; \alpha^*, \pi^*)$ where $\mathbf{Z}^* = (Z_1^*, \ldots, Z_n^*)$ is the true group label sequence and (α^*, π^*) are the true parameters. Let $(\hat{\alpha}, \hat{\pi})$ be the maximum likelihood estimator defined as

$$(\hat{\alpha}, \hat{\pi}) = \arg\max_{(\alpha, \pi)} l(\mathbf{X}^{1:K}; \alpha, \pi),$$

where

$$l(\mathbf{X}^{1:K}; \alpha, \pi) = \sum_{\mathbf{Z} \in \{1, \ldots, Q\}^n} \exp[\log\{p(\mathbf{X}^{1:K} | \mathbf{Z}; \pi)\}] \prod_{i=1}^{n} \alpha_{Z_i},$$

$$\log\{p(\mathbf{X}^{1:K} | \mathbf{Z}; \pi)\} = \sum_{i,j,i \neq j}^{n} \log\{P(X_{ij}^{1:K} | Z_i, Z_j; \pi)\} = \sum_{i,j,i \neq j}^{n} \sum_{w \in \{0,1\}^K} \mathbb{I}_{X_{ij}^{1:K} = w} \log(\pi_{Z_i Z_j}^{(w)}).$$

Theorem 2. Let assumptions 1–3 hold. Then, for any distance $d(\cdot, \cdot)$ on the set of parameter π, we have

$$d(\hat{\pi}, \pi^*) \xrightarrow[n \to \infty]{\mathbb{P}} 0.$$

Moreover, assume that $\|\hat{\pi} - \pi^*\|_\infty = o_{\mathbb{P}}\{\sqrt{\log(n)/n}\}$; then, for any distance $d(\cdot, \cdot)$ in $\mathbb{R}^Q$,

$$d(\hat{\alpha}, \alpha^*) \xrightarrow[n \to \infty]{\mathbb{P}} 0.$$

Note that the rate $o_{\mathbb{P}}\{\sqrt{\log(n)/n}\}$ has not been proved yet. However, in the uniplex context, there is empirical evidence that the rate of convergence on $\hat{\pi}$ is $1/n$ (Gazal *et al.*, 2012). This result relies on a direct extension of the results given in Celisse *et al.* (2012) (see Appendix B.4).

Appendix B: Variational expectation–maximization algorithm for multiplex stochastic block models: principle, details and convergence

B.1. *General principle of the variational expectation–maximization*

SBMs belong to the incomplete-data models class, the non-observed data being the block indices $(Z_i)_{i=1,\ldots,n} \in \{1,\ldots,Q\}^n$. As written before, the likelihood has a marginal expression:

$$l(\mathbf{X}^{1:K}; \theta) = \sum_{\mathbf{Z} \in \{1,\ldots,Q\}^n} \exp[\log\{p(\mathbf{X}^{1:K}|\mathbf{Z}; \boldsymbol{\pi})\}] p(\mathbf{Z}; \boldsymbol{\alpha}), \qquad (11)$$

which is not tractable as soon as n and Q become large. The variational EM method is an alternative method to maximize the marginal likelihood with respect to θ. The variational EM algorithm (applied in the SBM context by Daudin *et al.* (2008)) relies on the following decomposition of equation (11). Let $\mathcal{R}_{\mathbf{X}^{1:K}}$ be any probability distribution on $\mathbf{Z}$; we have

$$\log\{l(\mathbf{X}^{1:K}; \theta)\} = \sum_{\mathbf{Z}} \mathcal{R}_{\mathbf{X}^{1:K}}(\mathbf{Z}) \log\{l_c(\mathbf{X}^{1:K}, \mathbf{Z}; \theta)\} - \sum_{\mathbf{Z}} \mathcal{R}_{\mathbf{X}^{1:K}}(\mathbf{Z}) \log\{\mathcal{R}_{\mathbf{X}^{1:K}}(\mathbf{Z})\}$$
$$+ \mathrm{KL}\{\mathcal{R}_{\mathbf{X}^{1:K}}, p(\cdot|\mathbf{X}^{1:K}; \theta)\}, \qquad (12)$$

where KL is the Kullback–Leibler distance, $l_c(\mathbf{X}^{1:K}, \mathbf{Z}; \theta)$ is the joint density of $\mathbf{X}^{1:K}$ and $\mathbf{Z}$ (namely the complete likelihood) and $p(\cdot|\mathbf{X}^{1:K}; \theta)$ is the posterior distribution of $\mathbf{Z}$ given the data $\mathbf{X}^{1:K}$ and the parameters θ. Instead of maximizing $\log\{l(\mathbf{X}^{1:K}; \theta)\}$, the variational EM algorithm optimizes a lower bound $\mathcal{I}_\theta(\mathcal{R}_{\mathbf{X}^{1:K}})$ of $\log\{l(\mathbf{X}^{1:K}; \theta)\}$ where

$$\mathcal{I}_\theta(\mathcal{R}_{\mathbf{X}^{1:K}}) = \log\{l(\mathbf{X}^{1:K}; \theta)\} - \mathrm{KL}\{\mathcal{R}_{\mathbf{X}^{1:K}}, p(\cdot|\mathbf{X}^{1:K}; \theta)\},$$
$$= \sum_{\mathbf{Z}} \mathcal{R}_{\mathbf{X}^{1:K}}(\mathbf{Z}) \log\{l_c(\mathbf{X}^{1:K}, \mathbf{Z}; \theta)\} - \sum_{\mathbf{Z}} \mathcal{R}_{\mathbf{X}^{1:K}}(\mathbf{Z}) \log\{\mathcal{R}_{\mathbf{X}^{1:K}}(\mathbf{Z})\} \qquad (13)$$
$$\leqslant \log\{l(\mathbf{X}^{1:K}; \theta)\}.$$

Thanks to equality (13), optimizing $\mathcal{I}_\theta(\mathcal{R}_{\mathbf{X}^{1:K}})$ with respect to θ no longer requires the computation of the marginal likelihood. Note also that the equality $\mathcal{I}_\theta(\mathcal{R}_{\mathbf{X}^{1:K}}) = \log\{l(\mathbf{X}^{1:K}; \theta)\}$ holds if and only if $\mathcal{R}_{\mathbf{X}^{1:K}} = p(\cdot|\mathbf{X}^{1:K}; \theta)$. As a consequence, $\mathcal{R}_{\mathbf{X}^{1:K}}$ will be taken as an approximation to $p(\cdot|\mathbf{X}^{1:K}; \theta)$ in a certain class of distributions. Jaakkola (2000) proposed to optimize it in the following class:

$$\mathcal{R}_{\mathbf{X}^{1:K}, \boldsymbol{\tau}}(\mathbf{Z}) = \prod_{i=1}^n h(Z_i, \hat{\boldsymbol{\tau}}_i),$$

where $h(\cdot; \boldsymbol{\tau}_i)$ is the multinomial distribution of parameter $\boldsymbol{\tau}_i = (\tau_{i1}, \ldots, \tau_{iq})$. Finally, the variational EM algorithm updates alternately θ and $\boldsymbol{\tau}$ in the following way. At iteration (t), given the current state $(\theta^{(t-1)}, \boldsymbol{\tau}^{(t-1)})$:

   step 1, compute

$$\boldsymbol{\tau}^{(t)} = \underset{\tau}{\arg\min}\, \mathrm{KL}\{\mathcal{R}_{\mathbf{X}^{1:K}, \tau}, p(\cdot|\mathbf{X}^{1:K}; \theta^{(t-1)})\} = \underset{\tau}{\arg\max}\, \mathcal{I}_{\theta^{(t-1)}}(\mathcal{R}_{\mathbf{X}^{1:K}, \tau});$$

   step 2, compute

$$\theta^{(t)} = \underset{\theta}{\arg\max}\, \mathcal{I}_{\theta^{(t)}}(\mathcal{R}_{\mathbf{X}^{1:K}, \tau^{(t)}}).$$

The details of steps 1 and 2 directly depend on the statistical model considered. For uniplex SBMs without covariates, they are given in Daudin *et al.* (2008). The details for the multiplex SBM are now given here.

B.2. *Details of the calculus for multiplex stochastic block models*

We now detail step 1 and step 2 for multiplex SBMs.

   Step 1: $\boldsymbol{\tau}^{(t)}$ verifies

$$\boldsymbol{\tau}^{(t)} = \underset{\tau}{\arg\min}\, \mathrm{KL}\{\mathcal{R}_{\mathbf{X}^{1:K}, \tau}, p(\cdot|\mathbf{X}^{1:K}; \theta^{(t-1)})\} = \underset{\tau}{\arg\max}\, \mathcal{I}_{\theta^{(t-1)}}(\mathcal{R}_{\mathbf{X}^{1:K}, \tau}).$$

We first rewrite $\mathcal{I}_\theta(\mathcal{R}_{\mathbf{X}^{1:K},\tau})$ for this special context:

$$\mathcal{I}_\theta(\mathcal{R}_{\mathbf{X}^{1:K},\tau}) = \sum_{\mathbf{Z}} \mathcal{R}_{\mathbf{X}^{1:K},\tau}(\mathbf{Z}) \log\{l_c(\mathbf{X}^{1:K}, \mathbf{Z}; \theta)\} - \sum_{\mathbf{Z}} \mathcal{R}_{\mathbf{X}^{1:K},\tau}(\mathbf{Z}) \log\{\mathcal{R}_{\mathbf{X}^{1:K},\tau}(\mathbf{Z})\},$$

with

$$\log\{l_c(\mathbf{X}^{1:K}, \mathbf{Z}; \theta)\} = \log\{p(\mathbf{X}^{1:K}|\mathbf{Z}; \theta)\} + \log\{p(\mathbf{Z}; \theta)\}$$

$$= \sum_{i,j,i\neq j} \log\{p(X_{ij}^{1:K}|Z_i, Z_j; \theta)\} + \sum_{i=1}^{n} \log(\alpha_{Z_i}).$$

This quantity must be integrated over $\mathbf{Z}$ where $\mathbf{Z} \sim \mathcal{R}_{\mathbf{X}^{1:K},\tau}$ which means that $\mathbf{Z} = (Z_i)_{i=1,\dots,n}$ are independent variables such that $P(Z_i = q) = \tau_{iq}$. We obtain

$$\mathcal{I}_\theta(\mathcal{R}_{\mathbf{X}^{1:K},\tau}) = \sum_{\mathbf{Z}} \mathcal{R}_{\mathbf{X}^{1:K},\tau}(\mathbf{Z}) \Big[\sum_{i,j,i\neq j} \log\{p(X_{ij}^{1:K}|Z_i, Z_j; \theta)\} + \sum_{i=1}^{n} \log(\alpha_{Z_i}) \Big] - \sum_{\mathbf{Z}} \mathcal{R}_{\mathbf{X}^{1:K},\tau}(\mathbf{Z}) \log\{\mathcal{R}_{\mathbf{X}^{1:K},\tau}(\mathbf{Z})\},$$

$$= \sum_{q,l} \sum_{i,j,i\neq j} \log\{p(X_{ij}^{1:K}|Z_i = q, Z_j = l; \theta)\}\tau_{iq}\tau_{jl} + \sum_{i=1}^{n}\sum_{q=1}^{Q} \tau_{iq}\log(\alpha_q) - \sum_{i=1}^{n}\sum_{q=1}^{Q} \tau_{iq}\log(\tau_{iq}),$$

where $\log\{p(X_{ij}^{1:K}|Z_i = q, Z_j = l; \theta)\}$'s expression is given in equation (1).

$\mathcal{I}_\theta(\mathcal{R}_{\mathbf{X}^{1:K},\tau})$ must be maximized with respect to τ under the constraint, $\forall i = 1,\dots,n$, $\Sigma_{q=1}^{Q} \tau_{iq} = 1$. As a consequence, we compute the derivatives of $\mathcal{I}_\theta(\mathcal{R}_{\mathbf{X}^{1:K},\tau}) + \Sigma_{i=1}^{n} \lambda_i(\Sigma_{q=1}^{Q}\tau_{iq} - 1)$ with respect to $(\lambda_i)_{i=1,\dots,n}$ and $(\tau_{iq})_{i=1,\dots,n,q=1,\dots,Q}$ where λ_i are the Lagrange multipliers, leading to the following collection of equations: for $i = 1,\dots,n$ and $q = 1,\dots,Q$,

$$\sum_{l}\sum_{j=1,j\neq i}^{n} \log\{p(X_{ij}^{1:K}|Z_i = q, Z_j = l; \theta)\}\tau_{jl} + \log(\alpha_q) - \log(\tau_{iq}) + 1 + \lambda_i = 0.$$

This leads to the following fixed point problem:

$$\hat{\tau}_{iq} = \exp(1 + \lambda_i)\alpha_q \prod_{j=1,j\neq i}^{n}\prod_{l=1}^{Q} p(X_{ij}^{1:K}|Z_i = q, Z_j = l; \theta)^{\hat{\tau}_{jl}}, \qquad \forall i = 1,\dots,n, \quad \forall q = 1,\dots,Q,$$

which must be solved under the constraints, $\forall i = 1,\dots,n$, $\Sigma_{q=1}^{Q}\tau_{iq} = 1$. This optimization problem is solved using a standard fixed point algorithm.

Step 2: compute $\theta^{(t)} = \arg\max_\theta \mathcal{I}_{\theta^{(t)}}(\mathcal{R}_{\mathbf{X}^{1:K},\tau^{(t)}})$.

Once the $\hat{\tau}$ have been optimized, the parameters θ maximizing $\mathcal{I}_\theta(\mathcal{R}_{\mathbf{X}^{1:K},\hat{\tau}})$ must be computed under the constraints $\Sigma_{q=1}^{Q} \alpha_q = 1$ and $\Sigma_{w\in\{0,1\}^L} \pi_{ql}^{(w)} = 1$ for all $(q,l) \in \{1,\dots,Q\}^2$.

The maximization with respect to $\boldsymbol{\alpha}$ is quite direct and, in any case, we obtain

$$\hat{\alpha}_q = \frac{1}{n}\sum_{i=1}^{n}\hat{\tau}_{iq},$$

$$\hat{\pi}_{ql}^{(w)} = \frac{\sum_{ij}\hat{\tau}_{iq}\hat{\tau}_{jl}\mathbb{1}_{X_{ij}^{1:K}=w}}{\sum_{ij}\hat{\tau}_{iq}\hat{\tau}_{jl}}.$$

Remark 8. If the edge probabilities depend on covariates

$$\mathrm{logit}(\pi_{ql}^{(w)}) = \mu_{ql}^{(w)} + (\beta_{ql}^{(w)})^{\mathrm{T}}\mathbf{y}_{ij},$$

then the optimization of $(\mu_{ql}^{(w)})$ and $(\beta_{ql}^{(w)})$ at step 2 of the variational EM algorithm is not explicit anymore and one should resort to optimization algorithms such as the Newton–Raphson algorithm.

B.3.   *Convergence of the variational expectation–maximization estimates*

We now consider the consistency of the estimates that are obtained by the variational EM algorithm. Using the variational EM algorithm that was previously described is equivalent to maximizing the so-called *variational likelihood* $\mathcal{I}$ where

$$\mathcal{I}(\mathbf{X}^{1:K};\boldsymbol{\tau},\boldsymbol{\alpha},\boldsymbol{\pi}) = \sum_{i\neq j}\tau_{iq}\tau_{jl}\sum_{w\in\{0,1\}^K}\mathbb{1}_{X_{ij}=w}\log(\pi_{ql}^{(w)}) - \sum_{iq}\tau_{iq}\{\log(\tau_{iq})-\log(\alpha_q)\}.$$

The variational estimators are obtained by

$$\tilde{\boldsymbol{\tau}}(\boldsymbol{\alpha},\boldsymbol{\pi}) = \arg\max_{\boldsymbol{\tau}}\mathcal{I}(\mathbf{X}^{1:K};\boldsymbol{\tau},\boldsymbol{\alpha},\boldsymbol{\pi}),$$

$$(\tilde{\boldsymbol{\alpha}},\tilde{\boldsymbol{\pi}}) = \arg\max_{\boldsymbol{\alpha},\boldsymbol{\pi}}\mathcal{I}(\mathbf{X}^{1:K};\tilde{\boldsymbol{\tau}},\boldsymbol{\alpha},\boldsymbol{\pi}).$$

Theorem 3. Assume that assumptions 1–3 hold. Then, for any distance on the set of parameters $\boldsymbol{\pi}$,

$$d(\tilde{\boldsymbol{\pi}},\boldsymbol{\pi}^*) \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Moreover, assume that $d(\tilde{\boldsymbol{\pi}},\boldsymbol{\pi}^*)=o_{\mathrm{P}}(1/n)$; then, for any distance on $\mathbb{R}^Q$,

$$d(\tilde{\boldsymbol{\alpha}},\boldsymbol{\alpha}^*) \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

B.4. About the proofs of theorems 2 and 3

The proofs of theorems 2 and 3 require many intermediate results which will not be given here because their adaptation to the multiplex context is quite direct. Indeed, at any step of the proof, the original Bernoulli distribution needs to be replaced by its K-dimensional version. More precisely, a central quantity in the proof is the ratio $\log\{p(\mathbf{Z}|\mathbf{X}^{1:K};\boldsymbol{\alpha},\boldsymbol{\pi})/p(\mathbf{Z}^*|\mathbf{X}^{1:K};\boldsymbol{\alpha},\boldsymbol{\pi})\}$. In the uniplex case, this quantity is

$$\log\left\{\frac{p(\mathbf{Z}|\mathbf{X}^{1:K};\boldsymbol{\alpha},\boldsymbol{\pi})}{p(\mathbf{Z}^*|\mathbf{X}^{1:K};\boldsymbol{\alpha},\boldsymbol{\pi})}\right\} = \log\left\{\frac{p(\mathbf{X}^{1:K}|\mathbf{Z};\boldsymbol{\pi})}{p(\mathbf{X}^{1:K}|\mathbf{Z}^*;\boldsymbol{\pi})}\right\} + \log\left\{\frac{p(\mathbf{Z};\boldsymbol{\alpha})}{p(\mathbf{Z}^*;\boldsymbol{\alpha})}\right\}$$

$$= \sum_{i\neq j}X_{ij}\log\left(\frac{\pi_{Z_i,Z_j}^{(w)}}{\pi_{z_i^*,z_j^*}^{(w)}}\right) + (1-X_{ij})\log\left(\frac{1-\pi_{Z_i,Z_j}^{(w)}}{1-\pi_{z_i^*,z_j^*}^{(w)}}\right) + \sum_{i=1}^{n}\log\left(\frac{\alpha_{Z_i}}{\alpha_{z_i^*}}\right).$$

In the multiplex case, the sum of two terms is replaced by a sum of 2^K terms:

$$\log\left\{\frac{p(\mathbf{Z}|\mathbf{X};\boldsymbol{\alpha},\boldsymbol{\pi})}{p(\mathbf{Z}^*|\mathbf{X};\boldsymbol{\alpha},\boldsymbol{\pi})}\right\} = \sum_{i\neq j}\sum_{w\in\{0,1\}^K}\mathbb{1}_{\{X_{ij}=w\}}\log\left(\frac{\pi_{Z_i,Z_j}^{(w)}}{\pi_{Z_i^*,Z_j^*}^{(w)}}\right) + \sum_{i=1}^{n}\log\left(\frac{\alpha_{Z_i}}{\alpha_{z_i^*}}\right).$$

where $\sum_{w\in\{0,1\}^K}\pi_{ql}^{(w)}=1$, for any (q,l). Going from two terms to 2^K terms does not imply any mathematical difficulty and so does not compromise the convergence results.

Appendix C: About the integrated completed likelihood criterion: numerical simulations

1000 networks were simulated according to a multiplex SBM with $n=95$ nodes, four groups and parameters fixed to the estimates that were obtained in Section 3.2. Table 2 displays the number of blocks Q selected by

Table 2. Number of blocks selected by the ICL criterion and adjusted Rand index

Estimated Q	ARI
3 for 178 data sets over 1000 4 for 822 data sets over 1000	Minimum: 0.4210 1st quartile: 0.9398 Median: 0.9646 Mean: 0.9393 3rd quartile: 1.0000 Maximum: 1.0000

the ICL criterion and the adjusted Rand index ARI (Hubert and Arabie, 1985) for measuring the adequacy of the clustering.

These simulations illustrate the efficiency of the ICL criterion for selecting the correct number of blocks and for recovering the latent block structure.

References

Ambroise, C. and Matias, C. (2012) New consistent and asymptotically normal parameter estimates for random-graph mixture models. *J. R. Statist. Soc.* B, **74**, 3–35.

Baker, W. E. (1992) The network organization in theory and practice. *Netwrks Organizns Struct. Form Action*, 396–429.

Baudry, J.-P., Celeux, G. and Marin, J.-M. (2008) Selecting models focussing on the modeller's purpose. In *COMPSTAT 2008* (ed. P. Brito), pp. 337–348. Heidelberg: Physica.

Bianconi, G. (2013) Statistical mechanics of multiplex networks: entropy and overlap. *Phys. Rev.* E, **87**, article 062806.

Bickel, P. and Chen, A. (2009) A nonparametric view of network models and Newman-Girvan and other modularities. *Proc. Natn. Acad. Sci. USA*, **106**, 21068–21073.

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

Biernacki, C., Celeux, G. and Govaert, G. (2000) Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Trans. Pattn Anal. Mach. Intell.*, **22**, 719–725.

Brummitt, C. D., D'Souza, R. M. and Leicht, E. (2012) Suppressing cascades of load in interdependent networks. *Proc. Natn. Acad. Sci. USA*, **109**, E680–E689.

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

Coleman, J. S. (1982) *The Asymmetric Society*. Syracuse: Syracuse University Press.

Côme, E. and Latouche, P. (2014) Model selection and clustering in stochastic block models with the exact integrated complete data likelihood. *Preprint arXiv:1303.2962v2*.

Cozzo, E., Arenas, A. and Moreno, Y. (2012) Stability of boolean multilevel networks. *Phys. Rev.* E, **86**, article 036115.

Daudin, J. J., Picard, F. and Robin, S. (2008) A mixture model for random graphs. *Statist. Comput.*, **18**, 173–183.

Gazal, S., Daudin, J.-J. and Robin, S. (2012) Accuracy of variational estimates for random graph mixture models. *J. Statist. Computn Simuln*, **82**, 849–862.

Goldenberg, A., Zheng, A. X., Fienberg, S. E. and Airoldi, E. M. (2010) A survey of statistical network models. *Found. Trends Mach. Learn.*, **2**, 129–233.

Holland, P. W., Laskey, K. B. and Leinhardt, S. (1983) Stochastic blockmodels: first steps. *Socl Netwrks*, **5**, 109–137.

Hubert, L. and Arabie, P. (1985) Comparing partitions. *J. Classificn*, **2**, 193–218.

Jaakkola, T. S. (2000) Tutorial on variational approximation methods. In *Advanced Mean Field Methods: Theory and Practice*, pp. 129–159. Cambridge: MIT Press.

Keribin, C., Brault, V., Celeux, G. and Govaert, G. (2014) Estimation and selection for the latent block model on categorical data. *Statist. Comput.*, **25**, no. 6, 1–16.

Komlós, J. and Simonovits, M. (1996) Szemerédi's regularity lemma and its applications in graph theory. *Technical Report*. Rutgers University, Piscataway.

Kunszenti-Kovács, D., Lovász, L. and Szegedy, B. (2014) Multigraph limits, unbounded kernels, and banach space decorated graphs. *Preprint arXiv:1406.7846*.

Lazega, E. (2012) Analyses de réseaux et classes sociales. *Rev. Fr. Socio-écon.*, **2**, 273–279.

Lazega, E., Jourda, M.-T. and Mounier, L. (2007) Des poissons et des mares: l'analyse de réseaux multi-niveaux. *Rev. Fr. Sociol.*, **48**, 93–131.

Lazega, E., Jourda, M.-T. and Mounier, L. (2013) Network lift from dual alters: extended opportunity structures from a multilevel and structural perspective. *Eur. Sociol. Rev.*, **29**, 1226–1238.

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

Leger, J.-B. (2014) Wmixnet: software for clustering the nodes of binary and valued graphs using the stochastic block model. *Preprint arXiv:1402.3410*. AgroParisTech, Paris.

Loe, C. W. and Jeldtoft Jensen, A. H. (2014) Comparison of communities detection algorithms for multiplex. *Preprint arXiv:1406.2205*. Imperial College London, London.

Maggioni, M. A., Breschi, S. and Panzarasa, P. (2013) Multiplexity, growth mechanisms and structural variety in scientific collaboration networks. *Indust. Innovn*, **20**, 185–194.

Mariadassou, M. and Matias, C. (2015) Convergence of the groups posterior distribution in latent or stochastic block models. *Bernoulli*, **21**, 537–573.

Mariadassou, M., Robin, S. and Vacher, C. (2010) Uncovering latent structure in valued graphs: a variational approach. *Ann. Appl. Statist.*, **4**, 715–742.

Matias, C. and Robin, S. (2014a) Modeling heterogeneity in random graphs through latent space models: a selective review. *ESAIM Proc. Surv.*, **47**, 55–74.

Matias, C. and Robin, S. (2014b) Modeling heterogeneity in random graphs through latent space models: a selective review. *Preprint arXiv:1402.4296v2*.

Mucha, P. J., Richardson, T., Macon, K., Porter, M. A. and Onnela, J.-P. (2010) Community structure in time-dependent, multiscale, and multiplex networks. *Science*, **328**, 876–878.

Nicosia, V., Bianconi, G., Latora, V. and Barthelemy, M. (2013) Growing multiplex networks. *Phys. Rev. Lett.*, **111**, article 058701.

Nowicki, K. and Snijders, T. A. B. (2001) Estimation and prediction for stochastic blockstructures. *J. Am. Statist. Ass.*, **96**, 1077–1087.

Perrow, C. (1991) A society of organizations. *Theor. Soc.*, **20**, 725–762.

Presthus, R. (1962) *The Organizational Society: an Analysis and a Theory*. New York: Vintage Books–Random House.

Rank, O. N., Robins, G. L. and Pattison, P. E. (2010) Structural logic of intraorganizational networks. *Organizn Sci.*, **21**, 745–764.

Saumell-Mendiola, A., Serrano, M. Á. and Boguñá, M. (2012) Epidemic spreading on interconnected networks. *Phys. Rev.* E, **86**, article 026106.

Snijders, T. A. and Baerveldt, C. (2003) A multilevel network study of the effects of delinquent behavior on friendship evolution. *J. Math. Sociol.*, **27**, 123–151.

Snijders, T. A. B. and Nowicki, K. (1997) Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *J. Classificn*, **14**, 75–100.

Szell, M., Lambiotte, R. and Thurner, S. (2010) Multirelational organization of large-scale social networks in an online world. *Proc. Natn. Acad. Sci. USA*, **107**, 13636–13641.

Tilly, C. (1998) *Durable Inequality*. Berkeley: University of California Press.

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

Wasserman, S. (1994) *Social Network Analysis: Methods and Applications*, vol. 8. Cambridge: Cambridge University Press.

Wyse, J., Friel, N. and Latouche, P. (2014) Inferring structure in bipartite networks using the latent block model and exact ICL. *Preprint arXiv:1404.2911*.