

# Stochastic Block Models for Multiplex networks: an application to networks of researchers

Pierre Barbillon <sup>\*1</sup>, Sophie Donnet<sup>1</sup>, Emmanuel Lazega<sup>2</sup>, and Avner Bar-Hen<sup>3</sup>

<sup>1</sup>AgroParisTech / UMR INRA MIA 518, 16 rue Claude Bernard, 75231 Paris Cedex 05, France

<sup>2</sup>Institut d'Études Politiques de Paris (Sciences Po), Département de Sociologie, Centre de Sociologie des Organisations, 19 rue Amélie, 75007 Paris, France

<sup>3</sup>MAP5, UFR de Mathématiques et Informatique Université Paris Descartes 45 rue des Saints-Pères 75270 Paris cedex 06

## Abstract

Modeling relations between individuals is a classical question in social sciences and clustering individuals according to the observed patterns of interactions allows to uncover a latent structure in the data. Stochastic block model (SBM) is a popular approach for grouping the individuals with respect to their social comportment. When several relationships of various types can occur jointly between the individuals, the data are represented by multiplex networks where more than one edge can exist between the nodes. In this paper, we extend the SBM to multiplex networks in order 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–

---

<sup>\*</sup>pierre.barbillon@agroparistech.fr

through a variational Expectation-Maximization procedure. Consistency of the estimates as well as statistical properties of the model are obtained. The number of groups is chosen thanks to the Integrated Completed Likelihood criteria, a penalized likelihood criterion. Multiplex Stochastic Block Model arises 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 labs. Our results show strong interactions between these two kinds of connections and the groups that are obtained are discussed to emphasize the common features of researchers grouped together.

**Keywords**— Bivariate Stochastic Block Model, Multilevel / 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 relations 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, 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 \(2013\)](#); [Nicosia et al. \(2013\)](#)). In these multiplex networks, different kinds of links (or connections) 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, political groups, etc.), with some individuals possibly belonging to the same institution. While the actors will exchange resources (such as advice for instance) 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 inter-organizational context of their actions) would allow us to identify the individuals that 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 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 relations 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 links 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 able to 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 observed in networks. These models allow to derive properties of the network at the macro-scale and to understand the outcomes of interactions. To explore heterogeneity at others scales (such as micro or meso-scale) in social networks, specific models such as the stochastic block models (SBM) ([Snijders and Nowicki \(1997\)](#)) have been developed for uniplex networks. In this paper, we propose an original extension of the SBMs to the multiplex case. Our model is efficient to model 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 using an extension of the variational EM algorithm. Consistency of the estimation of the parameters is proved. As for uniplex SBM, a key issue is to choose the number of blocks. We use a penalized likelihood criterion, namely Integrated Completed Likelihood (ICL). The inference procedure is performed on the cancer researchers / laboratories dataset.

The paper is organized as follows. The extension of SBM to multiplex network is presented in Section 2, the proofs of model identifiability and the consistency of variational EM procedure are postponed in Appendices A and B. In Section 3, we describe [Lazega et al. \(2008\)](#)'s dataset, apply the new modeling and discuss the results. Eventually, the contribution of multiplex SBM to the analysis of multiplex networks is highlighted in Section 4.

## 2 Multiplex stochastic block model

The main objective is to cluster the individuals (or nodes) into blocks sharing connection properties with the other individuals of the multiplex-network. Stochastic block models ([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 the following, we propose an extension of the Stochastic Block Model (SBM) to multiplex networks. The SBM for multiplex networks is derived from a multiplex Erdős-Rényi model which is described in subsection 2.1. The SBM for multiplex networks is derived in subsection 2.2.

## 2.1 Erdős-Rényi model for multiplex networks

Let  $\mathbf{X}^1, \dots, \mathbf{X}^K$  be  $K$  directed graphs relying on the same set of nodes  $E = \{1, \dots, n\}$ . We assume that  $\forall(i, j), i \neq j, \forall k \in \{1, \dots, K\}, X_{ij}^k \in \{0, 1\}$  and  $X_{ii} \neq 0$ . We define a joint distribution on  $\mathbf{X}^{1:K} = (\mathbf{X}^1, \dots, \mathbf{X}^K)$  as:  $\forall(i, j) \in \{1, \dots, n\}^2, i \neq j, \forall w \in \{0, 1\}^K$ ,

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

and  $(X_{ij}^{1:K})_{i,j}$  are mutually independent.

The maximum likelihood estimate of the parameter of interest  $\boldsymbol{\pi} = (\pi^{(w)})_{w \in \{0,1\}^K}$  is, for all  $w \in \{0, 1\}^K$ :

$$\hat{\pi}_w = \frac{1}{n(n-1)} \sum_{i,j, i \neq j} \mathbb{I}_{\{X_{ij}^{1:K} = w\}}.$$

This model is quite simple since any relation between two individuals (a relation being a collection of edges) does not depend on the relations between the other individuals. However, the different kind of relations between two individuals (edges) are not assumed to be independent.

**Remark 1.** This model is clearly an extension of the Erdős-Rényi model since the marginal distribution of  $X_{ij}^k$  (for any  $k = 1 \dots K$ ) is Bernoulli with density :

$$\mathbb{P}(X_{ij}^k = x_{ij}^k) = \left( \sum_{w \in \{0,1\} \setminus \{w_k=1\}} \pi^{(w)} \right)^{x_{ij}^k} \left( \sum_{w \in \{0,1\} \setminus \{w_k=0\}} \pi^{(w)} \right)^{1-x_{ij}^k}.$$

Moreover, any conditional distribution of  $X_{ij}^k$  given  $(X_{ij}^l)_{l \in \mathcal{S}_{\setminus k}}$  (where  $\mathcal{S}_{\setminus k}$  is a subset of

$\{1, \dots, K\}$  not containing  $k$ ) is also univariate Bernoulli. For instance, if  $K = 2$  the conditional distribution of  $X_{ij}^1$  given  $X_{ij}^2$  is

$$\mathbb{P}(X_{ij}^1 = x_{ij}^1 | X_{ij}^2 = x_{ij}^2) = \left( \frac{\pi^{(1, x_{ij}^2)}}{\pi^{(1, x_{ij}^2)} + \pi^{(0, x_{ij}^2)}} \right)^{x_{ij}^1} \left( \frac{\pi^{(0, x_{ij}^2)}}{\pi^{(1, x_{ij}^2)} + \pi^{(0, x_{ij}^2)}} \right)^{(1-x_{ij}^1)}.$$

Moreover, the components of the bivariate Bernoulli random vector  $(X_{ij}^1, X_{ij}^2)$  are independent if and only if  $\pi^{(00)}\pi^{(11)} = \pi^{(10)}\pi^{(01)}$ .

**Introduction of explanatory variables.** Naturally, the Erdős-Rényi model for multiplex networks can be extended to take into account explanatory variables. Let  $\mathbf{y}_{ij}$  denote the covariates characterizing the couple of nodes  $(i, j)$ , the model is defined by the probabilities:

$$\begin{aligned} \mathbb{P}(X_{ij}^{1:K} = w) &= \frac{\exp(\mu^{(w)} + (\beta^{(w)})^\top \mathbf{y}_{ij})}{1 + \sum_{v \neq (0, \dots, 0)} \exp(\mu^{(v)} + (\beta^{(v)})^\top \mathbf{y}_{ij})} \quad \forall w \neq (0, \dots, 0), \\ \mathbb{P}(X_{ij}^{1:K} = (0, \dots, 0)) &= \frac{1}{1 + \sum_{v \neq (0, \dots, 0)} \exp(\mu^{(v)} + (\beta^{(v)})^\top \mathbf{y}_{ij})}, \end{aligned} \quad (2)$$

where  $x^\top$  denotes the transposed vector of  $x$ .

**Remark 2.** Note that in the multiplex Erdős-Rényi model, the modeling 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. As a consequence, the covariates  $\mathbf{y}_{ij}$  only depend on the couple  $(i, j)$  and are not linked to the network under consideration.

Since the multiplex Erdős-Rényi model belongs to exponential models, the generalized linear model theory applies when we introduce the covariates as in model (2) and the estimates are obtained using standard optimization strategies.

## 2.2 Stochastic block model for multiplex networks

When the goal is to cluster individuals according to their social comportment, we can derive a Stochastic Block Model version of the multiplex Erdős-Rényi model. Let  $Q$  be

the number of blocks and  $Z_i$  the latent variable such that  $Z_i = q$  if the individual  $i$  belongs to block  $q$  (note that an individual can only belong to one block in this version).

The multiplex version of SBM is written as follows:  $\forall(i,j) \in \{1, \dots, n\}^2, i \neq j$ ,  $\forall w \in \{0, 1\}^K, \forall(q, l) \in \{1, \dots, Q\}^2$ ,

$$\begin{aligned}\mathbb{P}(X_{ij}^{1:K} = w | Z_i = q, Z_j = l) &= \pi_{ql}^{(w)} \\ P(Z_i = q) &= \alpha_q.\end{aligned}\tag{3}$$

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

$$\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_Q), \quad \boldsymbol{\pi} = (\pi_{ql}^{(w)})_{w \in \{0,1\}^K, (q,l) \in \{1, \dots, Q\}^2}, \quad \theta = (\boldsymbol{\alpha}, \boldsymbol{\pi}),$$

the likelihood function is written as:

$$\begin{aligned}\ell(\mathbf{X}^{1:K}; \theta) &= \int_{\mathbf{Z} \in \{1, \dots, Q\}^n} p(\mathbf{X}^{1:K} | \mathbf{Z}; \boldsymbol{\pi}) p(\mathbf{Z}; \boldsymbol{\alpha}) d\mathbf{Z}, \\ &= \sum_{\mathbf{Z} \in \{1, \dots, Q\}^n} \prod_{i,j,i \neq j} \pi_{Z_i Z_j}^{(X_{ij}^{1:K})} \prod_{i=1}^n \alpha_{Z_i},\end{aligned}\tag{4}$$

where the latent variable (the block affectations) are integrated out. The identifiability of the model can be proved (see Appendix A, theorem A.1) and the maximum likelihood estimators are consistent (theorem A.2).

**Remark 3.** Note that, as before, covariates on the couple  $(i, j)$  can be introduced in the model:  $\forall w \neq (0, \dots, 0)$ ,

$$\begin{aligned}\mathbb{P}(X_{ij}^{1:K} = w | Z_i = q, Z_j = l) &= \frac{\exp(\mu_{ql}^{(w)} + (\beta_{ql}^{(w)})^\top \mathbf{y}_{ij})}{1 + \sum_{v \neq (0, \dots, 0)} \exp(\mu_{ql}^{(v)} + (\beta_{ql}^{(v)})^\top \mathbf{y}_{ij})}, \\ \mathbb{P}(X_{ij}^{1:K} = (0, \dots, 0) | Z_i = q, Z_j = l) &= \frac{1}{1 + \sum_{v \neq (0, \dots, 0)} \exp(\mu_{ql}^{(v)} + (\beta_{ql}^{(v)})^\top \mathbf{y}_{ij})}.\end{aligned}\tag{5}$$

However, the number of parameters in this new model will increase drastically, leading to estimation issues.

**Maximum likelihood and model selection** As soon as  $n$  or  $Q$  are large, the observed likelihood (4) is not tractable (due to 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, 2014](#)), both in the frequentist and 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 \(2009\)](#) which relying on a profile-likelihood optimization or the moment estimation proposed by [Ambroise and Matias \(2012\)](#), to name but a few .

The variational EM in the context of SBM proposed by [Daudin et al. \(2008\)](#) is a flexible tool to tackle the computational challenge in many types of graphs. Simulation studies showed its practical efficiency ([Mariadassou et al., 2010](#)). Moreover, its theoretical convergence towards the maximum likelihood estimates has been studied by [Celisse et al. \(2012\)](#) for binary graphs. In this paper, we adapt the variational EM to multiplex networks. The algorithm is described in Appendix B and its convergence towards the true parameter is proved (Theorem B.1).

The selection of the most adequate number of blocks  $Q$  is performed using a modification of the ICL criterion (as in [Daudin et al., 2008; Mariadassou et al., 2010](#)). Let  $\mathcal{M}_Q$  denote the model defined (5) with  $Q$  blocks.

$$ICL(\mathcal{M}_Q) = \max_{\theta} \log p(\mathbf{X}^{1:K}, \tilde{\mathbf{Z}}; \theta) - \frac{1}{2} \{Q^2(2^K - 1) \log(Kn(n-1)) + (Q-1) \log n\} .$$

where  $\tilde{\mathbf{Z}}$  are the predictions of the assignments  $\mathbf{Z}$  obtained as a sub-product of the variational EM algorithm (see Appendix B). As in the BIC criteria, the log refers to the number of data. Thus, the  $n$  nodes are used to estimate the  $Q-1$  probabilities

$\alpha_1, \dots, \alpha_{Q-1}$ . The  $Kn(n - 1)$  edges are used to estimate  $\boldsymbol{\pi}$ . No theoretical results exist for the ICL properties, but this criteria has proved its efficiency in practice.

**Remark 4.** *If we consider covariates as in equation (5), then the ICL is adapted :*

$$ICL(\mathcal{M}_Q) = \max_{\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\beta}} \log p(\mathbf{X}^{1:K}, \tilde{\mathbf{Z}}; \boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\beta}) - \frac{1}{2} \{P_Q \log(Kn(n - 1)) + (Q - 1) \log n\} ,$$

where  $P_Q$  is the dimension of  $(\boldsymbol{\mu}, \boldsymbol{\beta})$ .

### 3 Analysis for the laboratory-researcher data

#### 3.1 The data

French scandals during the 1990s involving the voluntary sector around the cancer research dried up large donations that funded research laboratories. In the 2000s, the 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 coordinate the cancer research and to promote collaborations about top researchers. In this context, [Lazega et al. \(2008\)](#) studied the relations of advice between French cancer researchers identified as “Elite” conjointly with the relations of their respective laboratories.

At the inter-individual level, the actors (researchers) were submitted a list of cancer researchers and asked in interviews whom they sought advice from. The advice were of five types, namely 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. The five advice networks are too sparse to be studied separately. That is why they were aggregated. Therefore two researchers are considered as linked if at least one kind of relationship exists. Obviously the links are directed.

At the laboratory level (concerning only laboratories with “Elite” researchers), the lab-

oratory directors were asked to specify what type of resources they exchanged with the other laboratories on the list. The examined resources were the recruitment of post-docs and researchers, the development of programs 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. Once again, to avoid over-sparsity, the various networks were aggregated and two laboratories are said to be linked if there is at least one link between them. From this network on labs, 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/ organization) into a multiplex network (several kinds of relations among 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 covariates are available to describe the researchers: their age, their specialty, two publication performance scores based on two periods of five years, their status (director of the lab or not). Auxiliary covariates are also available the laboratories: their size (number of researchers) and their location.

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

### 3.2 Statistical inference through SBM for multiplex

To estimate the parameters of the multiplex SBM model on this dataset, we use a modified version of the variational EM algorithm ([Daudin et al., 2008](#)) described in the Appendix Section [B](#). The optimization of the ICL criterion derived from likelihood [\(4\)](#)

leads to four blocks (indistinctly denominated clusters or groups). For the sake of clarity, we index by  $R$  and  $L$  (rather than  $\mathbf{X}^1, \mathbf{X}^2$ ) the two adjacency matrices (respectively the direct and indirect ones)

In Figures 1 and 2 we plot the marginal and conditional probabilities of the connections of researchers (respectively labs) 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 lab for the researcher network structure. The obtained blocks are described in Table 1: (1a) gives the sizes of the four blocks; (1b), (1c) and (1d) describe the blocks with respect to the covariates “location”, “director of not” and “specialty”. The estimations by the variational EM procedure were conducted by wmixnet ([Leger, 2014](#)) with our specific implementation of the bivariate Bernoulli model.

We now discuss the results. 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, Figure 1 shows that the existence of a connection (exchange of resources) between labs clearly increases the probability of connection (sharing advice) 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 to 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 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

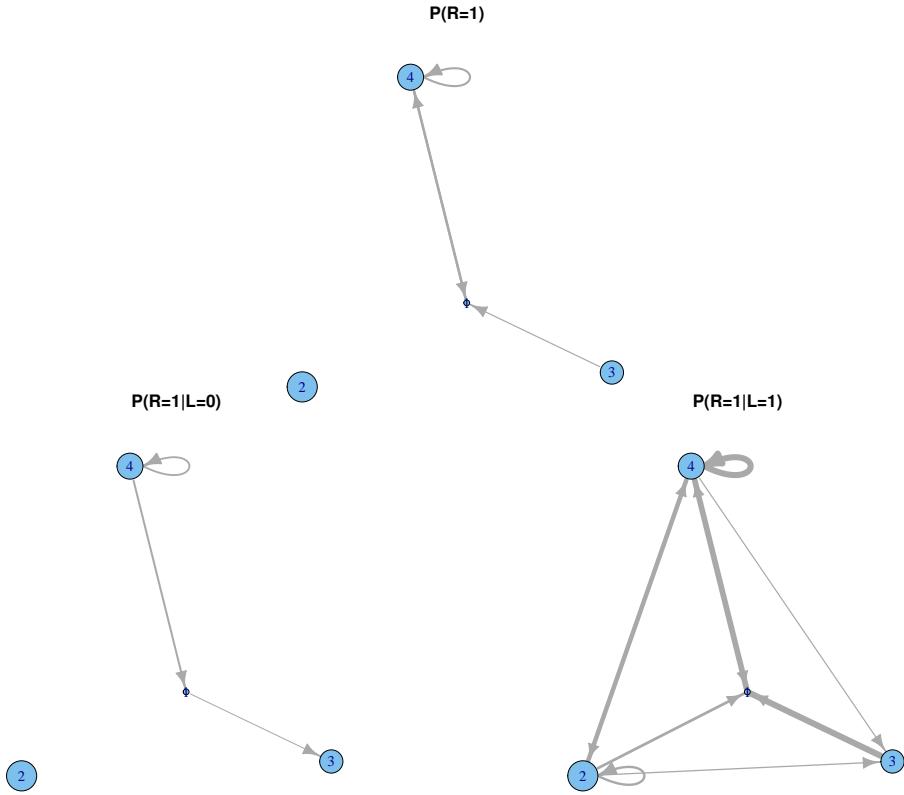


Figure 1: Marginal probabilities of Researcher connections between and within blocks (top) and probabilities of Researcher connections between and within blocks conditionally on absence (bottom left-hand-side) or presence (bottom right-hand-side) of Lab connection. Vertex size is proportional to the block size. Edge width is proportional to the probabilities of connection; if this probability is smaller than 0.1, edges are not displayed.

dependent on their laboratories in terms of connections. Furthermore, Figure 2 shows 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

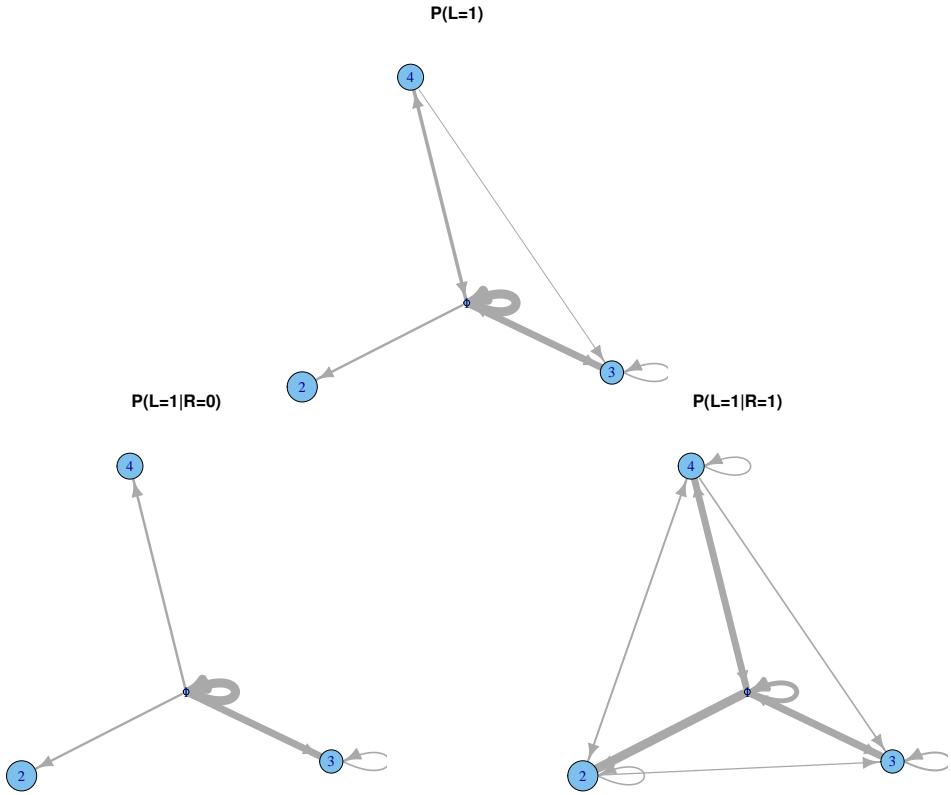


Figure 2: Marginal probabilities of Lab connections between and within blocks (top) and probabilities of Lab connections between and within blocks conditionally on absence (bottom left-hand-side) or presence (bottom right-hand-side) of Researcher connections. Vertex size is proportional to the block size. Edge width is proportional to the probabilities of connection; if this probability is smaller than 0.1, edges are not displayed.

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 labs in terms of size (Figure 3(b)). 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 labs that have highest average indegrees and outdegrees (Figures 3(e), 3(f), 3(g), 3(h)). They

Table 1: (a) Size of blocks obtained by multiplex SBM. (b) Cross frequencies of blocks versus the lab location (in Île-de-France (IdF) or not). (c) Cross frequencies of blocks versus the researcher’s status (lab director or not). (d) Cross frequencies of blocks versus the researcher’s specialties (PH: Public Health; Su: Surgery; He: Hematology; ST: Solid Tumours; FPh: Fundamental pharmacology; FMo: Fundamental molecular research; FMoG: Fundamental molecular genetic research).

block		not idf    idf			not director    director		
		1	1	1	1	1	1
(a)	1	2					
	2	48					
	3	19					
	4	26					

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

have the highest performance in terms of publication performance scores in both periods (Figures 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 labs that have the lowest average indegrees and outdegrees (Figures 3(e), 3(f), 3(g), 3(h)), slightly older than the others (Figure 3(a)), mainly in the smallest labs in terms of size. This block is heterogeneous in terms of specialties (especially 40% clinicians and 25% diagnostic/prevention/epidemiology specialists) except fundamental research (Table 1(d)). They also have among the lowest performance levels for both periods (Figures 3(c), 3(d)), although this is increasing. This is the biggest block. Their behavior may be described as fusional as proposed 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 (Figure 3(a) and Table 1(d)). They have

relatively low indegrees and outdegrees, in labs that have the highest indegrees (after Block 1) and average outdegrees (Figures 3(e), 3(f), 3(g), 3(h)). 70% are among the top performers of this population, i.e. highest performance levels after Block 1 members, for both periods (Figures 3(c), 3(d)). Their dominant relational strategies are individualist or independentists (as shown in Figure 1) since the probabilities of connection between researchers remain quite unchanged, no matter if their laboratories are connected.

Block 4 is also heterogeneous in terms of specialties but its largest subgroup is composed of hematologists (clinical and fundamentalists) (Table 1(d)). Researchers have average indegrees and outdegrees in laboratories that have relatively low indegrees and outdegrees and that are also of average size (Figures 3(e), 3(f), 3(g), 3(h)). There are proportionally more directors of laboratories in this block than in the others (Table 1(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 (Figures 3(c), 3(d)). In this block, researchers can take advantage of their laboratories to be connected with colleagues but they can also hold relations apart from their laboratories.

In this case, SBM highlight in the data a specific kind of block structure that provides an interesting understanding of the effect of dual positioning. SBM mixes people and laboratories that previous analyses used to separate. It is interesting to notice, for example, that SBM partitions the population regardless of geographic location (Table 1(b)) –a criterion that was previously shown to be meaningful to understand collective action in this research milieu. In fact this partition may highlight the links between laboratories and between individuals that cut across the boundaries and remoteness 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.

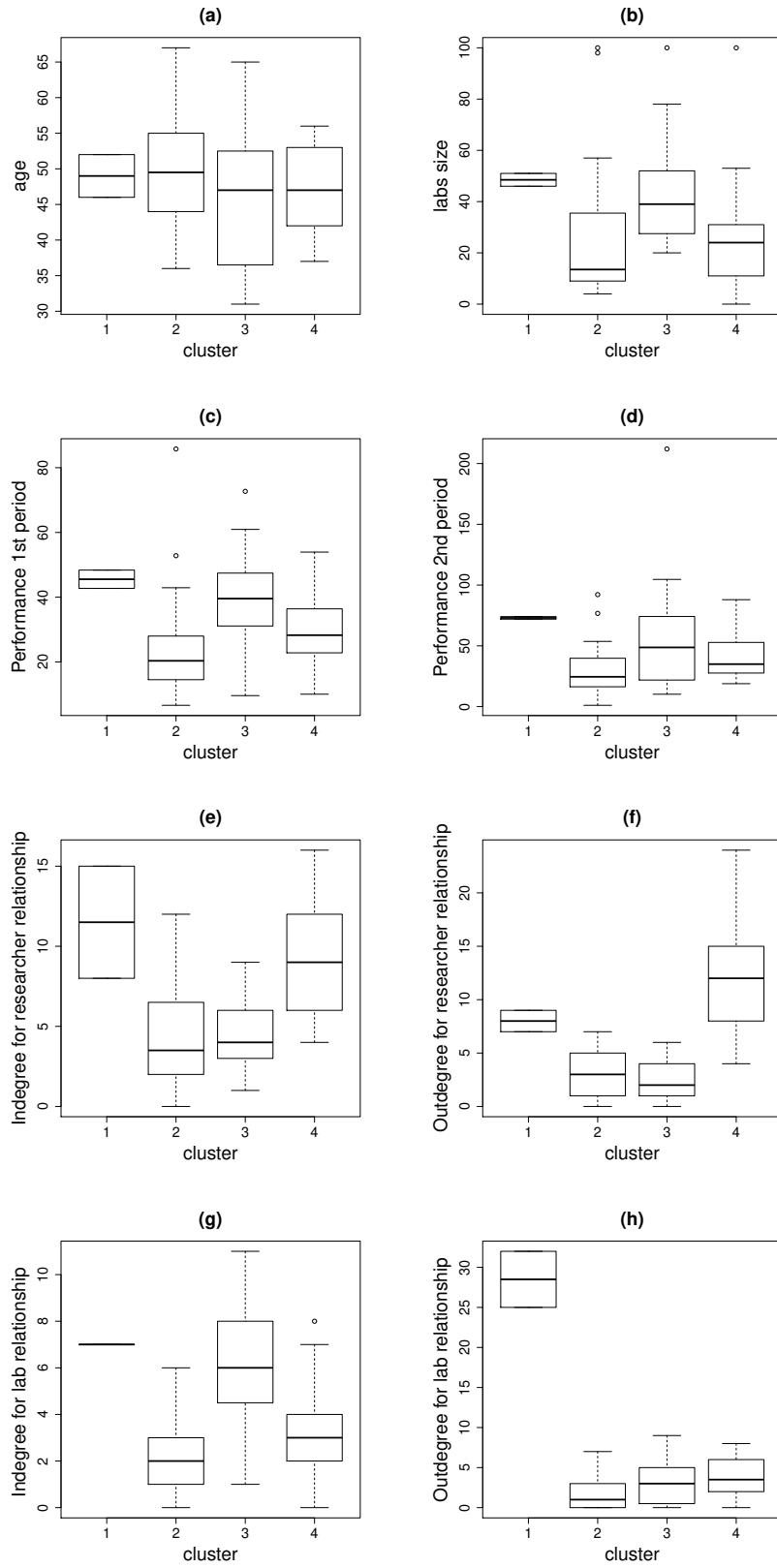


Figure 3: Boxplots of age (a), lab size (b), performance in 1st period (c), performance in 2nd period (d), indegree for researcher relationship (e), outdegree for researcher relationship (f), indegree for lab relationship (g) and outdegree for lab relationship (h) grouped by blocks.

## 4 Discussion

The essence of ‘networks’ is to help actors cut across organizational boundaries to create new relationships (Baker, 1992), to identify new opportunities and, eventually, to create new organizations to use or hoard these new opportunities (Lazega, 2012; Lazega et al., 2013; Tilly, 1998). This work is a step forward a more precise comprehension of such mechanisms.

In this paper, we proposed a SBM modelisation 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 to MERGMs (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 inter-individual level conditioned by the characteristics of the inter-organizational network. Our analysis found a general structure that shows the ways in which these rules can be derived in this case and help 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 inter-individual network does not apply to all actors in the network identically.

In this paper, the model is applied on  $K = 2$  networks. Applying with  $K$  much larger would imply estimation difficulties since the number of parameters exponentially depends on  $K$ . However, particular assumption (such as Markovian dependencies) 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 than the number of individuals than maybe it would be interesting to introduce an other kind of relation, specifying if 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) .

In this version of the analysis, the covariates were studied a posteriori, the classification being purely done on the network. From a practical point of view, including the covariates in the model would imply estimation difficulty (and even more if the effect of the covariates on the connexion probabilities depends on the blocks). However, even if attractive at first sight, the choice of including the covariates in the modeling has to be questioned. Including them will cluster the data beyond the effect of these covariates, while not including them will provide a description of blocks on the basis of the relation, may this relation 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” catches more attention. Indeed, this modelisation 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 relations. However, one could dream of a model which would help the actors 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/her interdependencies at different levels in a strategic manner-, then the study of interdependencies jointly at the inter-individual and the inter-organizational 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 listed before, this analysis of multiplex networks seems therefore adapted to certain types of questions that sociologists ask when they try to combine both individual and contextual factors in order to estimate the likelihood of an individual or a group to adopt a given behavior or to reach a given level of performance. More generally, this approach explores a complex meso-social level of accumulation, of appropriation and of sharing of multiple resources. This level, still poorly known, is difficult to observe without a structural approach.

## A Identifiability of the multiplex SBM and convergence of the maximum likelihood estimates

### A.1 Identifiability

Celisse et al. (2012) have proved the identifiability of the parameters for the uniplex Bernoulli Stochastic Block models. We extend their results to the multiplex SBM. First we recall the notations :  $\forall(i, j) \in \{1, \dots, n\}^2, i \neq j, \forall(q, l) \in \{1, \dots, Q\}^2, \forall w \in \{0, 1\}^K$ , we set:

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

$$\alpha_q = P(Z_i = q)$$

Let us introduce  $\boldsymbol{\pi}^{(w)} = (\pi_{ql}^{(w)})_{(q,l) \in \{1\dots Q\}^2}$  and  $\boldsymbol{\pi} = (\boldsymbol{\pi}^{(w)})_{w \in \{0,1\}^K}$ .  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_Q)$ .

Theorem A.1 sets the identifiability of  $\theta = (\boldsymbol{\alpha}, \boldsymbol{\pi})$  in the multiplex SBM.

**Theorem A.1.** *Let  $n \geq 2Q$ . Assume that for any  $q \in \{1, \dots, Q\}$ ,  $\alpha_q > 0$  and for every  $w \in \{0, 1\}^K$ , the coordinates 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 the original paper, the identifiability can be proved using algebraic tools. The only difference in the multiplex context is that the proof has to 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 for a member of block  $q$  to have a tie of type  $w$  with an other individual:  $r_q^{(w)}(\theta) = \sum_{l=1}^Q \pi_{ql}^{(w)} \alpha_l$ . Let  $R^{(w)}(\theta)$  be the  $Q$ -square matrix such that  $R_{iq}^{(w)}(\theta) = (r_q^{(w)}(\theta))^i$  for  $i = 0 \dots Q - 1$  and  $q = 1 \dots Q$ .  $R^{(w)}$  is a Vandermonde matrix, which is invertible by assumptions on the coordinates of

$\mathbf{r}^{(w)}(\theta)$ . Now, for any  $w$  and any  $i = 0 \dots 2Q - 1$ , we set:

$$u_i^{(w)}(\theta) = \mathbb{P}(X_{11}^{1:K} = \dots = X_{1i}^{1:K} = w; \theta) = \sum_{q=1}^Q \alpha_q (r_q^{(w)}(\theta))^i, \quad (6)$$

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

$$M_{ij}^{(w)}(\theta) = u_{i+j}^{(w)}(\theta), \quad i = 0 \dots Q, j = 0 \dots Q - 1. \quad (7)$$

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

$$M^{Q(w)}(\theta) = R^{(w)}(\theta) A(\theta) R^{(w)}(\theta)^T, \quad (8)$$

where  $A(\theta)$  is the  $\alpha$ -diagonal matrix. All the  $\alpha_q$  being non-null and  $R^{(w)}(\theta)$  being 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  $V^{i(w)}(\theta) = (1, r_i^{(w)}(\theta), \dots, (r_i^{(w)}(\theta))^Q)$ , then

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

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

$$B(x; \theta) = \det(M^{Q(w)}(\theta)) \prod_{i=0}^{Q-1} (x - r_i^{(w)}(\theta)). \quad (9)$$

Now assume that  $\theta = (\boldsymbol{\pi}, \alpha)$  and  $\theta' = (\boldsymbol{\pi}', \alpha')$  are two sets or parameters such that

for any multiplex graph  $\mathbf{X}^{1:K}$ ,  $\ell(\mathbf{X}^{1:K}; \theta) = \ell(\mathbf{X}^{1:K}; \theta')$ . Consequently, from equation (6), we get  $u_i^{(w)}(\theta) = u_i^{(w)}(\theta')$  and from equation (7), we deduce  $M^{i(w)}(\theta) = M^{i(w)}(\theta')$  for any  $i = 0, \dots, Q - 1$ . The polynomial function  $B$  depending 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  $i = 0, \dots, Q - 1$  thanks to equation (9). Thus  $R^{(w)}(\theta) = R^{(w)}(\theta')$  and

$$A(\theta) = (R^{(w)}(\theta)^\top)^{-1} M^{Q(w)}(\theta) R^{(w)}(\theta) = A(\theta').$$

As a consequence,  $\boldsymbol{\alpha} = \boldsymbol{\alpha}'$ .

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

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

We can write the  $Q \times Q$  matrix  $U^{(w)}(\theta)$  as  $U^{(w)}(\theta) = R^{(w)}(\theta) A(\theta) \boldsymbol{\pi}^{(w)} A(\theta) (R^{(w)}(\theta))^\top$ . Using the fact that  $R^{(w)}(\theta) = R^{(w)}(\theta')$  and  $A(\theta) = A(\theta')$ , we get

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

And the theorem is demonstrated. □

## A.2 Consistency of the maximum likelihood estimator in multiplex SBM models

We study the asymptotic properties of the maximum likelihood estimator in multiplex SBM models. Let  $\mathcal{A}_1$ ,  $\mathcal{A}_2$  and  $\mathcal{A}_3$  be the following assumptions:

Assumption  $\mathcal{A}_1$ : for every  $q \neq q'$ , there exists  $l \in \{1, \dots, Q\}$  such that  $\boldsymbol{\pi}_{ql} \neq \boldsymbol{\pi}_{q'l}$ , or  $\boldsymbol{\pi}_{lq} \neq \boldsymbol{\pi}_{lq'}$ .

Assumption  $\mathcal{A}_2$ : there exists  $\zeta > 0$  such that  $\forall (q, l) \in 1, \dots, Q$ ,  $\boldsymbol{\pi}_{ql}^{(w)} \in ]0, 1[ \Rightarrow \boldsymbol{\pi}_{ql}^{(w)} \in$

$[\zeta, 1 - \zeta]$ .

Assumption A<sub>3</sub>: there exists  $\gamma \in (0, 1/Q)$  such that  $\forall q \in 1, \dots, Q$ ,  $\alpha_q \in [\gamma, 1 - \gamma]$ .

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

$$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\pi}}) = \text{Argmax}_{(\boldsymbol{\alpha}, \boldsymbol{\pi})} \mathcal{L}(\mathbf{X}^{1:K}; \boldsymbol{\alpha}, \boldsymbol{\pi}),$$

where

$$\begin{aligned} \ell(\mathbf{X}^{1:K}; \boldsymbol{\alpha}, \boldsymbol{\pi}) &= \sum_{\mathbf{Z} \in \{1, \dots, Q\}^n} e^{\ell_1(\mathbf{X}^{1:K} | \mathbf{Z}; \boldsymbol{\pi})} \prod_{i=1}^n \alpha_{Z_i}, \\ \ell_1(\mathbf{X} | \mathbf{Z}; \boldsymbol{\pi}) &= \sum_{i,j, i \neq j}^n \log \mathbb{P}(X_{ij}^{1:K} | Z_i, Z_j; \boldsymbol{\pi}) = \sum_{i,j, i \neq j}^n \sum_{w \in \{0,1\}^K} \mathbb{I}_{X_{ij}^{1:K} = w} \log \pi_{Z_i Z_j}^{(w)}. \end{aligned}$$

**Theorem A.2.** *Let (A<sub>1</sub>), (A<sub>2</sub>) and (A<sub>3</sub>) hold. Then, for any distance  $d(\cdot, \cdot)$  on the set of parameter  $\boldsymbol{\pi}$ , we have:*

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

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

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

Note the rate  $o_{\mathbb{P}}(\sqrt{\log(n)}/n)$  has not been proved yet. However, in the unilevel context, there is empirical evidence that the rate convergence on  $\hat{\boldsymbol{\pi}}$  is  $1/n$  ([Gazal et al., 2012](#)).

## B Variational EM algorithm for multiplex SBM : principle, details and convergence

### B.1 General principal of the variational EM

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

$$\ell(\mathbf{X}^{1:K}; \theta) = \sum_{\mathbf{Z} \in \{1, \dots, Q\}^n} e^{\ell_1(\mathbf{X}^{1:K} | \mathbf{Z}; \boldsymbol{\pi})} p(\mathbf{Z}; \boldsymbol{\alpha}), \quad (10)$$

which is not tractable as soon as  $n$  and  $Q$  are large. The variational EM is an alternative method to maximize the marginal likelihood with respect to  $\theta$ . The variational EM (applied in the SBM context by [Daudin et al. \(2008\)](#)) relies on the following decomposition of (10). Let  $\mathcal{R}_{\mathbf{X}^{1:K}}$  be any probability distribution on  $\mathbf{Z}$ , we have:

$$\begin{aligned} \log \ell(\mathbf{X}^{1:K} | \theta) &= \sum_{\mathbf{Z}} \mathcal{R}_{\mathbf{X}^{1:K}}(\mathbf{Z}) \log p(\mathbf{X}^{1:K}, \mathbf{Z}; \theta) - \sum_{\mathbf{Z}} \mathcal{R}_{\mathbf{X}^{1:K}}(\mathbf{Z}) \log \mathcal{R}_{\mathbf{X}^{1:K}}(\mathbf{Z}) \\ &\quad + \mathbf{KL}[\mathcal{R}_{\mathbf{X}^{1:K}}, p(\cdot | \mathbf{X}^{1:K}; \theta)], \end{aligned} \quad (11)$$

where  $\mathbf{KL}$  is the Kullback-Leibler distance,  $p(\mathbf{X}^{1:K}, \mathbf{Z}; \theta)$  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  $\theta$ . Instead of maximizing  $\log \ell(\mathbf{X}^{1:K}; \theta)$ , the variational EM optimizes a lower bound  $\mathcal{I}_\theta(\mathcal{R}_{\mathbf{X}^{1:K}})$  of  $\log \ell(\mathbf{X}^{1:K} | \theta)$  where:

$$\begin{aligned} \mathcal{I}_\theta(\mathcal{R}_{\mathbf{X}^{1:K}}) &= \log \ell(\mathbf{X}^{1:K} | \theta) - \mathbf{KL}[\mathcal{R}_{\mathbf{X}^{1:K}}, p(\cdot | \mathbf{X}^{1:K}; \theta)], \\ &= \sum_{\mathbf{Z}} \mathcal{R}_{\mathbf{X}^{1:K}}(\mathbf{Z}) \log p(\mathbf{X}^{1:K}, \mathbf{Z}; \theta) - \sum_{\mathbf{Z}} \mathcal{R}_{\mathbf{X}^{1:K}}(\mathbf{Z}) \log \mathcal{R}_{\mathbf{X}^{1:K}}(\mathbf{Z}), \\ &\leq \log \ell(\mathbf{X}^{1:K} | \theta). \end{aligned} \quad (12)$$

Note that, thanks to equality (12), optimizing  $\mathcal{I}_\theta(\mathcal{R}_{\mathbf{X}^{1:K}})$  with respect to  $\theta$  no longer requires the computation of the marginal likelihood. Note also that the equality  $\mathcal{I}_\theta(\mathcal{R}_{\mathbf{X}^{1:K}}) = \log \ell(\mathbf{X}^{1:K}; \theta)$  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 of  $p(\cdot | \mathbf{X}^{1:K}; \theta)$  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  $\boldsymbol{\tau}_i = (\tau_{i1}, \dots, \tau_{iq})$ . Finally, the variational EM updates alternatively  $\theta$  and  $\boldsymbol{\tau}$  in the following way. At iteration  $(t)$ , given the current state  $(\theta^{(t-1)}, \boldsymbol{\tau}^{(t-1)})$ ,

- **Step 1** Compute

$$\boldsymbol{\tau}^{(t)} = \arg \min_{\boldsymbol{\tau}} \mathbf{KL}[\mathcal{R}_{\mathbf{X}^{1:K}, \boldsymbol{\tau}}, p(\cdot | \mathbf{X}^{1:K}; \theta^{(t-1)})] = \arg \max_{\boldsymbol{\tau}} \mathcal{I}_{\theta^{(t-1)}}(\mathcal{R}_{\mathbf{X}^{1:K}, \boldsymbol{\tau}}).$$

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

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

## B.2 Details of the calculus for multiplex SBM models

We now detail **Step 1** and **Step 2** for multiplex SBM models.

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

$$\boldsymbol{\tau}^{(t)} = \arg \min_{\boldsymbol{\tau}} \mathbf{KL}[\mathcal{R}_{\mathbf{X}^{1:K}, \boldsymbol{\tau}}, p(\cdot | \mathbf{X}^{1:K}; \theta^{(t-1)})] = \arg \max_{\boldsymbol{\tau}} \mathcal{I}_{\theta^{(t-1)}}(\mathcal{R}_{\mathbf{X}^{1:K}, \boldsymbol{\tau}}).$$

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

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

with

$$\begin{aligned} \log p(\mathbf{X}^{1:K}, \mathbf{Z}; \theta) &= \ell_1(\mathbf{X}^{1:K} | \mathbf{Z}; \theta) + \log p(\mathbf{Z}; \theta), \\ &= \sum_{i,j, i \neq j} \log p(X_{ij}^{1:K} | Z_i, Z_j; \theta) + \sum_{i=1}^n \log \alpha_{Z_i}. \end{aligned}$$

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

$$\begin{aligned} \mathcal{I}_\theta(\mathcal{R}_{\mathbf{X}^{1:K}, \boldsymbol{\tau}}) &= \sum_{\mathbf{Z}} \mathcal{R}_{\mathbf{X}^{1:K}, \boldsymbol{\tau}}(\mathbf{Z}) \left[ \sum_{i,j, i \neq j} \log p(X_{ij}^{1:K} | Z_i, Z_j; \theta) + \sum_{i=1}^n \log \alpha_{Z_i} \right] \\ &\quad - \sum_{\mathbf{Z}} \mathcal{R}_{\mathbf{X}^{1:K}, \boldsymbol{\tau}}(\mathbf{Z}) \log \mathcal{R}_{\mathbf{X}^{1:K}, \boldsymbol{\tau}}(\mathbf{Z}), \\ &= \sum_{q,l} \sum_{i,j, i \neq j} \log p(X_{ij}^{1:K} | Z_i = q, Z_j = l; \theta) \tau_{iq} \tau_{jl} \\ &\quad + \sum_{i=1}^n \sum_{q=1}^Q \tau_{iq} \log \alpha_q - \sum_{i=1}^n \sum_{q=1}^Q \tau_{iq} \log \tau_{iq}, \end{aligned}$$

where  $\log p(X_{ij}^{1:K} | Z_i = q, Z_j = l; \theta)$ 's expression is given in equation (3).

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

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

which leads to the following fixed point problem:

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

which has to be solved under the constraints  $\forall i = 1 \dots n, \sum_{q=1}^Q \widehat{\tau}_{iq} = 1$ . This optimization problem is solved using a standard fixed point algorithm.

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

Once the  $\widehat{\boldsymbol{\tau}}$  have been optimized, the parameters  $\theta$  maximizing  $\mathcal{I}_{\theta}(\mathcal{R}_{\mathbf{X}^{1:K}, \widehat{\boldsymbol{\tau}}})$  have to be computed under the constraints:  $\sum_{q=1}^Q \alpha_q = 1$  and  $\sum_{w \in \{0,1\}^L} \pi_{ql}^{(w)} = 1$  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:

$$\widehat{\alpha}_q = \frac{1}{n} \sum_{i=1}^n \widehat{\tau}_{iq}, \quad \widehat{\pi}_{ql}^{(w)} = \frac{\sum_{ij} \widehat{\tau}_{iq} \widehat{\tau}_{jl} \mathbb{I}_{X_{ij}^{1:K} = w}}{\sum_{ij} \widehat{\tau}_{iq} \widehat{\tau}_{jl}}$$

**Remark 5.** If the edge probabilities depend on covariates:

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

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

### B.3 Convergence of the VarEM estimates

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

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

The variational estimators (VE) are obtained by:

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

**Theorem B.1.** *Assume that  $(\mathcal{A}_1)$ ,  $(\mathcal{A}_2)$  and  $(\mathcal{A}_3)$  hold. Then for any distance on the set of parameters  $\boldsymbol{\pi}$ ,*

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

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

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

### B.4 About the proofs

The proofs of these results require many intermediate results which won't 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  $\log \frac{p(\mathbf{Z}|\mathbf{X}^{1:K}; \boldsymbol{\alpha}, \boldsymbol{\pi})}{p(\mathbf{Z}^*|\mathbf{X}^{1:K}; \boldsymbol{\alpha}, \boldsymbol{\pi})}$ . In

the unilevel case, this quantity is:

$$\begin{aligned} \log \frac{p(\mathbf{Z}|\mathbf{X}^{1:K}; \boldsymbol{\alpha}, \boldsymbol{\pi})}{p(\mathbf{Z}^*|\mathbf{X}^{1:K}; \boldsymbol{\alpha}, \boldsymbol{\pi})} &= \log \frac{p(\mathbf{X}^{1:K}|\mathbf{Z}; \boldsymbol{\pi})}{p(\mathbf{X}^{1:K}|\mathbf{Z}^*; \boldsymbol{\pi})} + \log \frac{p(\mathbf{Z}; \boldsymbol{\alpha})}{p(\mathbf{Z}^*; \boldsymbol{\alpha})} \\ &= \sum_{i \neq j} X_{ij} \log \frac{\pi_{Z_i, Z_j}^{(w)}}{\pi_{z_i^*, z_j^*}^{(w)}} + (1 - X_{ij}) \log \frac{1 - \pi_{Z_i, Z_j}^{(w)}}{1 - \pi_{z_i^*, z_j^*}^{(w)}} + \sum_{i=1}^n \log \frac{\alpha_{Z_i}}{\alpha_{z_i^*}} \end{aligned}$$

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

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

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.

## References

- Ambroise, C. and Matias, C. (2012). New consistent and asymptotically normal parameter estimates for random-graph mixture models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 74(1):3–35.
- Baker, W. E. (1992). The network organization in theory and practice. *Networks and organizations: Structure, form, and action*, 396:429.
- Bianconi, G. (2013). Statistical mechanics of multiplex networks: Entropy and overlap. *Phys. Rev. E*, 87:062806.
- Bickel, P. and Chen, A. (2009). A nonparametric view of network models and Newman-Girvan and other modularities. *JProc. Natl. Acad. Sci. U.S.A.*, 106(50):21068–21073.
- Brummitt, C. D., D’Souza, R. M., and Leicht, E. (2012). Suppressing cascades of load in interdependent networks. *Proceedings of the National Academy of Sciences*, 109(12):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. Stat.*, 6:1847–1899.
- Coleman, J. S. (1982). *The asymmetric society*. Syracuse University Press.
- Cozzo, E., Arenas, A., and Moreno, Y. (2012). Stability of boolean multilevel networks. *Phys. Rev. E*, 86:036115.
- Daudin, J. J., Picard, F., and Robin, S. (2008). A mixture model for random graphs. *Statistics and Computing*, 18(2):173–183.
- Gazal, S., Daudin, J.-J., and Robin, S. (2012). Accuracy of variational estimates for random graph mixture models. *Journal of Statistical Computation and Simulation*, 82(6):849–862.
- Jaakkola, T. S. (2000). Tutorial on variational approximation methods. In *In Advanced Mean Field Methods: Theory and Practice*, pages 129–159. MIT Press.
- Lazega, E. (2012). Analyses de réseaux et classes sociales. *Revue française de socio-économie*, 2:273–279.
- Lazega, E., Jourda, M.-T., and Mounier, L. (2007). Des poissons et des mares: l’analyse de réseaux multi-niveaux. *Revue française de sociologie*, 48(1):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. *European sociological review*, 29(6):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. *Social Networks*, 30(2):159 – 176.
- Leger, J.-B. (2014). Wmixnet: Software for clustering the nodes of binary and valued graphs using the stochastic block model. *ArXiv e-prints*.

- Loe, C. W. and Jeldtoft Jensen, H. (2014). Comparison of Communities Detection Algorithms for Multiplex. *ArXiv e-prints*.
- Maggioni, M. A., Breschi, S., and Panzarasa, P. (2013). Multiplexity, growth mechanisms and structural variety in scientific collaboration networks. *Industry and Innovation*, 20(3):185–194.
- Mariadassou, M., Robin, S., and Vacher, C. (2010). Uncovering latent structure in valued graphs: A variational approach. *The Annals of Applied Statistics*, 4(2):715–742.
- Matias, C. and Robin, S. (2014). Modeling heterogeneity in random graphs through latent space models: a selective review.
- 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(5980):876–878.
- Nicosia, V., Bianconi, G., Latora, V., and Barthelemy, M. (2013). Growing multiplex networks. *Physical review letters*, 111(5):058701.
- Nowicki, K. and Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455):1077–1087.
- Perrow, C. (1991). A society of organizations. *Theory and society*, 20(6):725–762.
- Presthus, R. (1962). *The organizational society: An analysis and a theory*. Vintage Books/Random House.
- Rank, O. N., Robins, G. L., and Pattison, P. E. (2010). Structural logic of intraorganizational networks. *Organization Science*, 21(3):745–764.
- Saumell-Mendiola, A., Serrano, M. Á., and Boguñá, M. (2012). Epidemic spreading on interconnected networks. *Physical Review E*, 86(2):026106.

- Snijders, T. A. and Baerveldt, C. (2003). A multilevel network study of the effects of delinquent behavior on friendship evolution. *Journal of mathematical sociology*, 27(2-3):123–151.
- Snijders, T. A. B. and Nowicki, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *J. Classification*, 14(1):75–100.
- Szell, M., Lambiotte, R., and Thurner, S. (2010). Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences*, 107(31):13636–13641.
- Tilly, C. (1998). *Durable inequality*. Univ of California Press.
- Wang, P., Robins, G., Pattison, P., and Lazega, E. (2013). Exponential random graph models for multilevel networks. *Social Networks*, 35(1):96–115.
- Wasserman, S. (1994). *Social network analysis: Methods and applications*, volume 8. Cambridge university press.