



HAL
open science

Consistency of maximum-likelihood and variational estimators in the Stochastic Block Model

Alain Celisse, J.-J. Daudin, Laurent Pierre

► **To cite this version:**

Alain Celisse, J.-J. Daudin, Laurent Pierre. Consistency of maximum-likelihood and variational estimators in the Stochastic Block Model. 2011. hal-00593644v2

HAL Id: hal-00593644

<https://hal.science/hal-00593644v2>

Preprint submitted on 29 Nov 2011 (v2), last revised 29 Sep 2012 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Consistency of maximum-likelihood and variational estimators in the Stochastic Block Model *

Alain Celisse[†]

*Laboratoire de Mathématique
UMR 8524 CNRS– Université Lille 1
F-59 655 Villeneuve d’Ascq Cedex, France
e-mail: celisse@math.univ-lille1.fr*

and

Jean-Jacques Daudin

*UMR 518 AgroParisTech/INRA MIA
16 rue Claude Bernard
F-75 231 Paris Cedex 5, France
e-mail: daudin@agroparistech.fr*

and

Laurent Pierre

*Université Paris X
200 avenue de la République
F- 92 001 Nanterre Cedex France
e-mail: laurent.pierre@u-paris10.fr*

Abstract: The *stochastic block model* (SBM) is a probabilistic model designed to describe heterogeneous *directed* and *undirected* graphs. In this paper, we address the asymptotic inference on SBM by use of maximum-likelihood and variational approaches. The identifiability of SBM is proved, while asymptotic properties of maximum-likelihood and variational estimators are provided. In particular, the consistency of these estimators is settled, which is, to the best of our knowledge, the first result of this type for variational estimators with random graphs.

AMS 2000 subject classifications: Primary 62G05, 62G20; secondary 62E17, 62H30.

Keywords and phrases: Random graphs, Stochastic Block Model, maximum likelihood estimators, variational estimators, consistency, concentration inequalities.

1. Introduction

In the last decade, networks have arisen in numerous domains such as social sciences and biology. They provide an attractive graphical representation of

* Authors thank Catherine Matias and Mahendra Mariadassou for their helpful comments.

[†]This work has been supported by the French Agence Nationale de la Recherche (ANR) under reference ANR-09-JCJC-0027-01

complex data. However, the increasing size of networks and their great number of connections have made it difficult to interpret *network representations of data* in a satisfactory way. This has strengthened the need for statistical analysis of such networks, which could raise latent patterns in the data.

Interpreting networks as random graphs, *unsupervised classification* (clustering) of the vertices of the graph has received much attention. It is based on the idea that vertices with a similar connectivity can be gathered in the same class. The initial graph can be replaced by a simpler one without losing too much information. This idea has been successfully applied to social (Nowicki and Snijders, 2001) and biological (Picard *et al.*, 2009) networks. It is out of the scope of the present work to review all of them.

Mixture models are a convenient and classical tool to perform unsupervised classification in usual statistical settings. Mixture models for random graphs were first proposed by Holland *et al.* (1983) who defined the so-called *stochastic block* model (SBM), in reference to an older *non stochastic block* model widely used in social science. Assuming each vertex belongs to only one class, a *latent variable* (called the *label*) assigns every vertex to its corresponding class. SBM is therefore a versatile means to infer underlying structures of the graph. Subsequently, several versions of SBM have been studied and it is necessary to formally distinguish between them. Three binary distinctions can be made to this end:

1. The graph can be *directed* or *undirected*.
2. The graph can be *binary* or *weighted*.
3. The model can (i) rely on *latent random variables* (the labels), or (ii) assume the labels are *unknown parameters*:
 - (i) SBM is a usual mixture model with random multinomial *latent variables* (Nowicki and Snijders, 2001; Daudin *et al.*, 2008; Ambroise and Matias, 2010). In this model, vertices are sampled in a population and the concern is on the population parameters, that is the frequency of each class and their connectivity parameters.
 - (ii) An alternative *conditional* version of *SBM* (called CSBM) has been introduced. In CSBM, former latent random variables (the labels) are considered as *fixed parameters*. The main concerns are then the estimation of between- and within-class connectivity parameters as well as of the unknown label associated to every vertex (see Rohe *et al.*, 2010; Choi *et al.*, 2011, for a use of CSBM to perform clustering).

The main interest of SBM is that it provides a *more realistic and versatile model than the famous Erdős-Rényi graph* while remaining easily interpretable. However unlike usual statistical settings where independence is assumed, one specificity of SBM is that *vertices are not independent*. Even if numerous approaches have been developed to overcome this challenging problem, most of them suffer some high computational cost. See for instance Snijders and Nowicki (1997) who studied maximum-likelihood estimators of SBM with only two classes and *binary undirected* graphs, or Nowicki and Snijders (2001) where Gibbs sampling is applied for more than

two classes. Other strategies also exist that only apply to *undirected* graphs (see for instance [Bickel and Chen, 2010](#); [Choi et al., 2011](#); [Rohe et al., 2010](#); [Ambroise and Matias, 2010](#), to name but a few of them).

A *variational approach* has been proposed by [Daudin et al. \(2008\)](#) to remedy this computational burden. It can be used with *binary directed* SBM and avoids the algorithmic complexity of the likelihood and bayesian approaches ([Mixnet \(2009\)](#), see) and also [Mariadassou et al. \(2010\)](#) for *weighted undirected* SBM analyzed with a variational approach). However even if its practical performance shows a great improvement, *variational approach remains poorly understood from a theoretical point of view*. For instance, no consistency result does exist for *maximum likelihood* or *variational* estimators of SBM parameters (see ([Ambroise and Matias, 2010](#); [Choi et al., 2011](#); [Rohe et al., 2010](#); [Bickel and Chen, 2010](#)) for various consistency results of other estimators of the SBM parameters). Nonetheless, empirical clues ([Gazal et al., 2011](#)) have already supported the consistency of variational estimators in SBM. Establishing such asymptotic properties is precisely the purpose of the present work.

In this paper, the identifiability of *binary directed* SBM is proved under very mild assumptions for the first time to our knowledge. The *asymptotics of maximum-likelihood* and *variational estimators* is also addressed by use of concentration inequalities. In particular, variational estimators are shown to be asymptotically equivalent to maximum-likelihood ones, and then consistent. The framework of the present work assumes the number Q of classes to be known and independent of the number of vertices. Some attempts exist to provide a data-driven choice of Q (see [Daudin et al., 2008](#)), but this question is out of the scope of the present work.

The rest of the paper is organized as follows. The main notation and assumptions are introduced in Section 2, where identifiability of SBM is settled. Section 3 is devoted to the consistency of the maximum-likelihood estimators (MLE), and Section 4 to the asymptotic equivalence between variational and maximum-likelihood estimators. In particular, the consistency of variational estimators (VE) is proved. Some concluding remarks are provided in Section 5 with some further important questions.

2. Model definition and identifiability

Let $\Omega = (\mathcal{V}, \mathcal{X})$ be the set of infinite random graphs where $\mathcal{V} = \mathbb{N}$ denotes the set of countable vertices and $\mathcal{X} = \{0, 1\}^{\mathbb{N}^2}$ the corresponding set of adjacency matrices. The random adjacency matrix, denoted by $X = \{X_{ij}\}_{i,j \in \mathbb{N}}$, is given by: for $i \neq j$, $X_{ij} = 1$ if an edge exists from vertex i to vertex j and $X_{ij} = 0$ otherwise, and $X_{ii} = 0$ (no loop). Let \mathbb{P} denote a probability measure on Ω .

2.1. Stochastic Block Model (SBM)

Let us consider a random graph with n vertices $\{v_i\}_{i=1, \dots, n}$. These vertices are assumed to be split into Q classes $\{C_q\}_{q=1, \dots, Q}$ depending on their structural

properties.

Set $\alpha = (\alpha_1, \dots, \alpha_Q)$ with $0 < \alpha_q < 1$ and $\sum_q \alpha_q = 1$. For every q , α_q denotes the probability for a given vertex to belong to the class C_q . For any vertex v_i , its label Z_i is generated as follows

$$\{Z_i\}_{1 \leq i \leq n} \stackrel{i.i.d.}{\sim} \mathcal{M}(n; \alpha_1, \dots, \alpha_Q) .$$

where $\mathcal{M}(n; \alpha_1, \dots, \alpha_Q)$ denotes the multinomial distribution. Let $Z_{[n]} = (Z_1, \dots, Z_n)$ denote the random vector of the labels of (v_1, \dots, v_n) .

The observation consists of an adjacency matrix $X_{[n]} = \{X_{i,j}\}_{1 \leq i,j \leq n}$, where $X_{i,i} = 0$ for every i and

$$X_{ij} \mid Z_i = q, Z_j = l \stackrel{i.i.d.}{\sim} \mathcal{B}(\pi_{q,l}), \quad \forall i \neq j ,$$

where $\mathcal{B}(\pi_{q,l})$ denotes the Bernoulli distribution with parameter $0 \leq \pi_{q,l} \leq 1$ for every (q, l) .

The log-likelihood is given by

$$\mathcal{L}_2(X_{[n]}; \alpha, \pi) = \log \left(\sum_{z_{[n]}} e^{\mathcal{L}_1(X_{[n]}; z_{[n]}, \pi)} \mathbb{P} [Z_{[n]} = z_{[n]}] \right) , \quad (1)$$

where

$$\mathcal{L}_1(X_{[n]}; z_{[n]}, \pi) = \sum_{i \neq j} X_{i,j} \log \pi_{z_i, z_j} + (1 - X_{i,j}) \log(1 - \pi_{z_i, z_j}) , \quad (2)$$

$\mathbb{P} [Z_{[n]} = z_{[n]}] = \prod_{i=1}^n \alpha_{z_i}$. In the following, let $\theta = (\alpha, \pi)$ denote the parameter and $\theta^* = (\alpha^*, \pi^*)$ be the true parameter value. Notice that the $X_{i,j}$ s are not independent. However, conditioning on $Z_i = q, Z_j = l$ yields independence.

Recall that the number Q of classes is assumed to be known and the purpose of the present work is to efficiently estimate the parameters of SBM.

2.2. Assumptions

In the present section, several assumptions are discussed, which will be used all along the paper .

Assumption 1 (A1). For every $(q, l) \neq (q', l')$,

$$\pi_{q,l} \neq \pi_{q',l'} .$$

This identifiability assumption is strong but not necessary. It will be used to simply proofs and get a first understanding of main involved quantities (Theorem 3.1). However, Assumption (A1) will be relaxed by (A2), which is a necessary condition for the identifiability of SBM.

Assumption 2 (A2). For every $q \neq q'$, there exists $l \in \{1, \dots, Q\}$ such that

$$\pi_{q,l} \neq \pi_{q',l} \text{ or } \pi_{l,q} \neq \pi_{l,q'} .$$

(A2) excludes the possibility that two columns are equal and that the corresponding rows are also equal. This condition is consistent with the goal of SBM, which is to define Q classes C_1, \dots, C_Q with different structural properties. For instance, the connectivity properties of vertices in C_q must be different from that of vertices in C_l with $q \neq l$. Therefore, settings where this assumption is violated correspond to ill-specified models with too many classes.

Assumption 3 (A3). There exists $\zeta > 0$ such that

$$\forall (q, l) \in \{1, \dots, Q\}^2, \quad \pi_{q,l} \in]0, 1[\Rightarrow \pi_{q,l} \in [\zeta, 1 - \zeta] .$$

SBM can deal with null probabilities of connection between vertices. However, the use of $\log \pi_{q,l}$ implies a special treatment for the case $\pi_{q,l} \in \{0, 1\}$. Therefore, a distinct analysis will be performed for $\pi_{q,l} \in \{0, 1\}$ and $\pi_{q,l} \notin \{0, 1\}$. Note that all along the present paper, (A3) is always assumed to hold with ζ not depending on n .

Assumption 4 (A4). There exists $0 < \gamma < 1/Q$ such that

$$\forall q \in \{1, \dots, Q\}, \quad \alpha_q \in [\gamma, 1 - \gamma] .$$

This assumption implies that no class is drained. Actually, the identifiability of SBM requires that every $\alpha_q \in (0, 1)$ for $q \in \{1, \dots, Q\}$, which is implied by (A4). In this paper, it is assumed that γ does not depend on n .

Assumption 5 (A5). There exists $0 < \gamma < 1/Q$ and $n_0 \in \mathbb{N}^*$ such that

$$\forall q \in \{1, \dots, Q\}, \forall n \geq n_0, \quad \frac{N_q(z_{[n]}^*)}{n} \geq \gamma ,$$

where $N_q(z_{[n]}^*) = |\{1 \leq i \leq n \mid z_i^* = q\}|$.

Note that (A5) is the empirical version of (A4). Indeed by definition of SBM, $z_{[n]}^*$ is the realization of a multinomial random variable with parameters $(\alpha_1, \dots, \alpha_Q)$. Then, any multinomial random variable will satisfy the requirement of (A5) with high probability. This assumption will be used in particular in Theorem 3.1.

2.3. Identifiability

The identifiability of the parameters in SBM have been first obtained by Allman *et al.* (2009) for *undirected* graphs (π is symmetric): if $Q = 2$, $n \geq 16$, and the coefficients of π are all different, the parameters are identifiable up to label switching. Allman *et al.* (2011) also established that for $Q > 2$, if n is even and $n \geq (Q - 1 + \frac{(Q+2)^2}{4})^2$ (with a similar condition if n is odd), the parameters

of SBM are generically identifiable, that is, identifiable except on a set with null Lebesgue measure.

First, generic identifiability (up to label switching) of the parameters of SBM is proved for *directed or undirected* graphs as long as $n \geq 2Q$.

Theorem 2.1. *Let $n \geq 2Q$ and assume that for any $1 \leq q \leq Q$, $\alpha_q > 0$ and the coordinates of $r = \pi\alpha$ are distinct. Then, SBM is identifiable.*

The assumption on r is not strongly restrictive since the set of vectors violating this assumption is of Lebesgue measure 0. Therefore, Theorem 2.1 actually asserts the generic identifiability of SBM (see Allman *et al.*, 2009). Moreover, Theorem 2.1 also holds with $r' = \pi^t\alpha$ (instead of $r = \pi\alpha$), and also with vectors r'' given by $r''_q = \sum_l \pi_{q,l}\pi_{l,q}\alpha_l$ for every $1 \leq q \leq Q$.

Let us further interpret that assumption on the coordinates of r on two examples. The first one is a particular instance of SBM called *Affiliation Model* (Allman *et al.*, 2011).

Example 1 (Affiliation model). *In this model, the matrix π is only parametrized by two coefficients π_1 and π_2 ($\pi_1 \neq \pi_2$), which respectively correspond to within-class and between-class connectivities between edges. With $Q = 2$, the matrix π is given by*

$$\pi = \begin{pmatrix} \pi_1 & \pi_2 \\ \pi_2 & \pi_1 \end{pmatrix} .$$

Then, requiring $(\pi\alpha)_1 = \pi_1\alpha_1 + \pi_2\alpha_2$ is not equal to $(\pi\alpha)_2 = \pi_2\alpha_1 + \pi_1\alpha_2$ amounts to impose that $\alpha_1 \neq \alpha_2$.

Since within- and between-class connectivities are the same for the two classes, distinguishing between them therefore requires a different proportion of edges in these classes ($\alpha_1 \neq \alpha_2$).

The second example describes a more general setting than Example 1 in which the assumption on the coordinates of r can be more deeply understood.

Example 2 (Permutation-invariant matrices). *For some matrices π , there exist permutations $\sigma : \{1, \dots, Q\} \rightarrow \{1, \dots, Q\}$ such that π remains unchanged if one permutes both its rows and columns according to σ . More precisely, let π^σ denote the matrix defined by*

$$\pi_{q,l}^\sigma = \pi_{\sigma(q),\sigma(l)} ,$$

for every $1 \leq q, l \leq Q$. Then, $\pi^\sigma = \pi$.

For a given matrix π , let us define the set of permutations letting π invariant by

$$\mathfrak{S}^\pi = \{ \sigma : \{1, \dots, Q\} \rightarrow \{1, \dots, Q\} \mid \pi^\sigma = \pi \} .$$

The matrix π is said permutation-invariant if $\mathfrak{S}^\pi \neq \{Id\}$, where Id denotes the identity permutation. For instance in the affiliation model (Example 1), π is permutation-invariant since \mathfrak{S}^π is the whole set of permutations on $\{1, \dots, Q\}$.

Let us first notice that “label-switching” translates into the following property. For any permutation of $\{1, \dots, Q\}$,

$$\pi^\sigma \alpha^\sigma = \pi \alpha \quad , \quad (3)$$

where $\alpha_q^\sigma = \alpha_{\sigma(q)}$ for every q . The main point is that label-switching arises whatever the choice of (α, π) , and for every σ .

By contrast, only permutation-invariant matrices satisfy the more specific following equality. Therefore, for any permutation-invariant matrix π , let $\sigma^\pi \in \mathfrak{S}^\pi$ denote one permutation whose support is of maximum cardinality. (Such a permutation is not necessarily unique, for instance with the affiliation model.) Then,

$$\pi \alpha^{\sigma^\pi} = \pi \alpha \quad . \quad (4)$$

Equation (4) amounts to impose equalities of the coordinates of $\pi \alpha$ in the support of σ^π . Let us recall that the support of σ^π corresponds to rows and columns of π that can be permuted without changing π . Then, assuming all coordinates of $\pi \alpha$ distinct leads to impose that classes with the same connectivity properties have different respective proportions (α_q) to be distinguished between one another.

Proof of Theorem 2.1. First, let $P_{[n]}$ denote the probability distribution function of the adjacency matrix $X_{[n]}$ of SBM. Let us show that there exists a unique (α, π) corresponding to $P_{[n]}$.

Up to reordering, let $r_1 < r_2 < \dots < r_Q$ denote the coordinates of the vector r in the increasing order: r_q is equal to the probability of an edge from a given vertex in the class C_q .

Let R denote the Van der Monde matrix defined by $R_{i,q} = r_q^i$, for $0 \leq i < Q$ and $1 \leq q \leq Q$. R is invertible since the coordinates of r are all different. For $i \geq 1$, $R_{i,q}$ is the probability that i given vertices in C_q have an edge.

Let us also define

$$u_i = \sum_{1 \leq k \leq Q} \alpha_k r_k^i, \quad i = 0, \dots, 2Q - 1 \quad .$$

For $i \geq 1$, u_i denotes the probability that the first i coefficients of the first row of $X_{[n]}$ are equal to 1. Note that $n \geq 2Q$ is a necessary requirement on n since $X_{ii} = 0$ by assumption. Hence given $P_{[n]}$, $u_0 = 1$ and u_1, \dots, u_{2Q-1} are known.

Futhermore, set M the $(Q + 1) \times Q$ matrix given by $M_{i,j} = u_{i+j}$ for every $0 \leq i \leq Q$ and $0 \leq j < Q$, and let M_i denote the square matrix obtained by removing the row i from M . The coefficients of M_Q are

$$M_{i,j} = u_{i+j} = \sum_{1 \leq k \leq Q} r_k^i \alpha_k r_k^j, \quad \text{with } 0 \leq i, j < Q \quad .$$

Defining the diagonal matrix $A = \text{Diag}(\alpha)$, it comes that $M_Q = R A R^t$, where R and A are invertible, but unknown at this stage. With $D_k = \det(M_k)$ and

the polynomial $B(x) = \sum_{k=0}^Q (-1)^{k+Q} D_k x^k$, it yields $D_Q = \det(M_Q) \neq 0$ and the degree of B is equal to Q .

Set $V_i = (1, r_i, \dots, r_i^Q)^t$ and let us notice that $B(r_i)$ is the determinant of the square matrix produced when appending V_i as last column to M . The $Q+1$ columns of this matrix are linearly dependent, since they are all linear combinations of the Q vectors V_1, V_2, \dots, V_Q . Hence $B(r_i) = 0$ and r_i is a root of B for every $1 \leq i \leq Q$. This proves that $B = D_Q \prod_{i=1}^Q (x - r_i)$. Then, one knows $r = (r_1, \dots, r_Q)$ (as the roots of B defined from M) and R . It results that $A = R^{-1} M_Q (R^t)^{-1}$, which yields a unique $(\alpha_1, \dots, \alpha_Q)$.

It only remains to determine π . For $0 \leq i, j < Q$, let us introduce $U_{i,j}$ the probability that the first row of $X_{[n]}$ begins with $i+1$ occurrences of 1, and the second row of X ends up with j occurrences of 1 ($i+1+j \leq n-1$ implies $n \geq 2Q$).

Then, $U_{i,j} = \sum_{k,l} r_k^i \alpha_k \pi_{k,l} \alpha_l r_l^j$, for $0 \leq i, j < Q$, and the $Q \times Q$ matrix $U = RA\pi AR^t$. The conclusion results from $\pi = A^{-1} R^{-1} U (R^t)^{-1} A^{-1}$.

□

The assumption of Theorem 2.1 on r (r' or r''), leading to *generic identifiability*, can be further relaxed in the particular case where $n = 4$ and $Q = 2$.

Theorem 2.2. *Set $n = 4$, $Q = 2$ and let us assume that $\alpha_q > 0$ for every $1 \leq q \leq Q$, and the coefficients of π are not all equal. Then, SBM is identifiable.*

The proof of this result is deferred to Appendix A.

Note that when $Q = 2$, (A2) implies that the coefficients of π are not all equal.

3. Maximum-likelihood estimation of SBM parameters

3.1. Asymptotics of $\mathbb{P}(Z_{[n]} = \cdot \mid X_{[n]})$

In this section we study the *a posteriori* probability distribution function of $Z_{[n]}$, $\mathbb{P}(Z_{[n]} = \cdot \mid X_{[n]})$, which is a random variable depending on $X_{[n]}$.

3.1.1. Equivalence classes between label sequences

Let us consider a realization of the SBM random graph generated with the sequence of true labels $Z = z^*$, where $z^* = \{z_i^*\}_{i \in \mathbb{N}^*}$.

Since a given matrix π can be *permutation-invariant* (see Example 2 Section 2.3), the mapping $z \mapsto \{\pi_{z_i, z_j}\}_{i, j \in \mathbb{N}^*}$ can be non injective. To remedy this problem, let us introduce an equivalence relation between two sequences z and z' of labels:

$$z \stackrel{\pi}{\sim} z' \Leftrightarrow \exists \sigma \in \mathfrak{S}^{\pi} \mid z'_i = \sigma(z_i), \quad \forall i \in \mathbb{N}^* .$$

Then, $[z]_{\pi} = [z']_{\pi}$, where $[z]_{\pi}$ denotes the equivalence class of z .

From now on, square-brackets in the equivalence class notation will be removed to simplify the reading as long as no confusion can be made. In such cases, z will be understood as the equivalence class of the label sequence z .

3.1.2. Main asymptotic result

Let $P^* := \mathbb{P}(\cdot \mid Z = z^*)$ denotes the conditional distribution given the (equivalence class of the) whole label sequence.

The following Theorem 3.1 provides the convergence rate of $\mathbb{P}(Z_{[n]} = z_{[n]}^* \mid X_{[n]})$ towards 1 with respect to P^* , that is given $Z = z^*$. It is an important result that will be repeatedly used along the paper.

Theorem 3.1. *Let us assume that assumptions (A2)–(A5) hold. For every $t > 0$,*

$$P^* \left[\sum_{z_{[n]} \neq z_{[n]}^*} \frac{\mathbb{P}(Z_{[n]} = z_{[n]} \mid X_{[n]})}{\mathbb{P}(Z_{[n]} = z_{[n]}^* \mid X_{[n]})} > t \right] = \mathcal{O}(ne^{-\kappa n}) ,$$

where $\kappa > 0$ is a constant depending on π^* , but not on z^* .

The proof of Theorem 3.1 is deferred to Appendix B.

A noticeable feature of this result is that the convergence rate does not depend on z^* . This point turns out to be crucial when deriving results for the MLE and the variational estimator (respectively Section 3.2 and Section 4.2). Besides, the exponential bound of Theorem 3.1 allows the use of Borel-Cantelli's lemma to get the P^* -almost sure convergence.

Corollary 3.2. *With the same notation as Theorem 3.1,*

$$\sum_{z_{[n]} \neq z_{[n]}^*} \frac{\mathbb{P}(Z_{[n]} = z_{[n]} | X_{[n]})}{\mathbb{P}(Z_{[n]} = z_{[n]}^* | X_{[n]})} \xrightarrow{n \rightarrow +\infty} 0, \quad P^* - a.s. .$$

Moreover,

$$\mathbb{P}(Z_{[n]} = z_{[n]}^* | X_{[n]}) \xrightarrow{n \rightarrow +\infty} 1, \quad P^* - a.s. ,$$

and for every $z_{[n]} \neq z_{[n]}^*$,

$$\mathbb{P}(Z_{[n]} = z_{[n]} | X_{[n]}) \xrightarrow{n \rightarrow +\infty} 0, \quad P^* - a.s. .$$

According to Corollary 3.2, observing the connections between the edges of the realization of the graph as n grows provides us enough information to be able to (asymptotically) recover, given $Z = z^*$, the sequence labels z^* used to generate this graph.

As a consequence of previous Corollary 3.2, one can also understand the above phenomenon in terms of the conditional distribution of the equivalence class $Z_{[n]}$ given $X_{[n]}$.

Corollary 3.3.

$$\mathcal{D}(Z_{[n]} | X_{[n]}) \xrightarrow[n \rightarrow +\infty]{w} \delta_{z^*}, \quad P^* - a.s. ,$$

where $\mathcal{D}(Z_{[n]} | X_{[n]})$ denotes the distribution of $Z_{[n]}$ given $X_{[n]}$, $\xrightarrow[n \rightarrow +\infty]{w}$ refers to the weak convergence in $\mathcal{M}_1(\mathcal{Z})$, the set of probability measures on $\mathcal{E}(\mathcal{Z})$ the set of equivalence classes on $\mathcal{Z} = \{1, \dots, Q\}^{\mathbb{N}}$ and δ_{z^*} is the Dirac measure at the equivalence class z^* .

Proof of Corollary 3.3. For every $n \in \mathbb{N}^*$, let us define $\mathcal{Z}_n = \{1, \dots, Q\}^n$ and $\mathcal{E}(\mathcal{Z}_n)$ the corresponding set of equivalence classes. Let us further introduce a metric space $(\mathcal{E}(\mathcal{Z}_n), d_n)$, where the distance d_n is given by

$$\forall z, z' \in \mathcal{E}(\mathcal{Z}_n), \quad d_n(z, z') = \min_{u \in z, v \in z'} \sum_{k=1}^n 2^{-k} \mathbf{1}_{(u_k \neq v_k)} .$$

Similarly, $(\mathcal{E}(\mathcal{Z}), d)$ denotes a metric space with

$$\forall z, z' \in \mathcal{E}(\mathcal{Z}), \quad d(z, z') = \min_{u \in z, v \in z'} \sum_{k \geq 1} 2^{-k} \mathbf{1}_{(u_k \neq v_k)} .$$

Then, $\mathcal{E}(\mathcal{Z}_n)$ can be embedded into $\mathcal{E}(\mathcal{Z})$ and $\mathcal{E}(\mathcal{Z}_n)$ can be identified to a subset of $\mathcal{E}(\mathcal{Z})$.

Let us introduce \mathcal{B} the Borel σ -algebra on $\mathcal{E}(\mathcal{Z})$, and \mathcal{B}_n the σ -algebra induced by \mathcal{B} on $\mathcal{E}(\mathcal{Z}_n)$. Let also $\mathbb{P}^n = P[\cdot | X_{[n]}]$ denote a probability measure on \mathcal{B} , and $\mathbb{E}_n[\cdot]$ is the expectation with respect to \mathbb{P}^n .

Set $h \in C_b(\mathcal{Z})$ (continuous bounded functions on $\mathcal{E}(\mathcal{Z})$) such that $\|h\|_\infty \leq M$ for $M > 0$. By continuity at point z^* , for every $\epsilon > 0$, there exists $\eta > 0$ such that

$$d(z, z^*) \leq \eta \quad \Rightarrow \quad |h(z^*) - h(z)| \leq \epsilon .$$

Then,

$$\begin{aligned} |\mathbb{E}_n [h (Z_{[n]})] - h(z^*)| &\leq \sum_{z_{[n]}} |h(z_{[n]}) - h(z^*)| \mathbb{P}^n (Z_{[n]} = z_{[n]}) \\ &\leq \epsilon + 2M \sum_{z_{[n]} \in (B_\eta^*)^c} \mathbb{P}^n (Z_{[n]} = z_{[n]}) \\ &\leq \epsilon + o_{\mathbb{P}}(1) \quad P^* - a.s. , \end{aligned}$$

where $B_\eta^* = B(z^*, \eta)$ denotes the ball in $\mathcal{E}(\mathcal{Z})$ with radius η with respect to d . In the last inequality, $o_{\mathbb{P}}(1)$ results from Corollary 3.2, which yields the result. \square

3.2. MLE consistency

The main focus of this section is to settle the consistency of the MLE of (α, π) .

Let us start by recalling the SBM log-likelihood (1):

$$\mathcal{L}_2(X_{[n]}; \alpha, \pi) = \log \left(\sum_{z_{[n]}} e^{\mathcal{L}_1(X_{[n]}; z_{[n]}, \pi)} \mathbb{P} [Z_{[n]} = z_{[n]}] \right) ,$$

where $\mathbb{P} [Z_{[n]} = z_{[n]}] = \prod_{i=1}^n \alpha_{z_i}$, and (α, π) are the SBM parameters. Note that $\mathcal{L}_2(X_{[n]}; \alpha, \pi)$ is an involved expression to deal with.

First, the $X_{i,j}$ s are not independent, which strongly differs from usual statistical settings. For this reason, no theoretical result has ever been derived for the MLE of SBM parameters.

Second, another non standard feature is the number of random variables which is $n(n-1)$ (and not n as usual). More precisely, there are $n(n-1)$ edges $X_{i,j}$ s but only n vertices. This unusual scaling difference implies a refined treatment of the normalizing constants n and $n(n-1)$, depending on the estimated parameter α and π respectively. As a consequence, the MLE consistency proof has been split into two parts.

First, the consistency of the π estimator is addressed by use of an approach based on M-estimators. Second, a result similar to Theorem 3.1 is combined with a “deconditioning” argument to get the expected consistency of the α estimator.

The consistency of the MLE of π strongly relies on a general theorem which is inspired from that for M-estimators (van der Vaart and Wellner, 1996).

Theorem 3.4. *Let (Θ, d) and (Ψ, d') denote metric spaces, and let $M_n : \Theta \times \Psi \rightarrow \mathbb{R}$ be a random function and $\mathbb{M} : \Theta \rightarrow \mathbb{R}$ a deterministic one such that for every $\epsilon > 0$,*

$$\sup_{d(\theta, \theta_0) \geq \epsilon} \mathbb{M}(\theta) < \mathbb{M}(\theta_0) \quad , \quad (5)$$

$$\sup_{(\theta, \psi) \in \Theta \times \Psi} |M_n(\theta, \psi) - \mathbb{M}(\theta)| := \|M_n - \mathbb{M}\|_{\Theta \times \Psi} \xrightarrow[n \rightarrow +\infty]{P} 0 \quad . \quad (6)$$

Moreover, set $(\widehat{\theta}, \widehat{\psi}) = \text{Argmax}_{\theta, \psi} M_n(\theta, \psi)$. Then,

$$d(\widehat{\theta}, \theta_0) \xrightarrow[n \rightarrow +\infty]{P} 0 \quad .$$

One important difference between Theorem 3.4 and its usual counterpart for M-estimators (van der Vaart and Wellner, 1996) is that M_n and \mathbb{M} do not depend on the same number of arguments. Our consistency result for the MLE of π strongly relies on this point.

Proof of Theorem 3.4. For every $\eta > 0$, there exists $\delta > 0$ such that

$$P \left[d(\widehat{\theta}, \theta_0) \geq \eta \right] \leq P \left[\mathbb{M}(\widehat{\theta}) \leq \mathbb{M}(\theta_0) - 3\delta \right] \quad .$$

Since $\|M_n - \mathbb{M}\|_{\Theta \times \Psi} \xrightarrow[n \rightarrow +\infty]{P} 0$, it comes that for large enough values of n ,

$$\begin{aligned} P \left[d(\widehat{\theta}, \theta_0) \geq \eta \right] &\leq P \left[M_n(\widehat{\theta}, \widehat{\psi}) \leq M_n(\theta_0, \psi_0) - \delta \right] + o(1) \\ &\leq o(1) \quad . \end{aligned}$$

□

The leading idea in what follows is to check the assumptions of Theorem 3.4.

The main point of our approach consists in using P^* as a reference probability measure, that is, working as if $Z_{[n]} = z_{[n]}^*$ were known. In this setting, a key quantity is

$$\mathcal{L}_1(X_{[n]}; z_{[n]}, \pi) = \sum_{i \neq j} X_{i,j} \log \pi_{z_i, z_j} + (1 - X_{i,j}) \log(1 - \pi_{z_i, z_j}) \quad ,$$

where $(z_{[n]}, \pi)$ are interpreted as parameters. For any $(z_{[n]}, \pi)$, let us introduce

$$\begin{aligned} \phi_n(z_{[n]}, \pi) &:= \frac{1}{n(n-1)} \mathcal{L}_1(X_{[n]}; z_{[n]}, \pi) \quad , \\ \Phi_n(z_{[n]}, \pi) &:= \mathbb{E} \left[\phi_n(z_{[n]}, \pi) \mid Z_{[n]} = z_{[n]}^* \right] \quad . \end{aligned}$$

Only a subset of the whole set of possible $(z_{[n]}, \pi)$ will be considered. Therefore, let

$$\mathcal{P} = \{(z_{[n]}, \pi) \mid |\Phi_n(z_{[n]}, \pi)| < +\infty\}$$

denote the set of *admissible* parameters. Indeed, any $(z_{[n]}, \pi) \in \mathcal{P}^c$ leads to $|\mathcal{L}_1(X_{[n]}; z_{[n]}, \pi)| = +\infty$, with large probability as $n \rightarrow +\infty$. In other words, this $(z_{[n]}, \pi)$ does not matter in the sum in the expression (1). Note that any $(z_{[n]}, \pi) \in \mathcal{P}$ satisfies for every (i, j) ,

$$\pi_{z_i, z_j} \in \{0, 1\} \Rightarrow \pi_{z_i, z_j} = \pi_{z_i^*, z_j^*}^* = \mathbb{E} \left[X_{i,j} \mid Z_{[n]} = z_{[n]}^* \right] .$$

The following Proposition 3.5 settles a uniform convergence result for $\phi_n - \Phi_n$. Its proof, which is deferred to Appendix C, strongly relies on Talagrand's (Massart, 2007) concentration inequality. This is a consequence of the uniform requirement, at least with respect to π .

Proposition 3.5. *With the above notation, let us assume that (A3) holds. Then,*

$$\sup_{\mathcal{P}} |\phi_n(z_{[n]}, \pi) - \Phi_n(z_{[n]}, \pi)| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0 .$$

The following theorem settles the desired properties for $\mathcal{L}_2(X_{[n]}; \alpha, \pi)$, that is (5) (uniform convergence) and (6) (well-identifiability). The proof of uniform convergence exploits the connection between $\phi_n(z_{[n]}, \pi)$ and $\mathcal{L}_2(X_{[n]}; \alpha, \pi)$ and subsequently Proposition 3.5.

Theorem 3.6. *Let us assume that (A3) and (A4) hold, and for every (α, π) , set $M_n(\alpha, \pi) = [n(n-1)]^{-1} \mathcal{L}_2(X_{[n]}; \alpha, \pi)$, and*

$$\begin{aligned} & \mathbb{M}(\pi) \\ &= \max_{\{a_{i,j}\} \in \mathcal{A}} \left\{ \sum_{q,l} \alpha_q^* \alpha_l^* \sum_{q',l'} [a_{q,q'} a_{l,l'} \pi_{q,l}^* \log \pi_{q',l'} + (1 - \pi_{q,l}^*) \log(1 - \pi_{q',l'})] \right\} , \end{aligned}$$

where (α^*, π^*) denotes the true parameter of SBM, and $\mathcal{A} = \left\{ A = (a_{i,j})_{1 \leq i,j \leq Q} \mid a_{q,q'} \geq 0, \sum_{q'} a_{q,q'} = 1 \right\} \subset \mathcal{M}_Q(\mathbb{R})$. Then for any $\eta > 0$,

$$\begin{aligned} & \sup_{d(\pi, \pi^*) \geq \eta} \mathbb{M}(\pi) < \mathbb{M}(\pi^*) , \\ & \sup_{\alpha, \pi} |M_n(\alpha, \pi) - \mathbb{M}(\pi)| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0 , \end{aligned}$$

where d denotes a distance.

The proof of Theorem 3.6 is given in Appendix D.

Let us now deduce the Corollary 3.7, which asserts the consistency of the MLE of π .

Corollary 3.7. *Under the same assumptions as Theorem 3.6, let us define the MLE of (α, π)*

$$(\hat{\alpha}, \hat{\pi}) := \text{Argmax}_{(\alpha, \pi)} \mathcal{L}_2(X_{[n]}; \alpha, \pi) .$$

Then for any distance $d(\cdot, \cdot)$ on the set of parameters π ,

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

Proof of Corollary 3.7. This is a straightforward consequence of Theorem 3.4 and Theorem 3.6. \square

A quick inspection of the proof of uniform convergence in Theorem 3.6 shows that the asymptotic behavior of log-likelihood \mathcal{L}_2 does not depend on α . Roughly speaking, this results from the expression of \mathcal{L}_2 in which the number of terms involving π is of order n^2 , whereas only n terms involve α . This difference of scaling with respect to n between π and α justifies to some extent a different approach for the MLE of α .

The proposed strategy heavily relies on an analogous result to Theorem 3.1, where the true value (α^*, π^*) of SBM parameters is replaced by an estimator $(\hat{\alpha}, \hat{\pi})$. Let us first state this result in a general framework. It will be also used in Section 4.

Proposition 3.8. *Let us assume that assumptions (A2)–(A4), and (A5) hold, and that there exists an estimator $\hat{\pi} = \pi^* + o_{\mathbb{P}}(v_n)$, with $nv_n = \mathcal{O}(1)$. Let also $\hat{\alpha}$ denote any estimator of α^* . Then for any $t > 0$,*

$$P^* \left[\sum_{z_{[n]} \neq z_{[n]}^*} \frac{\hat{P}(Z_{[n]} = z_{[n]} | X_{[n]})}{\hat{P}(Z_{[n]} = z_{[n]}^* | X_{[n]})} > t \right] = \mathcal{O}(ne^{-\kappa n} \vee \mathbb{P}[\|\hat{\pi} - \pi^*\|_{\infty} > v_n]) ,$$

where $\kappa > 0$ is a constant depending on π^* , and

$$\begin{aligned} & \log \left(\frac{\hat{P}(Z_{[n]} = z_{[n]} | X_{[n]})}{\hat{P}(Z_{[n]} = z_{[n]}^* | X_{[n]})} \right) \\ &= \sum_{i \neq j} \left\{ X_{ij} \log \left(\frac{\hat{\pi}_{z_i, z_j}}{\hat{\pi}_{z_i^*, z_j^*}} \right) + (1 - X_{ij}) \log \left(\frac{1 - \hat{\pi}_{z_i, z_j}}{1 - \hat{\pi}_{z_i^*, z_j^*}} \right) \right\} + \sum_i \log \frac{\hat{\alpha}_{z_i}}{\hat{\alpha}_{z_i^*}} . \end{aligned}$$

The proof of Proposition 3.8 is given in Appendix E.

Note that the novelty of this result compared to Theorem 3.1 lies in the convergence rate which depends on the behaviour of $\hat{\pi}$. This is the reliable price for estimating rather than knowing π^* . The requirement on the rate of convergence v_n of $\hat{\pi}$ arises from the proof as a necessary requirement to get the convergence in probability toward 0 as n tends to $+\infty$. However, we do not know whether this artificially results from the strategy of proof or whether this is essentially a necessary condition.

There is empirical evidence (see Gazal *et al.*, 2011) that the rate of convergence of $\hat{\pi}$ is of order $1/n$, but this property is assumed and not proved in this paper. Besides, in the same way as in Theorem 3.1, one crucial point in Proposition 3.8 is the independence of the convergence rate with respect to $z_{[n]}^*$. The strategy of proof presented in the sequel strongly relies on this property.

Let us now settle the consistency of the MLE of α^* on the basis of previous Proposition 3.8.

Theorem 3.9. *Let us assume the MLE $\hat{\pi} = \pi^* + o_{\mathbb{P}}(1/n)$. With the same assumptions as Theorem 3.6, and the notation of Corollary 3.7, then*

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

where d denotes any distance between vectors in \mathbb{R}^Q .

Proof of Theorem 3.9. As usually for mixture models, it is easy to see that the MLE of α is given for any q by

$$\hat{\alpha}_q = \frac{1}{n} \sum_{i=1}^n \hat{P}(Z_i = q \mid X_{[n]}) \quad .$$

First, let us work with respect to P^* , that is, as if $Z_{[n]} = z_{[n]}^*$ were known. Setting $N_q(z_{[n]}) = \sum_{i=1}^n \mathbb{1}_{(z_i=q)}$, it comes

$$\begin{aligned} \left| \hat{\alpha}_q - N_q(z_{[n]}^*)/n \right| &\leq \left| \frac{1}{n} \sum_{i=1}^n \hat{P}(Z_i = z_i^* \mid X_{[n]}) \mathbb{1}_{(z_i^*=q)} - N_q(z_{[n]}^*)/n \right| \\ &\quad + \hat{P}(Z_{[n]} \neq z_{[n]}^* \mid X_{[n]}) \\ &\leq \frac{1}{n} \sum_{i=1}^n \left(1 - \hat{P}(Z_i = z_i^* \mid X_{[n]}) \right) \mathbb{1}_{(z_i^*=q)} \\ &\quad + \hat{P}(Z_{[n]} \neq z_{[n]}^* \mid X_{[n]}) \\ &\leq \frac{1}{n} \sum_{i=1}^n \left(\hat{P}(Z_i \neq z_i^* \mid X_{[n]}) \right) \mathbb{1}_{(z_i^*=q)} + \hat{P}(Z_{[n]} \neq z_{[n]}^* \mid X_{[n]}) \\ &\leq 2\hat{P}(Z_{[n]} \neq z_{[n]}^* \mid X_{[n]}) \quad . \end{aligned}$$

Second, let us now use a deconditioning argument replacing P^* by \mathbb{P} . Let $N_q = N_q(Z_{[n]})$ denote a binomial random $\mathcal{B}(n, \alpha_q^*)$ for every q . Then for every $\epsilon > 0$,

$$\begin{aligned} &\mathbb{P} \left[|\hat{\alpha}_q - \alpha_q^*| > \epsilon \right] \\ &\leq \mathbb{P} \left[|\hat{\alpha}_q - N_q/n| > \epsilon/2 \right] + \mathbb{P} \left[|N_q/n - \alpha_q^*| > \epsilon/2 \right] \\ &\leq \mathbb{P} \left[|\hat{\alpha}_q - N_q/n| > \epsilon/2 \right] + o(1) \quad , \end{aligned}$$

by use of LLG. Finally, a straightforward use of Proposition 3.8 leads to

$$\begin{aligned}
 & \mathbb{P} [|\hat{\alpha}_q - N_q/n| > \epsilon/2] \\
 &= \mathbb{E}_{Z_{[n]}} [P (|\hat{\alpha}_q - N_q/n| > \epsilon/2 \mid Z_{[n]})] \\
 &\leq \sum_{z_{[n]}} P \left[\hat{P} (Z_{[n]} \neq z_{[n]} \mid X_{[n]}) > \epsilon/4 \mid Z_{[n]} = z_{[n]} \right] \mathbb{P} [Z_{[n]} = z_{[n]}] \\
 &= o(1) .
 \end{aligned}$$

□

4. Variational estimators of SBM parameters

In Section 3, consistency has been proved for the maximum likelihood estimators. However, this result is essentially theoretical since in practice the MLE can only be computed for very small graphs (with less than 20 vertices). Nevertheless, such results for the MLE are useful in at least two respects. First from a general point of view, they provide a new strategy to derive consistency of estimators obtained from likelihoods in non-*i.i.d.* settings. Second in the framework of the present paper, these results are exploited to settle the consistency of the variational estimators.

The main interest of *variational estimators* in SBM is that unlike the MLE ones, they are *useful in practice* since they enable to deal with very huge graphs (several thousands of vertices). Indeed, the log-likelihood $\mathcal{L}_2(X_{[n]}; \alpha, \pi)$ involves a sum over Q^n terms, which is intractable except for very small and unrealistic values of n :

$$\mathcal{L}_2(X_{[n]}; \alpha, \pi) = \log \left\{ \sum_{z_{[n]} \in \mathcal{Z}_n} e^{\sum_{i \neq j} b_{ij}(z_i, z_j)} P_{Z_{[n]}}(z_{[n]}) \right\},$$

with $b_{ij}(z_i, z_j) = X_{ij} \log \pi_{z_i, z_j} + (1 - X_{ij}) \log(1 - \pi_{z_i, z_j})$. To circumvent this problem, alternatives are Markov chain Monte Carlo (MCMC) algorithms (Andrieu and Atchadé, 2007) and variational approximation (Jordan *et al.*, 1999). However, MCMC algorithms suffer a high computational cost, which makes them unattractive compared to variational approximation. Actually the variational method is the only one which can deal with thousands of vertices in a reasonable computation time thanks to its complexity in $\mathcal{O}(n^2)$. For instance, the Mixnet (2009) package (based on variational approximation) works with up to 2000 vertices, whereas the STOCNET package (see Boer *et al.*, 2006) (Gibbs sampling) only deals with 200 vertices.

The purpose of the present section is to prove that the variational approximation yields consistent estimators of the SBM parameters. The resulting estimators will be called *variational estimators* (VE).

4.1. Variational approximation

To the best of our knowledge, the first use of variational approximation for SBM has been made by Daudin *et al.* (2008). The variational method consists in approximating $P^{X_{[n]}} = \mathbb{P}(Z_{[n]} = \cdot | X_{[n]})$ by a product of n multinomial distributions. This leads to approximate $\mathcal{L}_2(X_{[n]}; \alpha, \pi)$ by a sum of n^2 terms. The computational virtue of this trick is that a sum over Q^n terms is replaced by a sum over n^2 terms.

Let us define \mathcal{D}_n as a set of product multinomial distributions

$$\mathcal{D}_n = \left\{ D_{\tau_{[n]}} = \prod_{i=1}^n \mathcal{M}(1, \tau_{i,1}, \dots, \tau_{i,Q}) \mid \tau_{[n]} \in \mathcal{S}_n \right\}, \quad (7)$$

where

$$\mathcal{S}_n = \left\{ \tau_{[n]} = (\tau_1, \dots, \tau_n) \in ([0, 1]^Q)^n \mid \forall i, \tau_i = (\tau_{i,1}, \dots, \tau_{i,Q}), \sum_{q=1}^Q \tau_{i,q} = 1 \right\} .$$

For any $D_{\tau_{[n]}} \in \mathcal{D}_n$, the variational log-likelihood, $\mathcal{J}(\cdot; \cdot, \cdot, \cdot)$ is defined by

$$\mathcal{J}(X_{[n]}; \tau_{[n]}, \alpha, \pi) = \mathcal{L}_2(X_{[n]}; \alpha, \pi) - K(D_{\tau_{[n]}}; P^{X_{[n]}}) , \quad (8)$$

where $K(\cdot, \cdot)$ denotes the Kullback-Leibler divergence, and $P^{X_{[n]}} = \mathbb{P}(Z_{[n]} = \cdot \mid X_{[n]})$. With this choice of \mathcal{D}_n , $\mathcal{J}(X_{[n]}; \tau_{[n]}, \alpha, \pi)$ has the following expression (see [Daudin et al., 2008](#)):

$$\mathcal{J}(X_{[n]}; \tau_{[n]}, \alpha, \pi) = \sum_{i \neq j} \sum_{q,l} b_{ij}(q,l) \tau_{i,q} \tau_{j,l} - \sum_{iq} \tau_{i,q} (\log \tau_{i,q} - \log \alpha_q) . \quad (9)$$

Then, the variational approximation $R_{X_{[n]}}$ to $P^{X_{[n]}}$ is given by solving the minimization problem over \mathcal{D}_n :

$$R_{X_{[n]}} \in \operatorname{Argmin}_{D_{\tau} \in \mathcal{D}_n} K(D_{\tau}, P^{X_{[n]}}) ,$$

as long as such a minimizer exists.

Thus, minimizing $K(D_{\tau_{[n]}}; P^{X_{[n]}})$ with respect to $\tau_{[n]}$ is equivalent to maximizing $\mathcal{J}(X_{[n]}; \tau_{[n]}, \pi, \alpha)$, which leads to

$$\hat{\tau}_{[n]} = \hat{\tau}_{[n]}(\pi, \alpha) := \operatorname{Argmax}_{\tau_{[n]}} \mathcal{J}(X_{[n]}; \tau_{[n]}, \alpha, \pi) .$$

The variational estimators (VE) of (α, π) are

$$(\tilde{\alpha}, \tilde{\pi}) = \operatorname{Argmax}_{\alpha, \pi} \mathcal{J}(X_{[n]}; \hat{\tau}_{[n]}, \alpha, \pi) . \quad (10)$$

Note that in practice, the variational algorithm maximizes $\mathcal{J}(X_{[n]}; \tau, \alpha, \pi)$ alternatively with respect to τ and (α, π) (see [Daudin et al., 2008](#)).

In the sequel, the same notation as in Section 3 is used. In particular, it is often assumed that a realization of SBM is observed, which has been generated from the sequence of true labels $Z = z^*$. In this setting, P^* denotes the conditional distribution $\mathbb{P}(\cdot \mid Z = z^*)$ given the whole label sequence. The first result provides some assurance about the reliability of the variational approximation to $P^{X_{[n]}}$.

Proposition 4.1. *For every n , let \mathcal{D}_n denote the set defined by (7), and $P^{X_{[n]}}(\cdot)$ be the distribution of $Z_{[n]}$ given $X_{[n]}$. Then, assuming (A2) – –(A4) hold,*

$$K(R_{X_{[n]}}; P^{X_{[n]}}) := \inf_{D \in \mathcal{D}_n} K(D, P^{X_{[n]}}) \xrightarrow{n \rightarrow \infty} 0 , \quad P^* - a.s. .$$

Note that this convergence result is given with respect to P^* (and not to \mathbb{P}). Stronger results can be obtained (see Section 4.1) thanks to fast convergence rates. Proposition 4.1 yields some confidence in the reliability of the variational approximation, which gets closer to $P^{X_{[n]}}$ as n tends to $+\infty$. However, it does not provide any warranty about the good behavior of variational estimators, which is precisely the goal of following Section 4.2.

Proof of Proposition 4.1.

By definition of the variational approximation,

$$K(R_{X_{[n]}}, P^{X_{[n]}}) \leq K(\delta_{z_{[n]}^*}, P^{X_{[n]}}) ,$$

where $\delta_{z_{[n]}^*} = \prod_{1 \leq i \leq n} \delta_{z_i^*} \in \mathcal{D}_n$. Then,

$$K(R_{X_{[n]}}, P^{X_{[n]}}) \leq \sum_{1 \leq i \leq n} -\log(P(Z_i = z_i^* | X_{[n]})) = -\log \left[P \left(Z_{[n]} = z_{[n]}^* | X_{[n]} \right) \right] .$$

The conclusion results from Theorem 3.1, and Corollary 3.2 since $P(Z_{[n]} = z_{[n]}^* | X_{[n]}) \xrightarrow[n \rightarrow \infty]{} 1$ P^* -a.s.. \square

4.2. Consistency of the variational estimators

Since the variational log-likelihood $\mathcal{J}(\cdot; \cdot, \cdot, \cdot)$ (8) is defined from the log-likelihood $\mathcal{L}_2(\cdot; \cdot, \cdot)$, the properties of $\mathcal{J}(X_{[n]}; \tau_{[n]}, \alpha, \pi)$ are strongly connected to those of $\mathcal{L}_2(X_{[n]}; \alpha, \pi)$. Therefore, the strategy followed in the present section is very similar to that of Section 3. In particular, the consistency of $\tilde{\pi}$ (VE of π , see (10)) is addressed first. Then, the consistency of the VE of α ($\tilde{\alpha}$, see (10)) exploits the convergence of the estimator of π .

The first step consists in applying Theorem 3.4 to settle the $\tilde{\pi}$ consistency. Following results aim at justifying the use of this theorem by checking its assumptions.

Theorem 4.2 states that \mathcal{L}_2 and \mathcal{J} are asymptotically equivalent uniformly with respect to α and π .

Theorem 4.2. *With the same notation as Theorem 3.6 and Section 4.1, let us define*

$$J_n(\alpha, \pi) := \frac{1}{n(n-1)} \mathcal{J}(X_{[n]}; \hat{\tau}_{[n]}, \alpha, \pi) .$$

Then, (A3) and (A4) yield

$$\sup_{\alpha, \pi} \{|J_n(\alpha, \pi) - M_n(\alpha, \pi)|\} = o(1), \quad \mathbb{P}\text{-a.s.} ,$$

where the supremum is computed over sets given in (A3) and (A4).

This statement is stronger than Proposition 4.1 in several respects. On the one hand, convergence applies almost surely with respect to \mathbb{P} , and not with respect to P^* . On the other hand, Theorem 4.2 exhibits the convergence rate toward 0, which is not faster than $n(n-1)$.

Proof of Theorem 4.2.

Recall the definitions of \mathcal{L}_1 (2), \mathcal{L}_2 (1), \mathcal{J} (8) and let $\hat{z}_{[n]} = \hat{z}_{[n]}(\pi) = \text{Argmax}_{z_{[n]}} \mathcal{L}_1(X_{[n]}; z_{[n]}, \pi)$. Lemma F.1 yields

$$\mathcal{J}(X_{[n]}; \tau_{[n]}, \alpha, \pi) \leq \mathcal{L}_2(X_{[n]}; \alpha, \pi) \leq \mathcal{L}_1(X_{[n]}; \hat{z}_{[n]}, \pi) .$$

Then, applying Assumption (A4) and Lemma F.2, there exists $0 < \gamma < 1$ independent of (α, π) such that

$$|\mathcal{J}(X_{[n]}; \hat{\tau}_{[n]}, \alpha, \pi) - \mathcal{L}_1(X_{[n]}; \hat{z}_{[n]}, \pi)| \leq n \log(1/\gamma) .$$

The conclusion results straightforwardly. \square

The consistency of $\tilde{\pi}$ is provided by the following result, which is simple a consequence of Theorem 4.2, Proposition 3.5, and Theorem 3.4.

Corollary 4.3. *With the notation of Theorem 4.2 and assuming (A3) and (A4) hold, let us define the VE of (α, π)*

$$(\tilde{\alpha}, \tilde{\pi}) = \text{Argmax}_{\alpha, \pi} J_n(\alpha, \pi) .$$

Then for any distance $d(\cdot, \cdot)$ on the set of π parameters,

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

The proof is completely similar to that of Corollary 3.7 and is therefore not reproduced here.

The consistency of the VE of α^* results from the same deconditioning argument as the MLE of α^* (Section 3.2). There is some empirical evidence (see Gazal *et al.*, 2011) about the rate $1/n$ of convergence of $\tilde{\pi}$. This rate is assumed to hold in the following theorem.

Theorem 4.4. *Let us assume the VE $\tilde{\pi}$ converges at rate $1/n$ to π^* . With the same assumptions as Theorem 4.2 and assuming (A3) and (A4) hold, then*

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

where d denotes any distance between vectors in \mathbb{R}^Q .

The crux of the proof is an other use of Proposition 3.8.

Proof of Theorem 4.4.

First, let us show that given $Z_{[n]} = z_{[n]}^*$,

$$\left| \tilde{\alpha}_q - N_q(z_{[n]}^*)/n \right| \xrightarrow[n \rightarrow \infty]{P^*} 0 .$$

For every q ,

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

where $\tilde{\tau}_{i,q} = \hat{\tau}_{i,q}(\tilde{\alpha}, \tilde{\pi})$ (see (10)). Introducing z_i^* , it comes that

$$\tilde{\alpha}_q = \frac{1}{n} \sum_{i=1}^n \tilde{\tau}_{i,z_i^*} \mathbf{1}_{(z_i^*=q)} + \frac{1}{n} \sum_{i=1}^n \tilde{\tau}_{i,q} \mathbf{1}_{(z_i^* \neq q)} .$$

From (7), let us consider the *a posteriori* distribution of $\tilde{Z}_{[n]} = (\tilde{Z}_1, \dots, \tilde{Z}_n)$ denoted by

$$D_{\tilde{\tau}_{[n]}}(z_{[n]}) = \mathbb{P} \left[\tilde{Z}_{[n]} = z_{[n]} \mid X_{[n]} \right] = \prod_{i=1}^n \tilde{\tau}_{i,z_i} .$$

Then,

$$\begin{aligned} \left| \tilde{\alpha}_q - N_q(z_{[n]}^*)/n \right| &= \left| \frac{1}{n} \sum_{i=1}^n (\tilde{\tau}_{i,z_i^*} - 1) \mathbf{1}_{(z_i^*=q)} + \frac{1}{n} \sum_{i=1}^n \tilde{\tau}_{i,q} \mathbf{1}_{(z_i^* \neq q)} \right| \\ &\leq \frac{1}{n} \sum_{i=1}^n (1 - \tilde{\tau}_{i,z_i^*}) \mathbf{1}_{(z_i^*=q)} + \frac{1}{n} \sum_{i=1}^n \tilde{\tau}_{i,q} \mathbf{1}_{(z_i^* \neq q)} \\ &\leq \frac{1}{n} \sum_{i=1}^n (1 - \tilde{\tau}_{i,z_i^*}) , \end{aligned}$$

using that when $z_i^* \neq q$, $\tilde{\tau}_{i,q} \leq \sum_{q \neq z_i^*} \tilde{\tau}_{i,q} = 1 - \tilde{\tau}_{i,z_i^*}$. Hence,

$$\left| \tilde{\alpha}_q - N_q(z_{[n]}^*)/n \right| \leq \frac{1}{n} \sum_{i=1}^n \mathbb{P} \left[\tilde{Z}_{[n]} \neq z_{[n]}^* \mid X_{[n]} \right] = 1 - D_{\tilde{\tau}_{[n]}}(z_{[n]}^*) .$$

It remains to show that $D_{\tilde{\tau}_{[n]}}(z_{[n]}^*) \xrightarrow[n \rightarrow \infty]{P^*} 1$, at a rate which does not depend of $z_{[n]}^*$. Let $\tilde{P}(z_{[n]}^*)$ denote the *a posteriori* distribution of $Z_{[n]}$ with parameters $(\tilde{\alpha}, \tilde{\pi})$. According to Lemma 4.5, the asymptotic behavior of $D_{\tilde{\tau}_{[n]}}(z_{[n]}^*)$ is closely related to that of $\tilde{P}(z_{[n]}^*)$. Then, another use of Proposition 3.8 applied to $\tilde{P}(z_{[n]}^*)$ and $\tilde{\pi}$ yields

$$\tilde{P}(z_{[n]}^*) = 1 - \mathcal{O} \left(n e^{-\kappa n} \vee \mathbb{P} \left[\|\tilde{\pi} - \pi^*\|_\infty > v_n \right] \right) ,$$

where $\kappa > 0$ is a constant depending on π^* . Therefore, Lemma 4.5 implies

$$D_{\tilde{\tau}_{[n]}}(z_{[n]}^*) = 1 - \sqrt{\mathcal{O}(ne^{-\kappa n} \vee \mathbb{P}[\|\tilde{\pi} - \pi^*\|_\infty > v_n])} .$$

Result follows from the same reasoning as in the end of the proof of Proposition 3.8. □

Lemma 4.5.

$$\left| D_{\tilde{\tau}_{[n]}}(z_{[n]}^*) - \tilde{P}(z_{[n]}^*) \right| \leq \sqrt{-\frac{1}{2} \log [\tilde{P}(z_{[n]}^*)]} .$$

Proof of Lemma 4.5.

$$\begin{aligned} \left| D_{\tilde{\tau}_{[n]}}(z_{[n]}^*) - \tilde{P}(z_{[n]}^*) \right| &\leq \left\| D_{\tilde{\tau}_{[n]}} - \tilde{P} \right\|_{TV} \\ &\leq \sqrt{\frac{1}{2} K(D_{\tilde{\tau}_{[n]}} , \tilde{P})} \\ &\leq \sqrt{\frac{1}{2} K(\delta_{z_{[n]}^*} , \tilde{P})} = \sqrt{-\frac{1}{2} \log [\tilde{P}(z_{[n]}^*)]} . \end{aligned}$$

□

5. Conclusion

This paper provides theoretical (asymptotic) results about the stochastic block model (SBM) inference, especially applying to directed graphs, unlike most of existing results. This is typically the realistic setting of real applications such as biological networks.

In particular, asymptotic equivalence between maximum-likelihood and variational estimators is proved, as well as the consistency of resulting estimators. To the best of our knowledge, these are the first results of this type for variational estimators of the SBM parameters. Such theoretical properties are essential since they validate the empirical practice which uses variational approaches as a reliable means to deal with up to several thousands of vertices.

Besides, this work can be seen as a preliminary step toward a deeper analysis of maximum-likelihood and variational estimators of SBM parameters. In particular, a further interesting question is the choice of the number Q of classes in the mixture model. Indeed, it is important to develop a *data-driven* strategy to choose Q in order to make the variational approach fully applicable in practice, and validate the empirical practice.

Appendix A: Proof of Theorem 2.2

Proof of Theorem 2.2.

Let us just assume $Q = 2$, $n = 4$, and that no element of α is zero.

If the coordinates of $r = \pi\alpha$ are distinct, then Theorem 2.1 applies and the desired result follows.

Otherwise, the two coordinates are r , r' and r'' are not distinct. Set $r_1 = r_2 = a$ and $u_i = \alpha_1 r_1^i + \alpha_2 r_2^i$, for $i \geq 0$. Let us also define $b = r'_1 = r'_2$, and $c = r''_1 = r''_2$. Then, the following equalities hold:

$$\begin{aligned} a &= \pi_{11}\alpha_1 + \pi_{12}\alpha_2 = \pi_{21}\alpha_1 + \pi_{22}\alpha_2 \quad , \\ b &= \pi_{11}\alpha_1 + \pi_{21}\alpha_2 = \pi_{12}\alpha_1 + \pi_{22}\alpha_2 \quad , \\ c &= \pi_{11}^2\alpha_1 + \pi_{21}\pi_{12}\alpha_2 = \pi_{12}\pi_{21}\alpha_1 + \pi_{22}^2\alpha_2 \quad . \end{aligned}$$

From $a - b = (\pi_{12} - \pi_{21})\alpha_2 = -(\pi_{12} - \pi_{21})\alpha_1$ we deduce $\pi_{12} = \pi_{21}$ and $a = b$. Then,

$$\begin{aligned} \alpha_1\alpha_2(\pi_{11} - \pi_{12})^2 &= (\alpha_1 + \alpha_2)(\alpha_1\pi_{11}^2 + \alpha_2\pi_{12}^2) - (\alpha_1\pi_{11} + \alpha_2\pi_{12})^2 \\ &= c - a^2 \\ &= c - b^2 \\ &= \alpha_1\alpha_2(\pi_{22} - \pi_{12})^2 \quad . \end{aligned}$$

If $c = a^2$, then $\pi_{11} = \pi_{12} = \pi_{21} = \pi_{22} = a$ and α cannot be found.

If $c \neq a^2$, then $|\pi_{11} - \pi_{12}| = |\pi_{22} - \pi_{12}| \neq 0$. But $\alpha_1(\pi_{11} - \pi_{12}) = a - \pi_{12} = b - \pi_{12} = \alpha_2(\pi_{22} - \pi_{12})$ leads to $|\alpha_1| = |\alpha_2|$ and $\alpha_1 = \alpha_2 = 1/2$. Hence $\pi_{11} = \pi_{22}$. Then, π_{11} and π_{12} are the roots of the polynomial $x^2 - 2ax + 2a^2 - c$.

At this stage, we need to distinguish between π_{11} and π_{12} . Let us introduce the probability d that $X_{[n]}$ fits the pattern

$$\begin{array}{cccc} \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot \\ 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{array} .$$

Then, $d = (\pi_{11}^3 + 3\pi_{11}\pi_{12}^2)/4$ and one can compute $e = \sqrt[3]{d - a^3} = (\pi_{11} - \pi_{12})/2$. This leads to $\pi_{11} = \pi_{22} = a + e$ and $\pi_{12} = \pi_{21} = a - e$, which yields the conclusion. \square

Appendix B: Proof of Theorem 3.1 and Proposition 3.8

Let us first give the proof under Assumption (A1) and (A3). The proof in the more general setting of Assumptions (A2), (A3), and (A5) is given in Appendix B.5.

$$\mathbf{B.1. \textit{Upper bounding}} \quad \mathbb{P} \left[\sum_{z_{[n]} \neq z_{[n]}^*} \frac{\mathbb{P}[Z_{[n]}=z_{[n]} | X_{[n]}]}{\mathbb{P}[Z_{[n]}=z_{[n]}^* | X_{[n]}]} > t \mid Z = z^* \right]$$

Assuming (A1) holds, let us notice that π^* cannot be permutation-invariant (see Example 2). Thus, $(S) = \{Id\}$ and the equivalence classes defined in Section 3.1.1 are all singletons. Therefore in the sequel, $z_{[n]}$ denotes the vector and not the associated equivalence class.

Let $P^{X_{[n]}}(z_{[n]})$ denote $\mathbb{P}[Z_{[n]} = z_{[n]} \mid X_{[n]}]$ for every $z_{[n]}$. The sum on $z_{[n]}$ is partitioned according to the number r of differences between $z_{[n]}$ and $z_{[n]}^*$.

$$\sum_{z_{[n]} \neq z_{[n]}^*} \frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} = \sum_{r=1}^n \sum_{\|z - z^*\|_0 = r} \frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)},$$

where $\|z\|_0$ designates the number of non-zero components of the vector z .

$$\begin{aligned} & \mathbb{P} \left[\sum_{z \neq z^*} \frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} > t \mid Z = z^* \right] \\ &= \mathbb{P} \left[\sum_{r=1}^n \sum_{\|z_{[n]} - z_{[n]}^*\|_0 = r} \frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} > t \mid Z = z^* \right] \\ &\leq \mathbb{P} \left[\bigcup_{r=1}^n \left\{ \sum_{\|z_{[n]} - z_{[n]}^*\|_0 = r} \frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} > \frac{t}{n} \mid Z = z^* \right\} \right] \\ &\leq \sum_{r=1}^n \mathbb{P} \left[\sum_{\|z_{[n]} - z_{[n]}^*\|_0 = r} \frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} > \frac{t}{n} \mid Z = z^* \right] \\ &\leq \sum_{r=1}^n \sum_{\|z_{[n]} - z_{[n]}^*\|_0 = r} \mathbb{P} \left[\frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} > \frac{t}{n \binom{n}{r} (Q-1)^r} \mid Z = z^* \right]. \end{aligned}$$

Note that the number of $z_{[n]} \neq z_{[n]}^*$ such that $\|z_{[n]} - z_{[n]}^*\|_0 = r$ is equal to $\binom{n}{r} (Q-1)^r$ since $z_{[n]} \in \{1, \dots, Q\}^n$. This leads us to

$$\begin{aligned} & \mathbb{P} \left[\sum_{z \neq z^*} \frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} > t \mid Z = z^* \right] \\ &\leq \sum_{r=1}^n \sum_{\|z_{[n]} - z_{[n]}^*\|_0 = r} \mathbb{P} \left[\frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} > \frac{t}{n^{r+1} (Q-1)^r} \mid Z = z^* \right]. \end{aligned}$$

$$\mathbf{B.2. Upper bounding} \quad \mathbb{P} \left[\frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} > \frac{t}{n^{r+1}(Q-1)^r} \mid Z = z^* \right]$$

Let us first notice that

$$\begin{aligned} & \mathbb{P} \left[\frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} > \frac{t}{n^{r+1}(Q-1)^r} \mid Z = z^* \right] \\ = & \mathbb{P} \left[\log \frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} > \log \left(\frac{t}{n^{r+1}(Q-1)^r} \right) \mid Z = z^* \right] \\ = & \mathbb{P} \left\{ \frac{1}{r(2n-r-1)} \left(\log \frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} - \mathbb{E}^{Z=z^*} \left[\log \frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} \right] \right) > \right. \\ & \left. \frac{1}{r(2n-r-1)} \left(\log \frac{t}{n^{r+1}(Q-1)^r} - \mathbb{E}^{Z=z^*} \left[\log \frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} \right] \right) \mid Z = z^* \right\}. \end{aligned}$$

Moreover,

$$\begin{aligned} & \log \left(\frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} \right) - \mathbb{E}^{Z=z^*} \left[\log \left(\frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} \right) \right] \\ = & \sum_{i \neq j} \left\{ \left(X_{ij} - \pi_{z_i^*, z_j^*}^* \right) \log \left[\frac{\pi_{z_i, z_j}^* (1 - \pi_{z_i^*, z_j^*}^*)}{\pi_{z_i^*, z_j^*}^* (1 - \pi_{z_i, z_j}^*)} \right] \right\}. \end{aligned}$$

Under condition (A1), Lemma B.4 implies that there are $r(2n-r-1)$ non-zero term in the sum.

Let us then apply Proposition B.1 (Hoeffding's inequality), with $a_{ij} = -b_{ij} = \log \left[(1-\zeta)^2 \zeta^{-2} \right]$ (see Lemma B.2), $L = 2(b_{i,j} - a_{i,j})$. For any $s > 0$, it comes

$$\begin{aligned} & \mathbb{P} \left[\frac{1}{r(2n-r-1)} \left(\log \frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} - \mathbb{E}^{Z=z^*} \left[\log \frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} \right] \right) > s \mid Z = z^* \right] \\ & \leq \exp \left(\frac{-r(2n-r-1)s^2}{L^2} \right). \end{aligned}$$

B.3. Conclusion

One then apply the previous result with a particular choice of s . Thus,

$$s = \frac{1}{r(2n-r-1)} \left(\log \frac{t}{n^{r+1}(Q-1)^r} - \mathbb{E}^{Z=z^*} \left[\log \frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} \right] \right)$$

leads to

$$s = \frac{\log t - (r+1)\log(n) - r\log(Q-1)}{r(2n-r-1)} - \frac{1}{r(2n-r-1)} \mathbb{E}^{Z=z^*} \left[\log \frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} \right].$$

With Lemma B.3, it is not difficult to show that for large enough values of n ,

$$s^2 \geq \frac{3}{4} \left(\frac{1}{r(2n-r-1)} \mathbb{E}^{Z=z^*} \left[\log \frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} \right] \right)^2 \geq \frac{3}{4} (c^*)^2.$$

It results that

$$\exp\left(\frac{-r(2n-r-1)s^2}{L^2}\right) \leq \exp\left(\frac{-3r(2n-r-1)(c^*)^2}{4L^2}\right),$$

which implies

$$\begin{aligned} \mathbb{P} \left[\sum_{z \neq z^*} \frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} > t \mid Z = z^* \right] &\leq \sum_{r=1}^n \sum_{\|z_{[n]} - z_{[n]}^*\|_0 = r} \exp\left(\frac{-3r(2n-r-1)(c^*)^2}{4L^2}\right) \\ &\leq \sum_{r=1}^n \binom{n}{r} (Q-1)^r \exp\left(\frac{-3r(2n-r-1)(c^*)^2}{4L^2}\right) \\ &\leq \sum_{r=1}^n \binom{n}{r} [(Q-1)u_n]^r, \end{aligned}$$

where $u_n = \exp\left[\frac{-3(n-1)(c^*)^2}{4L^2}\right]$.

Using the inequality $e^x - 1 \leq xe^x$, one obtains

$$\begin{aligned} \mathbb{P} \left[\sum_{z \neq z^*} \frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} > t \mid Z = z^* \right] &\leq (1 + (Q-1)u_n)^n - 1 \\ &\leq e^{(Q-1)nu_n} - 1 \\ &\leq (Q-1)u_n n e^{(Q-1)nu_n} \xrightarrow[n \rightarrow +\infty]{} 0, \end{aligned}$$

which concludes the proof.

B.4. Hoeffding's inequality and related lemmas

Proposition B.1 (Hoeffding's inequality). *Let $\{Y_{i,j}\}_{1 \leq i \neq j \leq n}$ independent random variables such that for every $i \neq j$, $Y_{i,j} \in [a_{i,j}, b_{i,j}]$ almost surely. Then, for any $t > 0$,*

$$\mathbb{P} \left[\sum_{i \neq j}^n (Y_{i,j} - \mathbb{E}[Y_{i,j}]) > t \right] \leq \exp\left(\frac{-t^2}{\sum_{i \neq j} (b_{i,j} - a_{i,j})^2}\right).$$

Lemma B.2 (Values of $a_{i,j}$ and $b_{i,j}$). *Assuming (A3) holds for π^* with $\zeta > 0$, it comes for every $1 \leq i \neq j \leq n$,*

$$\left| X_{ij} \log \left[\frac{\pi_{z_i, z_j}^* (1 - \pi_{z_i^*, z_j^*}^*)}{\pi_{z_i^*, z_j^*}^* (1 - \pi_{z_i, z_j}^*)} \right] \right| \leq 2 \log \left[\left(\frac{1 - \zeta}{\zeta} \right) \right].$$

Lemma B.3 (Bounds for the exponent of the exponential bound). *Assuming (A1) and (A3) hold, there exist positive constants $c^* = c(\pi^*)$ and $C^* = C(\pi^*)$ such that*

$$0 < c^* \leq -\frac{1}{r(2n - r - 1)} \mathbb{E}^{Z=z^*} \left[\log \frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} \right] \leq C^*.$$

Proof of Lemma B.3.

$$\begin{aligned} & \mathbb{E}^{Z=z^*} \left[\log \frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} \right] \\ &= \mathbb{E}^{Z=z^*} \left[\sum_{i \neq j} \left\{ X_{ij} \log \left(\frac{\pi_{z_i, z_j}^*}{\pi_{z_i^*, z_j^*}^*} \right) + (1 - X_{ij}) \log \left(\frac{1 - \pi_{z_i, z_j}^*}{1 - \pi_{z_i^*, z_j^*}^*} \right) \right\} + \sum_i \log \frac{\alpha_{z_i}^*}{\alpha_{z_i^*}^*} \right] \\ &= \sum_{i \neq j} \mathbb{E}^{Z=z^*} \left[X_{ij} \log \left(\frac{\pi_{z_i, z_j}^*}{\pi_{z_i^*, z_j^*}^*} \right) + (1 - X_{ij}) \log \left(\frac{1 - \pi_{z_i, z_j}^*}{1 - \pi_{z_i^*, z_j^*}^*} \right) \right] + \sum_i \log \frac{\alpha_{z_i}^*}{\alpha_{z_i^*}^*} \\ &= \sum_{i \neq j} - \left[\pi_{z_i^*, z_j^*}^* \log \left(\frac{\pi_{z_i^*, z_j^*}^*}{\pi_{z_i, z_j}^*} \right) + (1 - \pi_{z_i^*, z_j^*}^*) \log \left(\frac{1 - \pi_{z_i^*, z_j^*}^*}{1 - \pi_{z_i, z_j}^*} \right) \right] + \sum_i \log \frac{\alpha_{z_i}^*}{\alpha_{z_i^*}^*}. \end{aligned}$$

Since $\|z_{[n]} - z_{[n]}^*\|_0 = r$, Lemma B.4 implies that there are $r(2n - r - 1)$ couples (i, j) such that $\pi_{z_i^*, z_j^*}^* \log \left(\frac{\pi_{z_i, z_j}^*}{\pi_{z_i^*, z_j^*}^*} \right) + (1 - \pi_{z_i^*, z_j^*}^*) \log \left(\frac{1 - \pi_{z_i, z_j}^*}{1 - \pi_{z_i^*, z_j^*}^*} \right) \neq 0$. Let us set

$$\begin{aligned} C^* &:= \max \left\{ \pi_{q,l}^* \log \left(\frac{\pi_{q,l}^*}{\pi_{q',l'}^*} \right) + (1 - \pi_{q,l}^*) \log \left(\frac{1 - \pi_{q,l}^*}{1 - \pi_{q',l'}^*} \right) \right\}, \\ c^* &:= \min \left\{ \pi_{q,l}^* \log \left(\frac{\pi_{q,l}^*}{\pi_{q',l'}^*} \right) + (1 - \pi_{q,l}^*) \log \left(\frac{1 - \pi_{q,l}^*}{1 - \pi_{q',l'}^*} \right) \right\}, \end{aligned}$$

where *maximum* and *minimum* are taken over the $\{(q, l), (q', l') \mid \pi_{q,l}^* \neq \pi_{q',l'}^*\}$. Then for each $(i, j) \in D(z_{[n]}^*, z_{[n]})$ (see Lemma B.4),

$$0 < c^* \leq \pi_{z_i^*, z_j^*}^* \log \left(\frac{\pi_{z_i^*, z_j^*}^*}{\pi_{z_i, z_j}^*} \right) + (1 - \pi_{z_i^*, z_j^*}^*) \log \left(\frac{1 - \pi_{z_i^*, z_j^*}^*}{1 - \pi_{z_i, z_j}^*} \right) \leq C^*.$$

Assumption (A4) implies that $\left| \log \frac{\alpha_{z_i}^*}{\alpha_{z_i^*}^*} \right| \leq \log \frac{1-\gamma}{\gamma}$.

Therefore $\frac{1}{r(2n-r-1)} \sum_i \log \frac{\alpha_{z_i}^*}{\alpha_{z_i^*}^*} \leq \frac{r}{r(2n-r-1)} \log \frac{1-\gamma}{\gamma} \xrightarrow{n \rightarrow +\infty} 0$.

Hence for n sufficiently high,

$$0 < c^* \leq -\frac{1}{r(2n-r-1)} \mathbb{E}^{Z=z^*} \left[\log \frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} \right] \leq C^* .$$

□

Lemma B.4. For every $z_{[n]}^*$ and $z_{[n]}$, let us define

$$D^* \left(z_{[n]}^*, z_{[n]} \right) = \left\{ (i, j) \mid i \neq j, \pi_{z_i^*, z_j^*}^* \neq \pi_{z_i, z_j}^* \right\} ,$$

and let $\left| D^* \left(z_{[n]}^*, z_{[n]} \right) \right|$ denote its cardinality. Let us further assume that there are $r > 0$ differences between $z_{[n]}^*$ and $z_{[n]}$, that is $\left\| z_{[n]}^* - z_{[n]} \right\|_0 = r$, and that assumptions (A1) and (A3) hold. Then,

$$\left| D^* \left(z_{[n]}^*, z_{[n]} \right) \right| = r(2n - r - 1) .$$

Proof of Lemma B.4.

Let T_n denote the $n \times n$ matrix defined by $T_n = \{t_{i,j}\}_{1 \leq i \neq j \leq n}$, where $t_{i,j} = \pi_{z_i^*, z_j^*}^* - \pi_{z_i, z_j}^*$, and $t_{i,j} = 0$ when $i = j$. Note that this last requirement results from the fact one is only interested in terms in the sum over couples (i, j) such that $i \neq j$.

Let us further assume that the r differences occur at the first r coordinates of $z_{[n]}$. Then, T_n is organized as a four-blocks matrix, where the bottom-right block (of order $n - r \times n - r$) is 0, as well as the whole diagonal. Then, the number of non-zero coefficients is equal to $n^2 - (n - r)^2 - r$.

Hence, $\left| D^* \left(z_{[n]}^*, z_{[n]} \right) \right| = r(2n - r)n^2 - [(n - r)^2 + r]$, which concludes the proof. □

B.5. (A1) is replaced by (A2) and (A5)

B.5.1. Strategy

The crux in the proof of Theorem 3.1 is the number $r(2n - r - 1)$ of couples (i, j) such that $\pi_{z_i^*, z_j^*}^* \neq \pi_{z_i, z_j}^*$, with $\left\| z_{[n]} - z_{[n]}^* \right\|_0 = r$. However, this is only true under (A1), which is quite restrictive.

In the present section, Proposition B.5 states that assuming (A2) and (A5) hold, there exists a constant $\tau > 0$ such that the number of couples (i, j) in the sum is at least τnr . Then following the same proof as with (A1), assumptions (A2) and (A5) provide the desired convergence rate (up to a constant).

B.5.2. Number of terms under (A2) and (A5)

Due to the relaxation of (A1) into (A2), one has to take into account the case where π^* is permutation-invariant (Section 3.1.1).

Let us start modifying the definition of the number r of differences between the equivalence classes $[z_{[n]}]_{\pi^*}$ and $[z_{[n]}^*]_{\pi^*}$ of vectors $z_{[n]}$ and $z_{[n]}^*$. For any equivalence classes $[z_{[n]}]$ and $[z'_{[n]}]$, let us introduce

$$N_0(z_{[n]}, z'_{[n]}) = N_0\left([z_{[n]}], [z'_{[n]}]\right) = \min_{u \in [z_{[n]}], v \in [z'_{[n]}]} \|u - v\|_0 .$$

In the following, only the notation $N_0(z_{[n]}, z'_{[n]})$ will be used. In the same line of the previous proof with Assumption (A1), the set of equivalence classes can be partitioned according to the number r of differences between $[z_{[n]}]$ and $[z_{[n]}^*]$: $N_0(z_{[n]}, z_{[n]}^*) = r$. Let us further define

$$d(z_{[n]}, z'_{[n]}) = \left| \left\{ (i, j) \mid i \neq j, \pi_{z_i, z_j}^* \neq \pi_{z'_i, z'_j}^* \right\} \right| ,$$

where z_i and z'_j respectively refer to the i -th (resp. j -th) coordinate of vector $z_{[n]}$ (resp. $z'_{[n]}$).

Then, Proposition B.5 leads to

$$d(z_{[n]}, z_{[n]}^*) \geq \frac{\gamma^2}{2} nr ,$$

and the same proof as under Assumption (A1) enables to conclude.

B.5.3. Proposition B.5

Proposition B.5. *If (A2) and (A5) hold, then*

$$d(z_{[n]}, z_{[n]}^*) \geq \frac{\gamma^2}{2} n N_0(z_{[n]}, z_{[n]}^*) ,$$

where $d(z_{[n]}, z'_{[n]}) = \left| \left\{ (i, j) \mid i \neq j, \pi_{z_i, z_j}^* \neq \pi_{z'_i, z'_j}^* \right\} \right|$.

Proof of Proposition B.5.

Let us define $\pi^\sigma = (\pi_{\sigma(q), \sigma(l)})_{1 \leq q, l \leq Q}$ with σ a permutation on $\{1, \dots, Q\}$. Note that for permutation-invariant matrix π , there exists a permutation $\sigma \neq Id$ on $\{1, \dots, Q\}$ such that $\pi^\sigma = \pi$. Then, the following equalities hold

$$d(z_{[n]}, z'_{[n]}) = d\left(\sigma(z_{[n]}), z'_{[n]}\right) = d\left(z_{[n]}, \sigma(z'_{[n]})\right) ,$$

with $\sigma(z_{[n]}) = (\sigma(z_1), \sigma(z_2), \dots, \sigma(z_n))$. Furthermore, neither $d(z_{[n]}, z_{[n]}^*)$ nor $N_0(z_{[n]}, z_{[n]}^*)$ will change if the same permutation is applied to the coordinates of vectors $z_{[n]}$ and $z_{[n]}^*$. Then, computing $d(z_{[n]}, z_{[n]}^*)$ can be made by reordering $z_{[n]}$ and $z_{[n]}^*$.

Assumption (A5) implies that the number of coordinates of $z_{[n]}^*$ that are equal to 1 is at least $n_\gamma := \lceil n\gamma \rceil$, where $\lceil n\gamma \rceil$ denotes the first integer larger than $n\gamma$. The same property holds for every $1 \leq q \leq Q$. Let us use a permutation of the coordinates of $z_{[n]}^*$ such that

$$z_{[n]}^* = (1, 2, \dots, Q, 1, 2, \dots, Q, \dots, 1, 2, \dots, Q, z_{Qn_\gamma+1}^*, z_{Qn_\gamma+2}^*, \dots, z_n^*) ,$$

and apply the same permutation to $z_{[n]}$. For each block k of Q coordinates $(1, \dots, Q)$ of $z_{[n]}^*$, let us introduce a mapping $\sigma_k(\cdot)$ where k denotes the number of the block in $z_{[n]}^*$ such that

$$\forall kQ + 1 \leq i \leq (k+1)Q, \quad \sigma_k(z_i^*) = z_i .$$

Then it comes

$$z_{[n]} = (\sigma_1(1), \sigma_1(2), \dots, \sigma_1(Q), \sigma_2(1), \sigma_2(2), \dots, \sigma_2(Q), \dots, \sigma_{n_\gamma}(1), \dots, \sigma_{n_\gamma}(Q), \tag{11}$$

$$z_{Qn_\gamma+1}, z_{Qn_\gamma+2}, \dots, z_n) . \tag{12}$$

Note that this reorganization of $z_{[n]}^*$ is not unique. For instance, it is possible to exchange $\sigma_1(3)$ with $\sigma_4(3)$. Each σ_k is a function from $\{1, \dots, Q\}$ to $\{1, \dots, Q\}$, which is a permutation provided it is injective. Let us choose a reorganization of the coordinates of z^* which minimizes the number of injective σ_k s.

Besides,

$$\begin{aligned} d(z_{[n]}, z_{[n]}^*) &= \left| \left\{ (i, j) \mid i \neq j, \pi_{z_i^*, z_j^*}^* \neq \pi_{z_i, z_j}^* \right\} \right| \\ &\geq \left| \left\{ (i, j) \mid i \neq j, i, j \leq Qn_\gamma, \pi_{z_i^*, z_j^*}^* \neq \pi_{z_i, z_j}^* \right\} \right| \\ &= \sum_{k, k'} \left| \left\{ (i, j) \mid i \neq j, i \in I_k, j \in I_{k'}, \pi_{z_i^*, z_j^*}^* \neq \pi_{z_i, z_j}^* \right\} \right| , \end{aligned}$$

where I_k denotes the k -th block of coordinates of $z_{[n]}^*$. If $k \neq k'$, the requirement that $i \neq j$ is fulfilled. Otherwise for $k = k'$, it is necessary to require that $z_i^* \neq z_j^*$ since every values in I_k are different. Let us denote by $B(k, k') = \left| \left\{ (q, l) \mid \pi_{q, l}^* \neq \pi_{\sigma_k(q), \sigma_{k'}(l)}^* \right\} \right|$ and by $B(k) =$

$\left| \left\{ (q, l) \mid q \neq l, \pi_{q,l}^* \neq \pi_{\sigma_k(q), \sigma_k(l)}^* \right\} \right|$. Then, it comes that

$$\begin{aligned} d(z_{[n]}, z_{[n]}^*) &\geq \sum_{k, k'} \left| \left\{ (i, j) \mid i \neq j, i \in I_k, j \in I_{k'}, \pi_{z_i^*, z_j^*}^* \neq \pi_{z_i, z_j}^* \right\} \right| \\ &= \sum_{k \neq k'} \left| \left\{ (q, l) \mid \pi_{q,l}^* \neq \pi_{\sigma_k(q), \sigma_{k'}(l)}^* \right\} \right| \\ &\quad + \sum_k \left| \left\{ (q, l) \mid q \neq l, \pi_{q,l}^* \neq \pi_{\sigma_k(q), \sigma_k(l)}^* \right\} \right| \\ &= \sum_{k \neq k'} B(k, k') + \sum_k B(k) . \end{aligned}$$

Therefore, lower bounding $d(z_{[n]}, z_{[n]}^*)$ amounts to assess the cardinality of $B(k, k')$ and $B(k)$, for $1 \leq k \neq k' \leq n_\gamma$.

Let us now distinguish between two cases:

1. either for every $k, k' \in \{1, \dots, n_\gamma\}$, $B(k, k') + B(k', k) > 0$ and $B(k) > 0$.
2. or there exist k, k' such that $B(k, k') + B(k', k) = 0$ or $B(k) = 0$.

First case: In this setting, let $r = N_0(z_{[n]}, z_{[n]}^*)$. Then,

$$\begin{aligned} d(z_{[n]}, z_{[n]}^*) &\geq \sum_{k \neq k'} B(k, k') + \sum_k B(k) \\ &= \sum_{k < k'} [B(k, k') + B(k', k)] + \sum_k B(k) \\ &\geq \frac{n_\gamma(n_\gamma - 1)}{2} + n_\gamma = \frac{n_\gamma(n_\gamma + 1)}{2} \\ &\geq \frac{n_\gamma^2}{2} \geq \frac{n^2 \gamma^2}{2} \geq \frac{\gamma^2}{2} n r , \end{aligned}$$

since $n_\gamma \geq n\gamma$ and $n \geq r$.

Second case: Let us assume that there exists k, k' such that $B(k, k') + B(k', k) = 0$. (The $B(k)$ s will be lower bounded by 0.) Then, Lemma B.6 implies that for every such k, k' , then σ_k and $\sigma_{k'}$ are permutations.

Furthermore, Lemma B.7 then settles that for any such k, k' , $\sigma_k = \sigma_{k'} = \sigma$, where σ denotes a permutation of $\{1, \dots, Q\}$. Note that the existence of such a unique permutation σ results from the last argument in the proof of Lemma B.7. As consequences, it also comes that $\pi^\sigma = \pi$ and that for every $i > Qn_\gamma$, $z_i = \sigma(z_i^*)$.

Let m denote the number of non-injective mappings σ_k . Note that for any non-injective mapping σ_k ($1 \leq k \leq n_\gamma$), there exists at least one difference between $z_{[n]}$ and $z_{[n]}^*$ in the corresponding block k . Then, the number of difference r satisfies

$$r \leq m \times Q \quad \Leftrightarrow \quad \frac{m}{r} \geq \frac{1}{Q} . \quad (13)$$

The conclusion results from

$$\begin{aligned} d(z_{[n]}, z_{[n]}^*) &\geq \sum_{k \neq k'} B(k, k') + \sum_k B(k) \geq \sum_{k \neq k'} B(k, k') \\ &\geq \frac{n_\gamma(n_\gamma - 1) - (n_\gamma - m)[n_\gamma - m - 1]}{2} \\ &= \frac{2mn_\gamma - m^2 - m}{2} = \frac{mn_\gamma + m[n_\gamma - m - 1]}{2}. \end{aligned}$$

Finally, let us notice that $m \leq n_\gamma$, and that $n_\gamma - 1 \geq m = n_\gamma$ amounts to say that no injective mapping σ_k exists in (11). However according to Lemma B.6, it means that for every $1 \leq k, k' \leq n_\gamma$, $B(k, k') + B(k', k) > 0$, which contradicts the assumption. Then, $n_\gamma - (m + 1) \geq 0$ and

$$d(z_{[n]}, z_{[n]}^*) \geq \frac{mn_\gamma}{2} \geq \frac{mn_\gamma}{2} = \frac{\gamma n r m}{2r} \geq \frac{\gamma n r}{2Q} \geq \frac{\gamma^2 n r}{2}.$$

by use of (13) and $\gamma \leq 1/Q$ (see Assumption (A5)).

□

Lemma B.6. *With the same notation as Proposition B.5 and its proof, let us assume (A2) and (A5) hold. Furthermore, if $B(k, k') = B(k', k) = 0$ for given $1 \leq k, k' \leq n_\gamma$, then σ_k and $\sigma_{k'}$ are permutations of $\{1, \dots, Q\}$.*

Proof of Lemma B.6.

If $B(k, k') = B(k', k) = 0$ for given k and k' . Then for every $q, l \in \{1, \dots, Q\}$, $\pi_{q,l} = \pi_{\sigma_k(q), \sigma_{k'}(l)} = \pi_{\sigma_{k'}(q), \sigma_k(l)}$.

Assume furthermore that $\sigma_k(q) = \sigma_{k'}(q')$ for some $q, q' \in \{1, \dots, Q\}$. Then for every $l \in \{1, \dots, Q\}$, $\pi_{q,l} = \pi_{\sigma_k(q), \sigma_{k'}(l)} = \pi_{\sigma_k(q'), \sigma_{k'}(l)} = \pi_{q',l}$. Hence, for every $l \in \{1, \dots, Q\}$, $\pi_{q,l} = \pi_{q',l}$, which implies $q = q'$ using (A2).

Therefore, σ_k is injective and thus a permutation of $\{1, \dots, Q\}$. The same property holds for $\sigma_{k'}$ which is also a permutation of $\{1, \dots, Q\}$.

□

Lemma B.7. *With the same notation and assumptions as Proposition B.5 and its proof, let us assume that σ_k and $\sigma_{k'}$ are two permutations of $\{1, \dots, Q\}$ defined in (11). Then, $\sigma_k = \sigma_{k'}$ and $\pi^{\sigma_k} = \pi$.*

Moreover, for every $i > Qn_\gamma$, $\sigma_k(z_i^) = z_i$.*

Proof of Lemma B.7.

Let us assume that $\sigma_k \neq \sigma_{k'}$. Then, there exists $q \in \{1, \dots, Q\}$ such that $\sigma_k(q) \neq \sigma_{k'}(q)$. If it holds, one can interchange coordinates of $z_{[n]}$: $\sigma_k(q)$ and $\sigma_{k'}(q)$. This results in new mappings σ_k and $\sigma_{k'}$ between $z_{[n]}^*$ and $z_{[n]}$, which are

no longer injective. Then, the number of injective mappings σ_k in the writing of π decreases by 2 and is no longer minimal as earlier assumed. This yields $\sigma_k = \sigma_{k'}$ and thus $\pi^{\sigma_k} = \pi$.

Besides, for every $i > Qn_\gamma$, $z_i = \sigma_k(z_i^*)$. Indeed if this was not true, the same reasoning as before applies: An interchange between z_i and $\sigma_k(z_i^*)$ would decrease the number of injective σ_k s in (11), which contradicts our assumption. \square

Appendix C: Proof of Proposition 3.5

Proof of Proposition 3.5. Let us first recall that

$$\begin{aligned}\phi_n(z_{[n]}, \pi) &:= \frac{1}{n(n-1)} \mathcal{L}_1(X_{[n]}; z_{[n]}, \pi) \ , \\ \Phi_n(z_{[n]}, \pi) &:= \mathbb{E} \left[M_n(z_{[n]}, \pi) \mid z_{[n]} = z_{[n]}^* \right] \ .\end{aligned}$$

Then,

$$\begin{aligned}|\phi_n(z_{[n]}, \pi) - \Phi_n(z_{[n]}, \pi)| &= \rho_n \left| \sum_{i \neq j} \left(X_{ij} - \pi_{z_i^*, z_j^*} \right) \log \left[\pi_{z_i, z_j} / (1 - \pi_{z_i, z_j}) \right] \right| \ , \\ &= \rho_n \left| \sum_{i \neq j} \xi_{ij} g(\pi_{z_i, z_j}) \right| \ ,\end{aligned}$$

where $\rho_n = n(n-1)^{-1}$, $\xi_{ij} = X_{ij} - \pi_{z_i^*, z_j^*}$, and $g(t) = \log(t/(1-t))$, $t \in]0, 1[$.

With $g_{i,j} = g(\pi_{z_i, z_j})$, let us introduce

$$Z(g) = \left| \sum_{i \neq j} \xi_{ij} g_{ij} \right| \ ,$$

where $g = \{g_{i,j}\}_{1 \leq i \neq j \leq n}$.

Let \mathcal{P}_1 denote the set of couples $(z_{[n]}, \pi)$ such that the number of couples (i, j) such that $g_{i,j} \neq 0$ is negligible with respect to $n(n-1)$, and $\mathcal{P}_2 = \mathcal{P} \setminus \mathcal{P}_1$. Then, $\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2$ and for any $\eta > 0$,

$$\begin{aligned}\mathbb{P} \left[\sup_{\mathcal{P}} \rho_n Z(g) > \eta \mid Z = z^* \right] &= P^* \left[\sup_{\mathcal{P}} \rho_n Z(g) > \eta \right] \\ &\leq P^* \left[\sup_{\mathcal{P}_1} \rho_n Z(g) > \eta \right] + P^* \left[\sup_{\mathcal{P}_2} \rho_n Z(g) > \eta \right] \ .\end{aligned}\tag{14}$$

On the one hand, the first probability in the right-hand side of (14) converges to 0 $\mathbb{P} - a.s.$, since Lemma C.2 implies that

$$\sup_{\mathcal{P}_1} \rho_n Z(g) \xrightarrow[n \rightarrow \infty]{} 0, \quad \mathbb{P} - a.s. \ .$$

On the other hand, the second probability in the right-hand side of (14) can be dealt with thanks to Talagrand's inequality. For every $z_{[n]}$ and $\epsilon > 0$, let us define

$$\Omega_n(\epsilon; z_{[n]}) = \left\{ \sup_{\mathcal{P}_2(z_{[n]})} \rho_n Z(g) \leq (1 + \epsilon) \sqrt{\rho_n} \Lambda + \sqrt{\rho_n \Gamma^2 x_n} + (1/\epsilon + 1/3) \rho_n \Gamma x_n \right\} \ ,$$

where Γ and Λ are constants respectively defined in Lemma C.2 and Lemma C.3, $\mathcal{P}_2(z_{[n]}) = \{\pi \mid (z_{[n]}, \pi) \in \mathcal{P}_2\}$, and $\{x_n\}_n$ is a sequence of positive real numbers to be chosen later. Talagrand's inequality (Theorem C.1) implies for any $z_{[n]}$

$$P^* \left[\Omega_n(\epsilon; z_{[n]})^c \right] \leq e^{-x_n} .$$

Combining the previous result with

$$P^* \left[\sup_{\mathcal{P}_2} |\phi_n(z_{[n]}, \pi) - \Phi_n(z_{[n]}, \pi)| > \eta \right] \leq \sum_{z_{[n]}} P^* \left[\sup_{\mathcal{P}_2(z_{[n]})} \rho_n Z(g) > \eta \right] ,$$

it comes

$$\begin{aligned} & P^* \left[\sup_{\mathcal{P}_2} |\phi_n(z_{[n]}, \pi) - \Phi_n(z_{[n]}, \pi)| > \eta \right] \\ & \leq \sum_{z_{[n]}} P^* \left[\left\{ \sup_{\mathcal{P}_2(z_{[n]})} \rho_n Z(g) > \eta \right\} \cap \Omega_n(\epsilon; z_{[n]}) \right] + \sum_{z_{[n]}} e^{-x_n} \\ & \leq \sum_{z_{[n]}} P^* \left[(1 + \epsilon) \sqrt{\rho_n} \Lambda + \sqrt{\rho_n \Gamma^2 x_n} + (1/\epsilon + 1/3) \rho_n \Gamma x_n > \eta \right] + \sum_{z_{[n]}} e^{-x_n} . \end{aligned}$$

Since $z_{[n]}$ belongs to a set of cardinality at most Q^n , choosing $x_n = n \log(n)$ entails the first sum is equal to 0 for large enough values of n , while the second sum converges to 0.

Finally, a quick inspection of the proof shows this convergence is uniform with respect to $z_{[n]}^*$, which provides the desired result. \square

Theorem C.1 (Talagrand). *Let $\{Y_{ij}\}_{1 \leq i \neq j \leq n}$ denote independent centered random variables, and define*

$$\forall g \in \mathcal{G}, \quad Z(g) = \sum_{i \neq j} Y_{ij} g_{ij} ,$$

where $\mathcal{G} \subset \mathbb{R}^{n^2}$. Let us further assume that there exist $b > 0$ and $\sigma^2 > 0$ such that $|Y_{ij} g_{ij}| \leq b$ for every (i, j) , and $\sup_{g \in \mathcal{G}} \sum_{i \neq j} \text{Var}(Y_{ij} g_{ij}) \leq \sigma^2$. Then, for every $\epsilon > 0$, and $x > 0$,

$$\mathbb{P} \left[\sup_g Z(g) \geq \mathbb{E} \left[\sup_g Z(g) \right] (1 + \epsilon) + \sqrt{2\sigma^2 x} + b(1/\epsilon + 1/3)x \right] \leq e^{-x} .$$

Proof. A proof can be found in Massart (2007) (p.95, (4.50)). \square

Lemma C.2. *With the same notation as Theorem C.1, Assumption (A3) entails that there exists $\Gamma(\zeta) > 0$ only depending on ζ such that*

$$\sup_{\mathcal{P}} \max_{i \neq j} |\xi_{ij} g_{ij}| \leq \Gamma, \quad \text{and} \quad \sup_{\mathcal{P}} \max_{i \neq j} \text{Var}(\xi_{ij} g_{ij}) \leq \frac{\Gamma^2}{4} .$$

Proof. If $(z_{[n]}, \pi) \in \mathcal{P}$, then

$$\left(\pi_{z_i, z_j} \in \{0, 1\} \Rightarrow \pi_{z_i^*, z_j^*}^* = \pi_{z_i, z_j} \right) \Rightarrow (g_{i,j} = 0) .$$

Then for every $(z_{[n]}, \pi) \in \mathcal{P}$, there exists $\Gamma = \Gamma(\zeta) > 0$ (Assumption (A3)) such that

$$\forall i \neq j, \quad |\xi_{ij} g_{ij}| \leq \Gamma ,$$

for every $(z_{[n]}, \pi) \in \mathcal{P}$. This also leads to

$$\forall i \neq j, \quad \text{Var}(\xi_{ij} g_{ij}) \leq \Gamma^2/4 .$$

□

Lemma C.3. *With the same notation as Proposition 3.5, for every $z_{[n]}$ such that $(z_{[n]}, \pi) \in \mathcal{P}_2$, there exists a constant $\Lambda = \Lambda(\zeta) > 0$ (Assumption (A3)) such that*

$$\mathbb{E} \left[\sup_{\mathcal{P}_2(z_{[n]})} \rho_n \left| \sum_{i \neq j} (X_{ij} - \pi_{z_i^*, z_j^*}^*) g_{ij} \right| \middle| Z = z^* \right] \leq \Lambda \sqrt{n(n-1)}^{-1} .$$

Proof of Lemma C.3. Let $\mathbb{E}^*[\cdot]$ denote the expectation given $Z = z^*$. Then,

$$\begin{aligned} & \mathbb{E}^* \sup_{\mathcal{P}_2(z_{[n]})} \rho_n \left| \sum_{i \neq j} (X_{ij} - \pi_{z_i^*, z_j^*}^*) g_{ij} \right| \\ & \leq \mathbb{E}_{X, X'}^* \left[\sup_{\mathcal{P}_2(z_{[n]})} \rho_n \left| \sum_{i \neq j} (X_{ij} - X'_{ij}) g_{ij} \right| \right] , \end{aligned}$$

where the $X'_{i,j}$ s are independent random variables with the same distribution as the $X_{i,j}$ s. A symetrization argument based on Rademacher variables

$\{\epsilon_{i,j}\}_{1 \leq i \neq j \leq n}$ leads to

$$\begin{aligned} & \mathbb{E}^* \sup_{\mathcal{P}_2(z_{[n]})} \rho_n \left| \sum_{i \neq j} \left(X_{ij} - \pi_{z_i^*, z_j^*}^* \right) g_{ij} \right| \\ & \leq 2 \mathbb{E}^* \left[\sup_{\mathcal{P}_2(z_{[n]})} \rho_n \mathbb{E}_\epsilon \left[\left| \sum_{i \neq j} \epsilon_{ij} X_{ij} g_{ij} \right| \right] \right] , \end{aligned}$$

where $\mathbb{E}_\epsilon[\cdot]$ denotes the expectation with respect to $\epsilon_{i,j}$ s. Then, Jensen's inequality yields

$$\begin{aligned} & \mathbb{E}^* \sup_{\mathcal{P}_2(z_{[n]})} \rho_n \left| \sum_{i \neq j} \left(X_{ij} - \pi_{z_i^*, z_j^*}^* \right) g_{ij} \right| \\ & \leq 2 \mathbb{E}^* \left[\sup_{\mathcal{P}_2(z_{[n]})} \rho_n \sqrt{\text{Var}_\epsilon \left[\sum_{i \neq j} \epsilon_{ij} X_{ij} g_{ij} \right]} \right] \\ & \leq 2 \mathbb{E}^* \left[\sup_{\mathcal{P}_2(z_{[n]})} \rho_n \sqrt{n(n-1)g_{ij}^2} \right] \\ & \leq \Lambda(\zeta) \sqrt{\rho_n} . \end{aligned}$$

□

Justification of Talagrand in the proof of Proposition 3.5. Let us first recall that

$$\begin{aligned} \phi_n(z_{[n]}, \pi) & := \frac{1}{n(n-1)} \mathcal{L}_1(X_{[n]}; z_{[n]}, \pi) , \\ \Phi_n(z_{[n]}, \pi) & := \mathbb{E} \left[M_n(z_{[n]}, \pi) \mid z_{[n]} = z_{[n]}^* \right] . \end{aligned}$$

Let Θ be the set of parameters $(z_{[n]}, \pi)$ defined by assumptions of Section 2.2. In particular, any $z_{[n]}$ is allowed whereas only values of π satisfying (A3) are allowed.

$$\begin{aligned} \phi_n(z_{[n]}, \pi) - \Phi_n(z_{[n]}, \pi) & = \rho_n \sum_{i \neq j} \left(X_{ij} - \pi_{z_i^*, z_j^*}^* \right) \log \left[\pi_{z_i, z_j} / (1 - \pi_{z_i, z_j}) \right] \\ & = \rho_n \sum_{q,l} \sum_{i \neq j} \left(X_{ij} - \pi_{q,l}^* \right) \mathbb{1}_{(z_i^*=q, z_j^*=l)} a_{i,j} , \end{aligned}$$

where $a_{i,j} = \log \left[\pi_{z_i, z_j} / (1 - \pi_{z_i, z_j}) \right]$. For every (q, l) , Hoeffding's inequality and (A3) yields

$$P^* \left[\left| \sum_{i \neq j} \left(X_{ij} - \pi_{q,l}^* \right) \mathbb{1}_{(z_i^*=q, z_j^*=l)} a_{i,j} \right| \geq t \right] \leq 2 \exp \left[- \frac{t^2}{4N_{q,l}(z^*)\Gamma^2} \right] ,$$

with $N_{q,l}(z_{[n]}^*) = \sum_{i \neq j} \mathbb{1}_{(z_i^*=q, z_j^*=l)}$. Then,

$$\begin{aligned} & P^* [|\phi_n(z_{[n]}, \pi) - \Phi_n(z_{[n]}, \pi)| \geq t\rho_n] \\ & \leq \sum_{q,l} P^* \left[\left| \sum_{i \neq j} (X_{ij} - \pi_{q,l}^*) \mathbb{1}_{(z_i^*=q, z_j^*=l)} a_{i,j} \right| \geq t/Q^2 \right] \\ & \leq 2 \sum_{q,l} \exp \left[-\frac{t^2}{4N_{q,l}(z_{[n]}^*)Q^4\Gamma^2} \right]. \end{aligned}$$

Thus for every $x > 0$,

$$P^* [|\phi_n(z_{[n]}, \pi) - \Phi_n(z_{[n]}, \pi)| \geq x] \leq 2 \sum_{q,l} \exp \left[-\frac{x^2}{4\rho_n^2 N_{q,l}(z_{[n]}^*)Q^4\Gamma^2} \right].$$

Furthermore,

$$\begin{aligned} & \mathbb{P} [|\phi_n(z_{[n]}, \pi) - \Phi_n(z_{[n]}, \pi)| \geq x] \\ & = \sum_{z_{[n]}} \mathbb{P} [|\phi_n(z_{[n]}, \pi) - \Phi_n(z_{[n]}, \pi)| \geq x \mid Z_{[n]} = z_{[n]}] \mathbb{P} [Z_{[n]} = z_{[n]}] \\ & \leq \sum_{z_{[n]}} 2 \sum_{q,l} \exp \left[-\frac{x^2}{4\rho_n^2 N_{q,l}(z_{[n]})Q^4\Gamma^2} \right] \mathbb{P} [Z_{[n]} = z_{[n]}] \\ & = 2 \sum_{q,l} \mathbb{E} \left[\exp \left[-\frac{x^2}{4\rho_n^2 N_{q,l}(Z_{[n]})Q^4\Gamma^2} \right] \right]. \end{aligned} \tag{15}$$

Then, random variables $N_{q,l}(Z_{[n]})$ follow a binomial distribution $\mathcal{B}(n, \pi_{q,l}^*)$. It entails that

$$\mathbb{P} [|N_{q,l}(Z_{[n]}) - \alpha_q^* \alpha_l^* \rho_n^{-1}| \geq t] \leq 2 \exp \left[-\frac{t^2}{\rho_n^{-1}} \right],$$

hence for every $u > 0$

$$\mathbb{P} [|\rho_n N_{q,l}(Z_{[n]}) - \alpha_q^* \alpha_l^*| \geq u] \leq 2 \exp \left[-\frac{u^2}{\rho_n} \right].$$

Plugging this in (15), it comes for every $u > 0$

$$\begin{aligned} & \mathbb{E} \left[\exp \left[-\frac{x^2}{4\rho_n^2 N_{q,l}(Z_{[n]})Q^4\Gamma^2} \right] \right] \\ & \leq 2 \exp \left[-\frac{u^2}{\rho_n} \right] + \mathbb{E} \left[\exp \left[-\frac{x^2}{4\rho_n (u + \alpha_q^* \alpha_l^*) Q^4\Gamma^2} \right] \right] \\ & \leq 2 \exp \left[-\frac{u^2}{\rho_n} \right] + \exp \left[-\frac{x^2}{4\rho_n (u + 1) Q^4\Gamma^2} \right]. \end{aligned}$$

It enables to conclude that for every $x > 0$,

$$\begin{aligned} & \mathbb{P} \left[\left| \phi_n(z_{[n]}, \pi) - \Phi_n(z_{[n]}, \pi) \right| \geq x \right] \\ & \leq 2Q^2 \exp \left[-\frac{x^2}{4\rho_n(u+1)Q^4\Gamma^2} \right] + 4Q^2 \exp \left[-\frac{u^2}{\rho_n} \right] . \end{aligned}$$

□

Appendix D: Theorem 3.6

Notation Let α^* and π^* be the true values of α and π , \mathcal{A} be the set of stochastic matrices of size Q given by $\mathcal{A} = \{A = (a_{k,l})_{1 \leq k,l \leq Q} \mid a_{k,l} \geq 0, \sum_{l=1}^Q a_{k,l} = 1\}$.

Furthermore, let us introduce the following quantities

$$\begin{aligned} \phi_n(\pi, z_{[n]}) &= \frac{1}{n(n-1)} \mathcal{L}_1(X_{[n]}; z_{[n]}, \pi), \quad \widehat{z}_{[n]}(\pi) = \operatorname{Argmax}_z \phi_n(z_{[n]}, \pi), \\ \Phi_n(\pi, z_{[n]}) &= \frac{1}{n(n-1)} \sum_{i,j \neq i} \pi_{z_i^* z_j^*}^* \log \pi_{z_i, z_j} + (1 - \pi_{z_i^* z_j^*}^*) \log(1 - \pi_{z_i, z_j}), \\ \widetilde{z}_{[n]}(\pi) &= \operatorname{Argmax}_z \Phi_n(z_{[n]}, \pi), \\ M_n(\alpha, \pi) &= \frac{1}{n(n-1)} \mathcal{L}_2(X_{[n]}; \alpha, \pi), \\ \mathbb{M}(\pi, A) &= \sum_{q,l} \alpha_q^* \alpha_l^* \sum_{q'l'} a_{q,q'} a_{l,l'} [\pi_{q,l}^* \log \pi_{q'l'} + (1 - \pi_{q,l}^*) \log(1 - \pi_{q'l'})], \\ \bar{A}_\pi &= \operatorname{Argmax}_{A \in \mathcal{A}} \mathbb{M}(\pi, A), \quad \mathbb{M}(\pi) = \mathbb{M}(\pi, \bar{A}_\pi). \end{aligned}$$

Note that $\bar{A}_{\pi^*} = I_Q$ and $\mathbb{M}(\pi^*) = \sum_{q,l} \alpha_q^* \alpha_l^* \mathbb{H}_{q,l}^*$, where $\mathbb{H}_{q,l}^* = \pi_{q,l}^* \log \pi_{q,l}^* + (1 - \pi_{q,l}^*) \log(1 - \pi_{q,l}^*)$.

Proof

First, let us prove that: $\forall \eta > 0, \sup_{d(\pi, \pi^*) \geq \eta} \mathbb{M}(\pi) < \mathbb{M}(\pi^*)$. In the sequel, let $(\bar{a}_{q,l})_{1 \leq q,l \leq Q}$ denote coefficient of \bar{A}_π . Without any further indication, $\bar{a}_{q,l}$ refers to the matrix π . One has

$$\begin{aligned} &\mathbb{M}(\pi) - \mathbb{M}(\pi^*) \\ &= \sum_{q,l} \alpha_q^* \alpha_l^* \sum_{q'l'} \bar{a}_{q,q'} \bar{a}_{l,l'} [\pi_{q,l}^* \log \frac{\pi_{q'l'}}{\pi_{q,l}^*} + (1 - \pi_{q,l}^*) \log \frac{1 - \pi_{q'l'}}{1 - \pi_{q,l}^*}] \\ &= - \sum_{q,l} \alpha_q^* \alpha_l^* \sum_{q'l'} \bar{a}_{q,q'} \bar{a}_{l,l'} K(\pi_{q,l}^*, \pi_{q'l'}) . \end{aligned}$$

Since $\{\pi \mid d(\pi, \pi^*) \geq \eta, (A2)\}$ is a compact set, there exists $\pi^0 \neq \pi^*$ satisfying (A2) such that

$$\sup_{d(\pi, \pi^*) \geq \eta} \mathbb{M}(\pi) - \mathbb{M}(\pi^*) = \mathbb{M}(\pi^0) - \mathbb{M}(\pi^*) < 0 .$$

Otherwise for every (q, l) , $\sum_{q'l'} \bar{a}_{q,q'} \bar{a}_{l,l'} K(\pi_{q,l}^*, \pi_{q'l'}^0) = 0$ would entail that for every (q', l') , $\bar{a}_{q,q'} \bar{a}_{l,l'} K(\pi_{q,l}^*, \pi_{q'l'}^0) = 0$. It implies that there exists $f : \{1, \dots, Q\} \rightarrow \{1, \dots, Q\}$, $\pi_{q,l}^* = \pi_{f(q), f(l)}^0$, which is excluded since $\pi^0 \neq \pi^*$ up to label switching.

Second, let us prove that $\|M_n - \mathbb{M}\|_{\Theta_n} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0$. Set

$$|M_n(\alpha, \pi) - \mathbb{M}(\pi)| \leq |M_n(\alpha, \pi) - \phi_n(\pi, \widehat{z}_{[n]})| \quad (16)$$

$$+ |\phi_n(\pi, \widehat{z}_{[n]}) - \Phi_n(\pi, \widetilde{z}_{[n]})| \quad (17)$$

$$+ |\Phi_n(\pi, \widetilde{z}_{[n]}) - \mathbb{M}(\pi)| . \quad (18)$$

The three terms of the preceding expression are controlled as follows:

- upper bound of (16): Lemma F.2 implies that $\mathbb{P} - a.s.$,

$$\begin{aligned} \sup_{\alpha, \pi} |M_n(\alpha, \pi) - \phi_n(\pi, \widehat{z}_{[n]})| &= \sup_{\alpha, \pi} \frac{|\mathcal{L}_2(X_{[n]}; \alpha, \pi) - \mathcal{L}_1(X_{[n]}; \pi, \widehat{z}_{[n]})|}{n(n-1)} \\ &\leq \frac{\log(1/\gamma)}{n-1} \xrightarrow[n \rightarrow \infty]{} 0 , \end{aligned}$$

- upper bound of (17): the definitions of $\widehat{z}_{[n]}$ and $\widetilde{z}_{[n]}$ imply that $\phi_n(\pi, \widehat{z}_{[n]}) \geq \phi_n(\pi, \widetilde{z}_{[n]})$ and $\Phi_n(\pi, \widehat{z}_{[n]}) \leq \Phi_n(\pi, \widetilde{z}_{[n]})$, hence

$$|\phi_n(\pi, \widehat{z}_{[n]}) - \Phi_n(\pi, \widetilde{z}_{[n]})| \leq \sup_{\pi, z_{[n]}} |\phi_n(\pi, z_{[n]}) - \Phi_n(\pi, z_{[n]})| .$$

Theorem 3.5 implies that for every $z_{[n]}^*$ and $\eta > 0$,

$$\mathbb{P} \left(\sup_{\pi, z_{[n]}} |\phi_n(\pi, z_{[n]}) - \Phi_n(\pi, z_{[n]})| > \eta \mid Z_{[n]} = z_{[n]}^* \right) \xrightarrow[n \rightarrow +\infty]{} 0 ,$$

where the rate of convergence does not depend on $z_{[n]}^*$. Therefore

$$\mathbb{P} \left(\sup_{\pi} |\phi_n(\pi, \widehat{z}_{[n]}) - \Phi_n(\pi, \widetilde{z}_{[n]})| > \eta \right) \xrightarrow[n \rightarrow +\infty]{} 0 .$$

- upper bound of (18): $\Phi_n(\pi, z_{[n]})$ can be expressed as:

$$\begin{aligned} &\Phi_n(\pi, z_{[n]}) \\ &= \sum_{qq'l'} \frac{N_{qq'}(z_{[n]})N_{l'l}(z_{[n]})}{n(n-1)} [\pi_{q,l}^* \log \pi_{q'l'} + (1 - \pi_{q,l}^*) \log(1 - \pi_{q'l'})] , \quad (19) \end{aligned}$$

where $N_{qq'}(z_{[n]}) = \#\{i \mid z_i^* = q, \text{ and } z_i = q'\}$.

Let $\widetilde{N}_{qq'}(\pi) = N_{qq'}(\widetilde{z}_{[n]}(\pi))$, $N_q^* = \#\{i \mid z_i^* = q\}$, $\widetilde{a}_{qq'}(\pi) = \frac{\widetilde{N}_{qq'}(\pi)}{N_q^*}$, and \widetilde{A}_π the stochastic matrix of $\widetilde{a}_{qq'}(\pi)$. Coefficient $\widetilde{a}_{qq'}(\pi)$ yield the proportion of vertices from class q attributed to class q' by $z_{[n]}$. Note that (19) shows that $\Phi_n(\pi, z_{[n]})$ only depends on $z_{[n]}$ through the matrix \widetilde{A}_π . Therefore, one uses the notation $\Phi_n(\pi, \widetilde{A}_\pi)$ in place of $\Phi_n(\pi, z_{[n]})$.

Definitions of \widetilde{A}_π and \bar{A}_π imply that $\Phi_n(\pi, \widetilde{A}_\pi) \geq \Phi_n(\pi, \bar{A}_\pi)$ and $\mathbb{M}(\pi) = \mathbb{M}(\pi, \bar{A}_\pi) \geq \mathbb{M}(\pi, \widetilde{A}_\pi)$. Therefore,

1. $\Phi_n(\pi, \tilde{A}_\pi) \geq \mathbb{M}(\pi)$
 $\Rightarrow 0 \leq \Phi_n(\pi, \tilde{A}_\pi) - \mathbb{M}(\pi) \leq \Phi_n(\pi, \tilde{A}_\pi) - \mathbb{M}(\pi, \tilde{A}_\pi),$
2. $\Phi_n(\pi, \tilde{A}_\pi) \leq \mathbb{M}(\pi)$
 $\Rightarrow 0 \leq \mathbb{M}(\pi) - \Phi_n(\pi, \tilde{A}_\pi) \leq \mathbb{M}(\pi, \bar{A}_\pi) - \Phi_n(\pi, \bar{A}_\pi).$

Then,

$$\left| \Phi_n(\pi, \tilde{A}_\pi) - \mathbb{M}(\pi) \right| \leq \max_{A \in \mathcal{A}} |\Phi_n(\pi, A) - \mathbb{M}(\pi, A)| .$$

Moreover for every $A \in \mathcal{A}$,

$$\begin{aligned} \Phi_n(\pi, A) - \mathbb{M}(\pi, A) = \\ \sum_{qq'l'l'} \left[\frac{N_q^* N_l^*}{n(n-1)} - \alpha_q^* \alpha_l^* \right] a_{qq'} a_{ll'} \left[\pi_{q,l}^* \log \pi_{q'l'} + (1 - \pi_{q,l}^*) \log(1 - \pi_{q'l'}) \right] . \end{aligned}$$

Since any $\pi_{q'l'} \in \{0, 1\}$ such that $\pi_{q,l}^* \neq \pi_{q'l'}$ is excluded, (A3) provides

$$\left| \pi_{q,l}^* \log \pi_{q'l'} + (1 - \pi_{q,l}^*) \log(1 - \pi_{q'l'}) \right| \leq \Delta(\zeta) < +\infty ,$$

where $\Delta(\zeta) > 0$ is independent of π and q , and only depends on ζ from Assumption (A3).

Then, the strong law of large numbers applied to N_q^* entails that $\sup_{\pi} \left\{ \left| \Phi_n(\pi, \tilde{z}_{[n]}(\pi)) - \mathbb{M}(\pi) \right| \right\} \xrightarrow[n \rightarrow +\infty]{} 0 \quad \mathbb{P} - a.s. .$

Appendix E: Proof of Proposition 3.8

Proof of Proposition 3.8.

$$\begin{aligned} & \log \frac{\widehat{P}^{X_{[n]}}(z_{[n]})}{\widehat{P}^{X_{[n]}}(z_{[n]}^*)} \\ &= \sum_{i \neq j} \left\{ X_{ij} \log \left(\frac{\widehat{\pi}_{z_i, z_j}}{\widehat{\pi}_{z_i^*, z_j^*}} \right) + (1 - X_{ij}) \log \left(\frac{1 - \widehat{\pi}_{z_i, z_j}}{1 - \widehat{\pi}_{z_i^*, z_j^*}} \right) \right\} + \sum_i \log \frac{\widehat{\alpha}_{z_i}}{\widehat{\alpha}_{z_i^*}}. \end{aligned}$$

Notice that

$$\log \widehat{\pi}_{z_i, z_j} = \log \pi_{z_i, z_j}^* + \log \left[1 + \frac{\widehat{\pi}_{z_i, z_j} - \pi_{z_i, z_j}^*}{\pi_{z_i, z_j}^*} \right]$$

and

$$\log(1 - \widehat{\pi}_{z_i, z_j}) = \log(1 - \pi_{z_i, z_j}^*) + \log \left[1 - \frac{\widehat{\pi}_{z_i, z_j} - \pi_{z_i, z_j}^*}{1 - \pi_{z_i, z_j}^*} \right].$$

Therefore,

$$\begin{aligned} & X_{ij} \log \widehat{\pi}_{z_i, z_j} + (1 - X_{ij}) \log(1 - \widehat{\pi}_{z_i, z_j}) \\ &= X_{ij} \log \pi_{z_i, z_j}^* + (1 - X_{ij}) \log(1 - \pi_{z_i, z_j}^*) \\ &+ \log \left[1 + \frac{(\widehat{\pi}_{z_i, z_j} - \pi_{z_i, z_j}^*)(X_{ij} - \pi_{z_i, z_j}^*)}{\pi_{z_i, z_j}^*(1 - \pi_{z_i, z_j}^*)} \right], \end{aligned}$$

and thus

$$\begin{aligned} & \log \frac{\widehat{P}^{X_{[n]}}(z_{[n]})}{\widehat{P}^{X_{[n]}}(z_{[n]}^*)} \\ &= \sum_{i \neq j} \left\{ X_{ij} \log \left(\frac{\pi_{z_i, z_j}^*}{\pi_{z_i^*, z_j^*}^*} \right) + (1 - X_{ij}) \log \left(\frac{1 - \pi_{z_i, z_j}^*}{1 - \pi_{z_i^*, z_j^*}^*} \right) \right\} + \sum_i \log \frac{\widehat{\alpha}_{z_i}}{\widehat{\alpha}_{z_i^*}} \\ &+ \sum_{i \neq j} \log \left[1 + \frac{(\widehat{\pi}_{z_i, z_j} - \pi_{z_i, z_j}^*)(X_{ij} - \pi_{z_i, z_j}^*)}{\pi_{z_i, z_j}^*(1 - \pi_{z_i, z_j}^*)} \right] \\ &- \sum_{i \neq j} \log \left[1 + \frac{(\widehat{\pi}_{z_i^*, z_j^*} - \pi_{z_i^*, z_j^*}^*)(X_{ij} - \pi_{z_i^*, z_j^*}^*)}{\pi_{z_i^*, z_j^*}^*(1 - \pi_{z_i^*, z_j^*}^*)} \right] \\ &= T_1 + T_2 - T_3. \end{aligned}$$

In the following, one successively upper bound T_1 , T_2 , and T_3 .

Upper bounding T_1

The magnitude of T_1 is given by a similar argument to that in the proof of Theorem 3.1. Let us consider

$$\begin{aligned}
 T_1 &= \sum_{i \neq j} \left\{ X_{i,j} \log \left(\frac{\pi_{z_i, z_j}^*}{\pi_{z_i^*, z_j^*}^*} \right) + (1 - X_{i,j}) \log \left(\frac{1 - \pi_{z_i, z_j}^*}{1 - \pi_{z_i^*, z_j^*}^*} \right) \right\} + \sum_i \log \frac{\widehat{\alpha}_{z_i}}{\widehat{\alpha}_{z_i^*}} \\
 &= \sum_{i \neq j} \left\{ (X_{i,j} - \pi_{z_i^*, z_j^*}^*) \log \left(\frac{\pi_{z_i, z_j}^*}{\pi_{z_i^*, z_j^*}^*} \frac{1 - \pi_{z_i^*, z_j^*}^*}{1 - \pi_{z_i, z_j}^*} \right) + \sum_i \log \frac{\widehat{\alpha}_{z_i}}{\widehat{\alpha}_{z_i^*}} \right\} \\
 &\quad + \sum_{i \neq j} \left\{ \pi_{z_i^*, z_j^*}^* \log \left(\frac{\pi_{z_i, z_j}^*}{\pi_{z_i^*, z_j^*}^*} \right) + (1 - \pi_{z_i^*, z_j^*}^*) \log \left(\frac{1 - \pi_{z_i, z_j}^*}{1 - \pi_{z_i^*, z_j^*}^*} \right) \right\} \\
 &= T_{1,1} + T_{1,2} .
 \end{aligned}$$

Then,

$$P^* [T_1 > t] = P^* [T_{1,1} + T_{1,2} > t] .$$

Since (with the assumption that every coefficients of π^* are different)

$$\begin{aligned}
 &T_{1,2} [r(2n - r - 1)]^{-1} \\
 &\leq \max_{(q,l) \neq (q',l'), \pi_{q,l}^* \notin \{0,1\}} \pi_{q,l}^* \log \left(\frac{\pi_{q',l'}^*}{\pi_{q,l}^*} \right) + (1 - \pi_{q,l}^*) \log \left(\frac{1 - \pi_{q',l'}^*}{1 - \pi_{q,l}^*} \right) \\
 &= K(\pi^*) = K^* < 0 ,
 \end{aligned}$$

it comes

$$P^* [T_1 > t] \leq P^* [T_{1,1} + K^* [r(2n - r - 1)] > t] .$$

Similarly, Assumption (A4) yields a constant $C(\gamma) > 0$ such that

$$\sum_i \log \frac{\widehat{\alpha}_{z_i}}{\widehat{\alpha}_{z_i^*}} \leq nC(\gamma) ,$$

which entails

$$\begin{aligned}
 &P^* [T_1 > t] \\
 &\leq P^* \left[\sum_{i \neq j} (X_{i,j} - \pi_{z_i^*, z_j^*}^*) \log \left(\frac{\pi_{z_i, z_j}^*}{\pi_{z_i^*, z_j^*}^*} \frac{1 - \pi_{z_i^*, z_j^*}^*}{1 - \pi_{z_i, z_j}^*} \right) > t - nC(\gamma) - r(2n - r - 1)K^* \right] .
 \end{aligned}$$

Another use of Hoeffding's inequality associated with (A3) provides a constant $C_\zeta > 0$ such that

$$P^* [T_1 > t] \leq \exp \left[- \frac{(t - nC(\gamma) - r(2n - r - 1)K^*)^2}{r(2n - r - 1)C_\zeta} \right] .$$

Upper bounding T_2

 Since $\widehat{\pi} = \pi^* + o_{\mathbb{P}}(1/n)$

$$\begin{aligned} T_2 &= \sum_{i \neq j} \log \left[1 + \frac{(\widehat{\pi}_{z_i, z_j} - \pi_{z_i, z_j}^*)(X_{i,j} - \pi_{z_i, z_j}^*)}{\pi_{z_i, z_j}^*(1 - \pi_{z_i, z_j}^*)} \right] \\ &= \sum_{i \neq j} \frac{(\widehat{\pi}_{z_i, z_j} - \pi_{z_i, z_j}^*)(X_{i,j} - \pi_{z_i, z_j}^*)}{\pi_{z_i, z_j}^*(1 - \pi_{z_i, z_j}^*)} (1 + o_{\mathbb{P}}(1)) . \end{aligned}$$

The terms of this sum have to be gathered in another way to yield the magnitude of T_2 :

$$\begin{aligned} |T_2| &\leq \|\widehat{\pi} - \pi^*\|_{\infty} \sum_{q,l} K_{q,l} (1 + o_{\mathbb{P}}(1)) \left\{ \left| \sum_{q',l'} \sum_{\substack{i \neq j \\ (z_i^*, z_j^*) = (q', l') \\ (z_i, z_j) = (q, l)}} (X_{i,j} - \pi_{q',l'}^*) \right| \right. \\ &\quad \left. + \left| \sum_{q',l'} N_{q',l'}^{q,l} (\pi_{q',l'}^* - \pi_{q,l}^*) \right| \right\} , \end{aligned}$$

where $N_{q',l'}^{q,l} = \sum_{i \neq j} \mathbb{1}_{(z_i^* = q', z_j^* = l')} \mathbb{1}_{(z_i = q, z_j = l)}$. Let us now introduce an event of large probability:

$$\Omega_n = \{\|\widehat{\pi} - \pi^*\|_{\infty} \leq 1/n\} . \quad (20)$$

Note that by assumption, $\mathbb{P}[\Omega_n^c] \xrightarrow[n \rightarrow \infty]{} 0$. Hence,

$$\begin{aligned} &P^*[\Omega_n \cap \{|T_2| > t\}] \\ &\leq P^* \left[\sum_{q,l} K_{q,l} (1 + o_{\mathbb{P}}(1)) \left\{ \left| \sum_{q',l'} \sum_{\substack{i \neq j \\ (z_i^*, z_j^*) = (q', l') \\ (z_i, z_j) = (q, l)}} (X_{i,j} - \pi_{q',l'}^*) \right| \right\} > nt/2 \right] \\ &\quad + P^* \left[1/n \sum_{q,l} K_{q,l} (1 + o_{\mathbb{P}}(1)) \left\{ \left| \sum_{q',l'} N_{q',l'}^{q,l} (\pi_{q',l'}^* - \pi_{q,l}^*) \right| \right\} > t/2 \right] . \end{aligned}$$

Let us deal with the first term.

$$\begin{aligned}
 & P^* \left[\sum_{q,l} K_{q,l} (1 + o_{\mathbb{P}}(1)) \left\{ \sum_{\substack{q',l' \\ (z_i^*, z_j^*) = (q', l') \\ (z_i, z_j) = (q, l)}} \sum_{i \neq j} (X_{i,j} - \pi_{q',l'}^*) \right\} > nt/2 \right] \\
 & \leq \sum_{q,l} P^* \left[(1 + o_{\mathbb{P}}(1)) \left| \sum_{\substack{q',l' \\ (z_i^*, z_j^*) = (q', l') \\ (z_i, z_j) = (q, l)}} \sum_{i \neq j} (X_{i,j} - \pi_{q',l'}^*) \right| > nt/(2K_{q,l}Q^2) \right] \\
 & \leq Q^2 \max_{q,l} P^* \left[\left| \sum_{\substack{q',l' \\ (z_i^*, z_j^*) = (q', l') \\ (z_i, z_j) = (q, l)}} \sum_{i \neq j} (X_{i,j} - \pi_{q',l'}^*) \right| > nt/(4K_{q,l}Q^2) \right].
 \end{aligned}$$

A straightforward use of Hoeffding inequality yields

$$\begin{aligned}
 & P^* \left[\left| \sum_{\substack{q',l' \\ (z_i^*, z_j^*) = (q', l') \\ (z_i, z_j) = (q, l)}} \sum_{i \neq j} (X_{i,j} - \pi_{q',l'}^*) \right| > nt/(4K_{q,l}Q^2) \right] \\
 & \leq \exp \left[-\frac{1}{(4K_{q,l}Q^2)^2} \frac{n^2 t^2}{N_{q,l}} \right].
 \end{aligned}$$

For the second term, it comes that

$$\begin{aligned}
 & P^* \left[1/n \sum_{q,l} K_{q,l} (1 + o_{\mathbb{P}}(1)) \left\{ \left| \sum_{q',l'} N_{q',l'}^{q,l} (\pi_{q',l'}^* - \pi_{q,l}^*) \right| \right\} > t/2 \right] \\
 & \leq Q^2 \max_{q,l} P^* \left[\left| \sum_{q',l'} N_{q',l'}^{q,l} (\pi_{q',l'}^* - \pi_{q,l}^*) \right| > nt/(4K_{q,l}Q^2) \right] \\
 & \leq Q^2 \max_{q,l} P^* \left[\max_{(q',l')} |\pi_{q,l}^* - \pi_{q',l'}^*| N_{q,l} > nt/(4K_{q,l}Q^2) \right].
 \end{aligned}$$

Hence,

$$\begin{aligned}
 P^* [\Omega_n \cap \{|T_2| > t\}] & \leq Q^2 \max_{q,l} \exp \left[-\frac{1}{(4K_{q,l}Q^2)^2} \frac{n^2 t^2}{N_{q,l}} \right] \\
 & \quad + Q^2 \max_{q,l} P^* \left[\max_{(q',l')} |\pi_{q,l}^* - \pi_{q',l'}^*| > nt/(4N_{q,l}K_{q,l}Q^2) \right].
 \end{aligned}$$

Upper bounding T_3

Since $\hat{\pi} = \pi^* + o_{\mathbb{P}}(1/n)$, it comes

$$\begin{aligned}
 & \log \left[1 + \frac{(\hat{\pi}_{z_i^*, z_j^*} - \pi_{z_i^*, z_j^*}^*)(X_{i,j} - \pi_{z_i^*, z_j^*}^*)}{\pi_{z_i^*, z_j^*}^*(1 - \pi_{z_i^*, z_j^*}^*)} \right] \\
 & = \frac{(\hat{\pi}_{z_i^*, z_j^*} - \pi_{z_i^*, z_j^*}^*)(X_{i,j} - \pi_{z_i^*, z_j^*}^*)}{\pi_{z_i^*, z_j^*}^*(1 - \pi_{z_i^*, z_j^*}^*)} (1 + o_{\mathbb{P}}(1)).
 \end{aligned}$$

Let us recall that since the likelihood is finite, the denominator is not null (and even bounded away from 0 and 1). Then, T_3 can be written as

$$T_3 = \sum_{q,l} \frac{(Y_{q,l} - \pi_{q,l}^* N_{q,l})(\hat{\pi}_{q,l} - \pi_{q,l}^*)}{\pi_{q,l}^*(1 - \pi_{q,l}^*)} (1 + o_{\mathbb{P}}(1)).$$

Using the event Ω_n defined by (20), one has

$$\begin{aligned}
 P^* [\Omega_n \cap \{|T_3| > t\}] &= P^* \left[\left| \sum_{q,l} \frac{(Y_{q,l} - \pi_{q,l}^* N_{q,l})1/n}{\pi_{q,l}^*(1 - \pi_{q,l}^*)} (1 + o_{P^*}(1)) \right| > t \right] \\
 &\leq \sum_{q,l} P^* \left[\left| \frac{(Y_{q,l} - \pi_{q,l}^* N_{q,l})}{n\pi_{q,l}^*(1 - \pi_{q,l}^*)} (1 + o_{P^*}(1)) \right| > t/Q^2 \right] \\
 &\leq Q^2 \max_{q,l} P^* \left[\left| \frac{(Y_{q,l} - \pi_{q,l}^* N_{q,l})}{n\pi_{q,l}^*(1 - \pi_{q,l}^*)} (1 + o_{P^*}(1)) \right| > t/Q^2 \right] \\
 &\leq Q^2 \max_{q,l} P^* \left[\frac{|Y_{q,l} - \pi_{q,l}^* N_{q,l}|}{\pi_{q,l}^*(1 - \pi_{q,l}^*)} > nt/(2Q^2) \right] \\
 &\leq Q^2 \max_{q,l} P^* \left[\frac{1}{N_{q,l}} \frac{|Y_{q,l} - \pi_{q,l}^* N_{q,l}|}{\pi_{q,l}^*(1 - \pi_{q,l}^*)} > \frac{nt}{2Q^2 N_{q,l}} \right] \\
 &\leq 2Q^2 \max_{q,l} \left\{ \exp \left[-N_{q,l} K^* \left(\frac{nt}{N_{q,l}} \right)^2 \right] \right\} \\
 &\leq 2Q^2 \max_{q,l} \left\{ \exp \left[-K^* \frac{n^2 t^2}{N_{q,l}} \right] \right\} .
 \end{aligned}$$

Gathering T_1 -, T_2 -, and T_3 -upper bounds

Then, for any $\epsilon > 0$

$$\begin{aligned}
 &P^* \left[\sum_{z_{[n]} \neq z_{[n]}^*} \frac{\widehat{P}^{X_{[n]}}(z_{[n]})}{\widehat{P}^{X_{[n]}}(z_{[n]}^*)} > \epsilon \right] \\
 &\leq P^* \left[\left\{ \sum_{z_{[n]} \neq z_{[n]}^*} \frac{\widehat{P}^{X_{[n]}}(z_{[n]})}{\widehat{P}^{X_{[n]}}(z_{[n]}^*)} > \epsilon \right\} \cap \Omega_n \right] + P^* [\Omega_n^c] .
 \end{aligned}$$

Furthermore,

$$\begin{aligned}
 & P^* \left[\left\{ \sum_{z_{[n]} \neq z_{[n]}^*} \frac{\widehat{P}^{X_{[n]}}(z_{[n]})}{\widehat{P}^{X_{[n]}}(z_{[n]}^*)} > \epsilon \right\} \cap \Omega_n \right] \\
 & \leq \sum_{r=1}^n \sum_{\|z_{[n]} - z_{[n]}^*\|_0 = r} P^* \left[\left\{ \frac{\widehat{P}^{X_{[n]}}(z_{[n]})}{\widehat{P}^{X_{[n]}}(z_{[n]}^*)} > \epsilon / (n^{r+1} Q^r) \right\} \cap \Omega_n \right] \\
 & \leq \sum_{r=1}^n \sum_{\|z_{[n]} - z_{[n]}^*\|_0 = r} P^* \left[\left\{ \log \frac{\widehat{P}^{X_{[n]}}(z_{[n]})}{\widehat{P}^{X_{[n]}}(z_{[n]}^*)} > -(r+1) \log n - r \log Q + \log \epsilon \right\} \cap \Omega_n \right] \\
 & \leq \sum_{r=1}^n \sum_{\|z_{[n]} - z_{[n]}^*\|_0 = r} P^* \left[\left\{ \log \frac{\widehat{P}^{X_{[n]}}(z_{[n]})}{\widehat{P}^{X_{[n]}}(z_{[n]}^*)} > -5r \log n \right\} \cap \Omega_n \right] \quad (n \gg 1) \\
 & = \sum_{r=1}^n \sum_{\|z_{[n]} - z_{[n]}^*\|_0 = r} P^* [\{T_1 + T_2 - T_3 > -5r \log n\} \cap \Omega_n] .
 \end{aligned}$$

It remains to deal with $P^* [\{T_1 + T_2 - T_3 > -5r \log n\} \cap \Omega_n]$:

$$\begin{aligned}
 & P^* [\{T_1 + T_2 - T_3 > -5r \log n\} \cap \Omega_n] \\
 & \leq P^* [\{T_1 + T_2 - T_3 > -5r \log n\} \cap \Omega_n \cap \{|T_3| \leq r \log n\}] \\
 & \quad + P^* [\{|T_3| > r \log n\} \cap \Omega_n] \\
 & \leq P^* [\{T_1 + T_2 > -6r \log n\} \cap \Omega_n] + P^* [\{|T_3| > r \log n\} \cap \Omega_n] \\
 & \leq P^* [T_1 > -7r \log n] + P^* [\{|T_2| > r \log n\} \cap \Omega_n] \\
 & \quad + P^* [\{|T_3| > r \log n\} \cap \Omega_n] .
 \end{aligned}$$

Combining the previous bounds of T_1 , T_2 , and T_3 with the above inequality, it

comes

$$\begin{aligned}
 & P^* [\{T_1 + T_2 - T_3 > -r \log n\} \cap \Omega_n] \\
 \leq & \exp \left[-\frac{[nC(\gamma) + 7r \log n + r(2n - r - 1)K^*]^2}{r(2n - r - 1)C_\zeta} \right] \\
 & + 2Q^2 \max_{q,l} \exp \left[-\frac{1}{(4K_{q,l}Q^2)^2} \frac{n^2 (r \log n)^2}{N_{q,l}} \right] \\
 & + 2Q^2 \max_{q,l} P^* \left[\zeta > \frac{1}{4K_{q,l}Q^2} \frac{nr \log n}{N_{q,l}} \right] \\
 & + 2Q^2 \max_{q,l} \exp \left[-K^* \frac{n^2 (r \log n)^2}{N_{q,l}} \right] \\
 \leq & \exp \left[-r(2n - r - 1) \frac{K^{*2}}{C_\zeta} \right] \\
 & + 2Q^2 \max_{q,l} \exp \left[-\frac{1}{(4K_{q,l}Q^2)^2} \frac{n^2 r (\log n)^2}{(2n - r - 1)} \right] \\
 & + 2Q^2 \max_{q,l} P^* \left[\zeta > \frac{1}{4K_{q,l}Q^2} \frac{n \log n}{(2n - r - 1)} \right] \\
 & + 2Q^2 \max_{q,l} \exp \left[-K^* \frac{n^2 r (\log n)^2}{(2n - r - 1)} \right] .
 \end{aligned}$$

One gets that for n large enough,

$$\begin{aligned}
 & P^* [\{T_1 + T_2 - T_3 > -r \log n\} \cap \Omega_n] \\
 \leq & \exp \left[-r(2n - r - 1) \frac{K^{*2}}{C_\zeta} \right] + 4Q^2 \exp [-\bar{K}r(\log n)^2] \\
 \leq & \bar{K}_2 \exp [-\bar{K}r(\log n)^2] .
 \end{aligned}$$

The end of the proof follows from the same argument as in the proof of Theorem 3.1.

□

Appendix F: Proof of Theorem 4.2

Lemma F.1. *Let $\hat{z}_{[n]} = \hat{z}_{[n]}(\pi) = \text{Argmax}_{z_{[n]}} \mathcal{L}_1(X_{[n]}; z_{[n]}, \pi)$. For every $X_{[n]} \in \mathcal{X}_n$, $(\alpha, \pi) \in \Theta$, and $\tau_{[n]} \in \mathcal{S}_n$, it comes that*

$$\mathcal{J}(X_{[n]}; \tau_{[n]}, \alpha, \pi) \leq \mathcal{L}_2(X_{[n]}; \alpha, \pi) \leq \mathcal{L}_1(X_{[n]}; \hat{z}_{[n]}, \pi) .$$

Proof of Lemma F.1. The first inequality comes from the definition of \mathcal{J} . The second one results from $\hat{z}_{[n]}(\pi) = \text{Argmax}_{z_{[n]}} \mathcal{L}_1(X_{[n]}; z_{[n]}, \pi)$.

For any (α, π) ,

$$\begin{aligned} \mathcal{L}_2(X_{[n]}; \alpha, \pi) &= \log \left\{ \sum_{z_{[n]} \in \mathcal{Z}_n} e^{\mathcal{L}_1(X_{[n]}; z_{[n]}, \pi)} P_{Z_{[n]}}(z_{[n]}) \right\} \\ &\leq \log \left\{ e^{\mathcal{L}_1(X_{[n]}; \hat{z}_{[n]}, \pi)} \sum_{z_{[n]} \in \mathcal{Z}_n} P_{Z_{[n]}}(z_{[n]}) \right\} \\ &\leq \mathcal{L}_1(X_{[n]}; \hat{z}_{[n]}, \pi) . \end{aligned}$$

□

Lemma F.2. *Lemma F.1 and Assumption (A4) entail that there exists $0 < \gamma < 1$ such that for every (α, π) ,*

$$\begin{aligned} |\mathcal{L}_2(X_{[n]}; \alpha, \pi) - \mathcal{L}_1(X_{[n]}; \hat{z}_{[n]}, \pi)| &\leq n \log(1/\gamma) , \\ |\mathcal{J}(X_{[n]}; \hat{\tau}_{[n]}, \alpha, \pi) - \mathcal{L}_1(X_{[n]}; \hat{z}_{[n]}, \pi)| &\leq n \log(1/\gamma) . \end{aligned}$$

Proof of Lemma F.2. From Lemma F.1 and definition of $\hat{\tau}_{[n]}$ it comes for every (α, π) :

$$\mathcal{J}(X_{[n]}; \hat{\tau}_{[n]}, \alpha, \pi) \leq \mathcal{J}(X_{[n]}; \hat{\tau}_{[n]}, \alpha, \pi) \leq \mathcal{L}_2(X_{[n]}; \alpha, \pi) \leq \mathcal{L}_1(X_{[n]}; \hat{z}_{[n]}, \pi) .$$

Combined with $\mathcal{J}(X_{[n]}; \hat{\tau}_{[n]}, \alpha, \pi) = \mathcal{L}_1(X_{[n]}; \hat{z}_{[n]}, \pi) + \sum_{i=1}^n \log \alpha_{\hat{z}_i}$, it leads to both

$$\begin{aligned} |\mathcal{L}_2(X_{[n]}; \alpha, \pi) - \mathcal{L}_1(X_{[n]}; \hat{z}_{[n]}, \pi)| &\leq - \sum_{i=1}^n \log \alpha_{\hat{z}_i} , \\ |\mathcal{J}(X_{[n]}; \hat{\tau}_{[n]}, \alpha, \pi) - \mathcal{L}_1(X_{[n]}; \hat{z}_{[n]}, \pi)| &\leq - \sum_{i=1}^n \log \alpha_{\hat{z}_i} . \end{aligned}$$

Assumption (A4) yields the conclusion.

□

References

- Allman, E., Matias, C., and Rhodes, J. (2009). Identifiability of parameters in latent structure models with many observed variables. *Annals of Statistics*, **37**, 3099–3132.
- Allman, E., Matias, C., and Rhodes, J. A. (2011). Parameter identifiability in a class of random graph mixture models. *Journal of Statistical Planning and Inference*, **141**(5), 1719–1736.
- Ambroise, C. and Matias, C. (2010). New consistent and asymptotically normal estimators for random graph mixture models. Technical Report 1003.5165, arXiv.
- Andrieu, C. and Atchadé, Y. F. (2007). On the efficiency of adaptive mcmc algorithms. *Elec. Comm. in Probab.*, **12**, 336–349.
- Bickel, P. and Chen, A. (2010). A nonparametric view of network models and newman-girvan and other modularities. *PNAS*, pages 1–6.
- Boer, P., Huisman, M., Snijders, T., Steglich, C., Wichers, L., and Zeggelink, E. (2006). Stocnet: an open software system for the advanced statistical analysis of social networks, version 1.7. groningen: Ics/scienceplus.
- Choi, D. S., Wolfe, P. J., and Airoldi, E. M. (2011). Stochastic blockmodels with growing number of classes. Technical Report 1011.4644v2, arXiv.
- Daudin, J.-J., Picard, F., and Robin, S. (2008). A mixture model for random graphs. *Stat Comput*, **18**, 173–183.
- Gazal, S., Daudin, J.-J., and Robin, S. (2011). Accuracy of variational estimates for random graph mixture models. *J. Comput. Simul.*
- Holland, P., Laskey, K., and Leinhardt, S. (1983). Stochastic blockmodels: Some first steps. *Social Networks*, **5**, 109–137.
- Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., and Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine Learning*, **37**, 183–233.
- Mariadassou, M., Robin, S., and Vacher, C. (2010). Uncovering latent structure in valued graphs: A variational approach. *Ann. Appl. Stat.*, **4**, 715–742.
- Massart, P. (2007). *Concentration Inequalities and Model Selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard.
- Mixnet (2009). <http://www.cran.rproject.org>.
- Nowicki, K. and Snijders, T. (2001). Estimation and prediction for stochastic block-structures. *J. Am. Stat. Assoc.*, **96**, 1077–1087.
- Picard, F., Miele, V., Daudin, J.-J., Cottret, L., and Robin, S. (2009). Deciphering the connectivity structure of biological networks using mixnet. *BMC Bioinformatics*, **10**.
- Rohe, K., Chatterjee, S., and Yu, B. (2010). Spectral clustering and the high-dimensional stochastic blockmodel. Technical Report 791, Berkeley.
- Snijders, T. A. B. and Nowicki, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification*, **14**, 75–100.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes With Applications to Statistics*. Springer Series in Statistics.