



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-00593644v3

HAL Id: hal-00593644

<https://hal.science/hal-00593644v3>

Preprint submitted on 29 Sep 2012

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 in 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 derived. In particular, the consistency of these estimators is settled for the probability of an edge between two vertices (and for the group proportions at the price of an additional assumption), which is to the best of our knowledge the first result of this type for variational estimators in 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.

*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

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 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 realizations of 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 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, 2012). 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 studied (Bickel and Chen, 2009). 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.*, 2011; Choi *et al.*, 2012).

The present work deals with directed (and undirected) binary edges in random graphs drawn from SBM.

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*. Numerous approaches

have been developed to overcome this challenging problem, but most of them suffer some high computational cost. For instance [Snijders and Nowicki \(1997\)](#) study maximum-likelihood estimators of SBM with only two classes and *binary undirected* graphs, while [Nowicki and Snijders \(2001\)](#) perform Gibbs sampling for more than two classes at the price of a large computational cost. Other strategies also exist relying for instance on *profile-likelihood optimization* in CSBM ([Bickel and Chen, 2009](#)), on a *spectral clustering algorithm* in CSBM ([Rohe et al., 2011](#)), or on *moment estimation* in a particular instance of SBM called *affiliation model* ([Ambroise and Matias, 2012](#), and also [Example 1](#) in the present paper).

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 (see [Mixnet \(2009\)](#) 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 already exist for *maximum likelihood* or *variational* estimators of SBM parameters. Note however that consistency results for maximum likelihood estimators in the CSBM have been derived recently by [Choi et al. \(2012\)](#) where the number of groups is allowed to grow with the number of vertices. 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. Note that identifiability of directed SBM is really challenging since existing strategies such as that of [Allman et al. \(2009\)](#) cannot be extended easily. 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 consistent for estimating the probability π of an edge between two vertices. When estimating the group proportions α , an additional assumption on the convergence rate of $\hat{\pi}$ is required to derive consistency. The present framework 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 $X = \{X_{i,j}\}_{i,j \in \mathbb{N}}$, is given by: for $i \neq j$, $X_{i,j} = 1$ if an edge exists from vertex i to vertex j and $X_{i,j} = 0$ otherwise, and $X_{i,i} = 0$ (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

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

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

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

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

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

where

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

and $\mathbb{P}[Z_{[n]} = z_{[n]}] = \mathbb{P}_\alpha[Z_{[n]} = z_{[n]}] = \prod_{i=1}^n \alpha_{z_i}$. 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. For every $q \neq q'$, there exists $l \in \{1, \dots, Q\}$ such that

$$\pi_{q,l} \neq \pi_{q',l} \text{ or } \pi_{l,q} \neq \pi_{l,q'} . \quad (\text{A1})$$

Following (A1), the matrix π cannot have two columns equal and the corresponding rows also equal. This constraint 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 2. There exists $\zeta > 0$ such that

$$\forall (q, l) \in \{1, \dots, Q\}^2, \quad \pi_{q,l} \in]0, 1[\Rightarrow \pi_{q,l} \in [\zeta, 1 - \zeta] . \quad (\text{A2})$$

SBM can deal with null probabilities of connection between vertices. However, the use of $\log \pi_{q,l}$ implies a special treatment for $\pi_{q,l} \in \{0, 1\}$. Note that all along the present paper, (A2) is always assumed to hold with ζ not depending on n .

Assumption 3. There exists $0 < \gamma < 1/Q$ such that

$$\forall q \in \{1, \dots, Q\}, \quad \alpha_q \in [\gamma, 1 - \gamma] . \quad (\text{A3})$$

This assumption implies that no class is drained. Actually the identifiability of SBM (Theorem 2.1) requires every $\alpha_q \in (0, 1)$ for $q \in \{1, \dots, Q\}$, which is implied by (A3). In this paper, it is assumed that γ does not depend on n .

Assumption 4. There exist $0 < \gamma < 1/Q$ and $n_0 \in \mathbb{N}^*$ such that any realization of SBM (Section 2.1) with label vector $z_{[n]}^* = (z_1^*, \dots, z_n^*)$ satisfies

$$\forall q \in \{1, \dots, Q\}, \forall n \geq n_0, \quad \frac{N_q(z_{[n]}^*)}{n} \geq \gamma , \quad (\text{A4})$$

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

Note that (A4) is the empirical version of (A3). By definition of SBM, $z_{[n]}^*$ is the realization of a multinomial random variable with parameters $(\alpha_1, \dots, \alpha_Q)$. Any multinomial random variable will satisfy the requirement of (A4) 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 SBM parameters 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 $r = \pi \cdot \alpha$ are distinct. Then, SBM parameters are identifiable.*

The assumption on vector $\pi \cdot \alpha$ 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 $r' = \pi^t \cdot \alpha$ (instead of $r = \pi \cdot \alpha$), and 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 also emphasize that Assumption (A1) is implied by assuming either $\pi \cdot \alpha$ or $\pi^t \cdot \alpha$ has distinct coordinates (which leads to identifiability). Note that Bickel *et al.* (2011, Theorem 2, Section 3.1) also recently derived an identifiability result for “block models” in terms of “wheel motifs”.

Let us further illustrate the assumption on $\pi \cdot \alpha$ through two examples. The first one is a particular instance of SBM called *Affiliation Model* (Allman *et al.*, 2011) restricted to the setting where $Q = 2$.

Example 1 (Affiliation model). *From a general point of view, affiliation model is used with Q populations of vertices and considers undirected graphs (π symmetric). The present example focuses on a particular instance where $Q = 2$. 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 $\alpha_1 \neq \alpha_2$. Indeed 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 ($\alpha_1 \neq \alpha_2$).

*Note that Allman *et al.* (2011) have derived a result on identifiability for affiliation models with equal group proportions.*

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 σ of $\{1, \dots, Q\}$,

$$\pi^\sigma \alpha^\sigma = \pi \alpha , \tag{3}$$

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. 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,

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

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

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

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 $X_{i,i} = 0$ by assumption. Hence given $P_{[n]}$, $u_0 = 1$ and u_1, \dots, u_{2Q-1} are known.

Furthermore, 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

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

Defining the diagonal matrix $A = \text{Diag}(\alpha)$, it comes that $M_Q = RAR^t$, 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 $Q = 2$, (A1) implies 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 of labels z and z' :

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

Then $z \stackrel{\pi}{\sim} z'$ is equivalent to $[z]_{\pi} = [z']_{\pi}$, where $[z]_{\pi}$ denotes the equivalence class of z . As a consequence, any vectors $z_{[n]}$ and $z'_{[n]}$ in the same class have the same conditional likelihood (2):

$$\mathcal{L}_1(X_{[n]}; z_{[n]}, \pi) = \mathcal{L}_1(X_{[n]}; z'_{[n]}, \pi) .$$

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.

3.1.2. Main asymptotic result

Let $P^* := \mathbb{P}(\cdot \mid Z = z^*) = P_{\alpha^*, \pi^*}^*$ denote the *true* conditional distribution given the (equivalence class of the) whole label sequence, the notation emphasizing that P^* depends on (α^*, π^*) .

The following Theorem 3.1 provides the convergence rate of $\mathbb{P}\left(Z_{[n]} = z_{[n]}^* \mid X_{[n]}\right) = \mathbb{P}_{\alpha^*, \pi^*}\left(Z_{[n]} = z_{[n]}^* \mid X_{[n]}\right)$ 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 (A1)–(A4) hold. For every $t > 0$,*

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

where $\kappa > 0$ is a constant depending on π^* but not on z^* , and the $\mathcal{O}(ne^{-\kappa n})$ is uniform with respect to z^* .

Furthermore, the same result holds with P^* replaced by \mathbb{P} under (A1)–(A3).

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 consistency 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 \mathbb{P} -almost sure convergence.

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

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

Moreover,

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

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

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

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.

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

where $\mathcal{D}(Z_{[n]} \mid X_{[n]})$ 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 for $\mathcal{Z} = \{1, \dots, Q\}^{\mathbb{N}^*}$, $(\mathcal{E}(\mathcal{Z}), d)$ 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 $\mathcal{E}(\mathcal{Z})$, so that $\mathcal{E}(\mathcal{Z}_n)$ is 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 $\mathbb{P}^n = \mathbb{P}[\cdot \mid X_{[n]}]$ 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,

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

by use of Corollary 3.2, 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. □

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 of \mathcal{L}_1 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: (i) the consistency of the π estimator is addressed by use of an approach based on M-estimators, (ii) a result similar to Theorem 3.1 is combined with a “deconditioning” argument to get the consistency of the α^* estimator (Theorem 3.9) at the price of an additional assumption on the rate of convergence of the estimator $\hat{\pi}$ of π^* .

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

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

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

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

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

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

□

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

The main point of our approach consists in using $P^* = P_{\alpha^*, \pi^*}^*$ (Section 3.1.2) 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

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

where the expectation is computed with respect to $P^* = P_{\alpha^*, \pi^*}^*$. Actually our strategy (using Theorem 3.4) only requires to prove ϕ_n and Φ_n are uniformly

close to each other on a subset of parameters denoted by \mathcal{P} (see also the proof of Theorem 3.6 for more details) and defined as follows

$$\mathcal{P} = \{(z_{[n]}, \pi) \mid (\mathbf{A1}), (\mathbf{A2}), |\Phi_n(z_{[n]}, \pi)| < +\infty\} . \quad (7)$$

Showing this uniform convergence between ϕ_n and Φ_n over \mathcal{P} is precisely the purpose of Proposition 3.5. Its proof, which is deferred to Appendix C, strongly relies on Talagrand's concentration inequality (Massart, 2007).

Proposition 3.5. *With the above notation, let us assume (A1) and (A2) hold true. Then,*

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

Actually Proposition 3.5 is crucial to prove the following theorem which settles the desired properties for $\mathcal{L}_2(X_{[n]}; \alpha, \pi)$, that is (5) (uniform convergence) and (6) (well-identifiability).

Theorem 3.6. *Let us assume that (A1), (A2), and (A3) 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 $\mathcal{A} = \{A = (a_{i,j})_{1 \leq i,j \leq Q} \mid a_{q,q'} \geq 0, \sum_{q'} a_{q,q'} = 1\} \subset \mathcal{M}_Q(\mathbb{R})$. 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. Its uniform convergence part exploits the connection between $\phi_n(z_{[n]}, \pi)$ and $\mathcal{L}_2(X_{[n]}; \alpha, \pi)$ (Proposition 3.5).

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 the 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^* .

Our proof heavily relies on an analogous result to Theorem 3.1, where the true value (α^*, π^*) of SBM parameters is replaced by an estimator $(\hat{\alpha}, \hat{\pi})$. In what follows, $\hat{P}(Z_{[n]} = z_{[n]} | X_{[n]}) = \mathbb{P}_{\hat{\alpha}, \hat{\pi}}(Z_{[n]} = z_{[n]} | X_{[n]})$ (Section 3.1.2 and Lemma E.2) denotes the same quantity as $\mathbb{P}(Z_{[n]} = z_{[n]} | X_{[n]})$ where (α^*, π^*) has been replaced by $(\hat{\alpha}, \hat{\pi})$. Let us state this result in a general framework since it will be successively used in proofs of Theorems 3.9 and 4.4.

Proposition 3.8. *Let us assume that assumptions (A1)–(A4) hold, and that there exists an estimator $\hat{\pi}$ such that $\|\hat{\pi} - \pi^*\|_\infty = o_{\mathbb{P}}(v_n)$, with $v_n = o(\sqrt{\log n/n})$. Let also $\hat{\alpha}$ denote any estimator of α^* . Then for every $\epsilon > 0$,*

$$P^* \left[\sum_{z_{[n]} \neq z_{[n]}^*} \frac{\hat{P}(Z_{[n]} = z_{[n]} | X_{[n]})}{\hat{P}(Z_{[n]} = z_{[n]}^* | X_{[n]})} > \epsilon \right] \leq \kappa_1 n e^{-\kappa_2 \frac{(\log n)^2}{n v_n^2}} + \mathbb{P}[\|\hat{\pi} - \pi^*\|_\infty > v_n]$$

for n large enough, where $\kappa_1, \kappa_2 > 0$ are constants independent of z^* , and

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

Moreover, the same result holds replacing P^* by \mathbb{P} under (A1)–(A3).

The proof of Proposition 3.8 is given in Appendix E.

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 $z_{[n]}^*$. Note that the novelty of Proposition 3.8 compared to Theorem 3.1 lies in the convergence rate which depends on the behavior of $\hat{\pi}$. This is the reliable price for estimating rather than knowing π^* .

We assume $v_n = o(\sqrt{\log n/n})$, which arises from the proof as a necessary requirement for consistency. However, we do not know whether this is a *necessary* or only a *sufficient* condition. Furthermore there is still 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 the present paper.

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

Theorem 3.9. *Let $(\hat{\alpha}, \hat{\pi})$ denote the MLE of (α^*, π^*) and assume $\|\hat{\pi} - \pi^*\|_\infty = o_{\mathbb{P}}(\sqrt{\log n/n})$. 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 .

Note that the rate $1/n$ would be reached in “classical” parametric models with n^2 independent random variables.

Proof of Theorem 3.9. In the mixture model framework of SBM, Lemma E.2 shows 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 given $Z_{[n]} = z_{[n]}^*$. Setting $N_q(z_{[n]}) = \sum_{i=1}^n \mathbb{1}_{(z_i=q)}$, it comes

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

Note the last inequality results from

$$\begin{aligned} \frac{1}{n} \sum_i \hat{P}(Z_i \neq z_i^* \mid X_{[n]}) &\leq \max_{i=1, \dots, n} \hat{P}(Z_i \neq z_i^* \mid X_{[n]}) \\ &\leq \hat{P}(\cup_{i=1}^n (Z_i \neq z_i^*) \mid X_{[n]}) = \hat{P}(Z_{[n]} \neq z_{[n]}^* \mid X_{[n]}) \quad . \end{aligned}$$

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

for every $\epsilon > 0$,

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

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

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

□

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 :

$$\mathcal{L}_2(X_{[n]}; \alpha, \pi) = \log \left\{ \sum_{z_{[n]} \in \mathcal{Z}_n} e^{\sum_{i \neq j} b_{ij}(z_i, z_j)} P_{Z_{[n]}}(z_{[n]}) \right\} \ , \quad (8)$$

with $b_{ij}(z_i, z_j) = X_{i,j} \log \pi_{z_i, z_j} + (1 - X_{i,j}) \log(1 - \pi_{z_i, z_j})$. To circumvent this problem, alternatives are for instance 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 can deal with thousands of vertices in a reasonable computation time thanks to its complexity in $\mathcal{O}(n^2)$. For instance, Mixnet (2009) package (based on variational approximation) deals with up to several thousands of vertices, whereas STOCNET package (see

Boer *et al.*, 2006) (Gibbs sampling) only deals with a few hundreds of vertices. Note that other approaches based on *profile-likelihood* have been recently developed and studied for instance by Bickel and Chen (2009).

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 (see (9)). The computational virtue of this trick is to replace an intractable sum over Q^n terms (see (8)) by a sum over only n^2 terms (Eq. (11)).

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

$$\mathcal{D}_n = \left\{ D_{\tau_{[n]}} = \prod_{i=1}^n \mathcal{M}(1, \tau_{i,1}, \dots, \tau_{i,Q}) \mid \tau_{[n]} \in \mathcal{S}_n \right\} , \quad (9)$$

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

$$\mathcal{J}(X_{[n]}; \tau_{[n]}, \alpha, \pi) = \mathcal{L}_2(X_{[n]}; \alpha, \pi) - K(D_{\tau_{[n]}}; P^{X_{[n]}}) , \quad (10)$$

where $K(\cdot, \cdot)$ denotes the Kullback-Leibler divergence, and $P^{X_{[n]}} = \mathbb{P}(Z_{[n]} = \cdot | X_{[n]})$. With this choice of \mathcal{D}_n , $\mathcal{J}(X_{[n]}; \tau_{[n]}, \alpha, \pi)$ has the following expression (see Daudin *et al.* (2008) and the proof of Lemma F.3):

$$\mathcal{J}(X_{[n]}; \tau_{[n]}, \alpha, \pi) = \sum_{i \neq j} \sum_{q,l} b_{ij}(q,l) \tau_{i,q} \tau_{j,l} - \sum_{iq} \tau_{i,q} (\log \tau_{i,q} - \log \alpha_q) , \quad (11)$$

where $b_{ij}(q,l) = X_{i,j} \log \pi_{q,l} + (1 - X_{i,j}) \log(1 - \pi_{q,l})$. The main improvement of Eq. (11) upon Eq. (8) is that $\mathcal{J}(X_{[n]}; \tau_{[n]}, \alpha, \pi)$ can be fully computed for every $(\tau_{[n]}, \alpha, \pi)$. The variational approximation $R_{X_{[n]}}$ to $P^{X_{[n]}}$ is given by solving the minimization problem over \mathcal{D}_n :

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

as long as such a minimizer exists, which amounts to maximizing $\mathcal{J}(X_{[n]}; \tau_{[n]}, \alpha, \pi)$ as follows

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

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

$$(\tilde{\alpha}, \tilde{\pi}) = \text{Argmax}_{\alpha, \pi} \mathcal{J}(X_{[n]}; \hat{\tau}_{[n]}, \alpha, \pi) . \quad (12)$$

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). Furthermore since $(\alpha, \pi) \mapsto \mathcal{J}(X_{[n]}; \hat{\tau}_{[n]}, \alpha, \pi)$ is not concave, this variational algorithm can lead to local optima in the same way as for likelihood optimization.

In the sequel, the same notation as in Section 3 is used. In particular it is assumed that a realization of SBM is observed, which has been generated from the sequence of true labels $Z = z^*$. In this setting, $P^* = P_{\alpha^*, \pi^*}^*$ (Section 3.1.2) denotes the conditional distribution $\mathbb{P}(\cdot | Z = z^*)$ given the whole label sequence. The first result provides some assurance about the reliability of the variational approximation to $P^{X_{[n]}} = \mathbb{P}_{\alpha^*, \pi^*} [Z_{[n]} = \cdot | X_{[n]}]$ (Section 3.1.2).

Proposition 4.1. *For every n , let \mathcal{D}_n denote the set defined by (9), and $P^{X_{[n]}}(\cdot)$ be the distribution of $Z_{[n]}$ given $X_{[n]}$. Then, assuming (A1)–(A3) 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,

$$K(R_{X_{[n]}}, P^{X_{[n]}}) \leq K(\delta_{z_{[n]}^*}, P^{X_{[n]}}) = -\log \left[P \left(Z_{[n]} = z_{[n]}^* | X_{[n]} \right) \right] ,$$

since

$$\begin{aligned} K(\delta_{z_{[n]}^*}, P^{X_{[n]}}) &= \sum_{z_{[n]}} \delta_{z_{[n]}^*}(z_{[n]}) \log \left[\frac{\delta_{z_{[n]}^*}(z_{[n]})}{P^{X_{[n]}}(z_{[n]})} \right] \\ &= -\log \left[P \left(Z_{[n]} = z_{[n]}^* | X_{[n]} \right) \right] . \end{aligned}$$

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

4.2. Consistency of the variational estimators

Since the variational log-likelihood $\mathcal{J}(\cdot; \cdot, \cdot, \cdot)$ (10) 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 (12)) is addressed first. The consistency of the VE of α^* ($\tilde{\alpha}$, see (12)) is handled in a second step and depends on the convergence rate of the estimator of $\tilde{\pi}$.

The first step consists in applying Theorem 3.4 to settle the $\tilde{\pi}$ consistency. Following results aim at justifying the use of Theorem 3.4 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, (A2) and (A3) lead to

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

where the supremum is computed over sets fulfilling (A2) and (A3).

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 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.

From the definitions of \mathcal{L}_1 , \mathcal{L}_2 , and \mathcal{J} (respectively given by Eq. (2), Eq. (1), and Eq. (10)) and recalling $\hat{z}_{[n]} = \hat{z}_{[n]}(\pi) = \text{Argmax}_{z_{[n]}} \mathcal{L}_1(X_{[n]}; z_{[n]}, \pi)$, Lemma F.1 leads to

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

Then applying (A3) 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 a simple consequence of Theorem 4.2, Proposition 3.5, and Theorem 3.4.

Corollary 4.3. *With the notation of Theorem 4.2 and assuming (A1), (A2), and (A3) 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.

Finally, the consistency for the VE of α^* is derived from the same *deconditioning argument* as that one used for the MLE (proof of Theorem 3.9). Consistency for $\tilde{\alpha}$ is stated by the following result where a convergence rate of $1/n$ is assumed for $\tilde{\pi}$. Note that at least some empirical evidence and heuristics exist (see Gazal *et al.*, 2011) in favour of this rate.

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 (A1), (A2), and (A3) hold, then*

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

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

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

Proof of Theorem 4.4.

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 (12)). 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 (9), 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,

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

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,

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

It remains to show $D_{\tilde{\tau}_{[n]}}(z_{[n]}^*) \xrightarrow[n \rightarrow \infty]{P^*} 1$ at a rate which does not depend of $z_{[n]}^*$. Let $\tilde{P} = \mathbb{P}_{\tilde{\alpha}, \tilde{\pi}}(Z_{[n]} = \cdot \mid X_{[n]})$ denote the *a posteriori* distribution of $Z_{[n]}$ with parameters $(\tilde{\alpha}, \tilde{\pi})$ (Section 3.1.2). Since Lemma F.4 provides

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

the conclusion results from another use of Proposition 3.8 applied with $\hat{\pi} = \tilde{\pi}$ and $v_n = 1/n$. □

5. Conclusion

This paper provides theoretical (asymptotic) results about the stochastic block model (SBM) inference. Identifiability of SBM parameters has been proved for directed (and undirected) graphs. This is typically the 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 (up to an additional assumption for the group proportions). 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 crucial 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

B.1. Preliminaries

Assuming **(A1)** holds true, π^* can be permutation-invariant (see Section 3.1.1). For this reason, we will consider equivalence classes denoted by $[z_{[n]}] = [z_{[n]}]_{\pi^*}$ for the label vector $z_{[n]}$.

Let us define $P^{X_{[n]}}(z_{[n]}) = \mathbb{P}[Z_{[n]} = z_{[n]} \mid X_{[n]}]$ for every label vector $z_{[n]}$, and $P^{X_{[n]}}([z_{[n]}]) = \mathbb{P}([Z_{[n]}] = [z_{[n]}] \mid X_{[n]})$ for corresponding class $[z_{[n]}]$. Since every $z'_{[n]} \in [z_{[n]}]$ satisfies $P^{X_{[n]}}(z'_{[n]}) = P^{X_{[n]}}(z_{[n]})$, it results that

$$P^{X_{[n]}}([z_{[n]}]) = \sum_{z'_{[n]} \in [z_{[n]}]} P^{X_{[n]}}(z'_{[n]}) = |[z_{[n]}]| P^{X_{[n]}}(z_{[n]}) , \quad (13)$$

where $|[z_{[n]}]|$ denotes the cardinality of $[z_{[n]}]$.

B.2. Upper bounding $\mathbb{P}\left[\sum_{[z_{[n]}] \neq [z^*_{[n]}]} \frac{\mathbb{P}([Z_{[n]}] = [z_{[n]}] \mid X_{[n]})}{\mathbb{P}([Z_{[n]}] = [z^*_{[n]}} \mid X_{[n]})} > t \mid \mathbf{Z} = z^*\right]$

Using P^* instead of $\mathbb{P}[\cdot \mid Z = z^*]$ for simplicity, let us first notice

$$\begin{aligned} \sum_{[z_{[n]}] \neq [z^*_{[n]}]} \frac{P^{X_{[n]}}([z_{[n]}])}{P^{X_{[n]}}([z^*_{[n]}])} &= \sum_{[z_{[n]}] \neq [z^*_{[n]}]} \frac{\sum_{z'_{[n]} \in [z_{[n]}]} P^{X_{[n]}}(z_{[n]})}{\sum_{z^0_{[n]} \in [z^*_{[n]}]} P^{X_{[n]}}(z^0_{[n]})} \\ &\leq \sum_{[z_{[n]}] \neq [z^*_{[n]}]} \sum_{z'_{[n]} \in [z_{[n]}]} \frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z^*_{[n]})} \\ &= \sum_{z_{[n]} \notin [z^*_{[n]}]} \frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z^*_{[n]})} , \end{aligned}$$

by (13) applied to $[z^*_{[n]}]$. Partitioning according to the number $\|z_{[n]} - z^*_{[n]}\|_0 = r$ of differences between $z_{[n]}$ and $z^*_{[n]}$, it comes

$$\sum_{[z_{[n]}] \neq [z^*_{[n]}]} \frac{P^{X_{[n]}}([z_{[n]}])}{P^{X_{[n]}}([z^*_{[n]}])} = \sum_{r=1}^n \sum_{\substack{z_{[n]} \notin [z^*_{[n]}] \\ \|z_{[n]} - z^*_{[n]}\|_0 = r}} \frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}([z^*_{[n]}])} . \quad (14)$$

Note that the number of vectors $z_{[n]}$ such that $z_{[n]} \notin [z^*_{[n]}]$ and $\|z_{[n]} - z^*_{[n]}\|_0 = r$ is roughly upper bounded by $\binom{n}{r} (Q-1)^r$, this upper bound being reached for instance when π^* is such that $\pi^*_{q,l} \neq \pi^*_{q',l'}$ for every $(q,l) \neq (q',l')$. Then, a

straightforward union bound leads to

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

by use of $\binom{n}{r} \leq n^r$, which is tight enough for our purpose.

B.3. Upper bounding $\mathbb{P} \left[\frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} > \frac{t}{n^{r+1}(Q-1)^r} \mid \mathbf{Z} = z^* \right]$

Let us first notice that for every vectors $z_{[n]}$ and $z_{[n]}^*$,

$$\begin{aligned} & \log \left(\frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} \right) - \mathbb{E}^{Z=z^*} \left[\log \left(\frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} \right) \right] \\ & = \sum_{i \neq j} \left\{ \left(X_{i,j} - \pi_{z_i^*, z_j^*}^* \right) \log \left[\frac{\pi_{z_i^*, z_j^*}^* (1 - \pi_{z_i^*, z_j^*}^*)}{\pi_{z_i^*, z_j^*}^* (1 - \pi_{z_i, z_j}^*)} \right] \right\}. \end{aligned} \quad (15)$$

Note that for any vector $z_{[n]}$ such that $\pi_{z_i^*, z_j^*}^* \in \{0, 1\}$ and $\pi_{z_i^*, z_j^*}^* \neq \pi_{z_i, z_j}^*$, $\log \left(\frac{P^{X_{[n]}}(z_{[n]})}{P^{X_{[n]}}(z_{[n]}^*)} \right) = -\infty$ and $P^{X_{[n]}}(z_{[n]}) = 0$. Then such vector $z_{[n]}$ can be removed from the sum in Eq. (14).

Second for any vectors $z_{[n]}$ and $z'_{[n]}$, let us further define

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

where z_i and z'_j respectively refer to the i -th (resp. j -th) coordinate of vector $z_{[n]}$ (resp. $z'_{[n]}$). Note that $D(z_{[n]}, z'_{[n]})$ remains unchanged if $z_{[n]}$ and $z'_{[n]}$ are replaced by any representatives of their respective classes.

If $N_r(z_{[n]}) = |D(z_{[n]}, z_{[n]}^*)|$ denotes the number of terms in the sum of

Eq. (15), then

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

Finally, Hoeffding's inequality (Proposition B.1) applied with $a_{ij} = -b_{ij} = \log \left[(1-\zeta)^2 \zeta^{-2} \right]$ (see Lemma B.2 and (A2)), and $L = 2(b_{i,j} - a_{i,j})$ provides for any $s > 0$,

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

B.4. Conclusion

One then apply this last inequality with a particular choice of s :

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

which leads to

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

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

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

and that

$$\exp \left(\frac{-N_r(z_{[n]})s^2}{L^2} \right) \leq \exp \left(\frac{-3N_r(z_{[n]})(c^*)^2}{4L^2} \right) .$$

Using Proposition B.4, it results that

$$N_r(z_{[n]}) \geq \frac{\gamma^2}{2} n \left\| z_{[n]} - z_{[n]}^* \right\|_0 = \frac{\gamma^2}{2} nr ,$$

and

$$\begin{aligned} P^* \left[\sum_{[z_{[n]}] \neq [z_{[n]}^*]} \frac{P^{X_{[n]}}([z_{[n]}])}{P^{X_{[n]}}([z_{[n]}^*])} > t \right] &\leq \sum_{r=1}^n \binom{n}{r} (Q-1)^r \exp \left(-\frac{3(\gamma c^*)^2}{8L^2} nr \right) \\ &= \sum_{r=1}^n \binom{n}{r} [(Q-1)u_n]^r , \end{aligned}$$

where $u_n = \exp \left(-\frac{3(\gamma c^*)^2}{8L^2} n \right)$.

Finally for every $t > 0$, $\log(1+x) \leq x$ for every $x \geq 0$ implies

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

since $nu_n \rightarrow 0$ as $n \rightarrow +\infty$. Further noticing that the upper bound does not depend on $z_{[n]}^*$, the same result holds with P^* replaced by \mathbb{P} .

B.5. 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 $t > 0$,*

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

Lemma B.2 (Values of $a_{i,j}$ and $b_{i,j}$). *Assuming (A2) holds for π^* with $\zeta > 0$, it comes for every $1 \leq i \neq j \leq n$,*

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

Lemma B.3 (Bounding the conditional expectation). *Let us assume (A1), (A2), (A3), and (A4) hold true. Then for every label vectors $z_{[n]}$ and $z_{[n]}^*$ such that $\|z_{[n]} - z_{[n]}^*\|_0 = r$ ($1 \leq r \leq n$), there exist positive constants $c^* = c(\pi^*)$ and $C^* = C(\pi^*)$ such that*

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

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

Proof of Lemma B.3. First,

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

Note that the first sum in the above expression is actually taken over (i, j) such that $\pi_{z_i, z_j}^* \neq \pi_{z_i^*, z_j^*}^*$.

Second, let us introduce

$$C^* := \max \{ 2k(\pi_{q,l}^*, \pi_{q',l'}^*) \} \quad \text{and} \quad c^* := \min \{ k(\pi_{q,l}^*, \pi_{q',l'}^*) / 2 \} ,$$

where *maximum* and *minimum* are taken over $\{((q, l), (q', l')) \mid \pi_{q,l}^* \neq \pi_{q',l'}^*\}$, and $k(x, y) = x \log(x/y) + (1-x) \log[(1-x)/(1-y)]$ for every $x, y \in (0, 1)$. Then for every (i, j) such that $\pi_{z_i, z_j}^* \neq \pi_{z_i^*, z_j^*}^*$,

$$0 < c^* < k\left(\pi_{z_i^*, z_j^*}^*, \pi_{z_i, z_j}^*\right) < C^* .$$

Third, (A3) implies that $\left| \log \frac{\alpha_{z_i}^*}{\alpha_{z_i^*}^*} \right| \leq \log \frac{1-\gamma}{\gamma}$. Therefore (A4) and Proposition B.4 entail

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

The conclusion follows for every n such that

$$\frac{1}{n} \frac{2}{\gamma^2} \log \frac{1-\gamma}{\gamma} < \frac{c^*}{2} < C^* .$$

□

Proposition B.4. *Let $z_{[n]}$ and $z_{[n]}^*$ denote two label vectors. If (A1) and (A4) hold true, then*

$$\left| D(z_{[n]}, z_{[n]}^*) \right| \geq \frac{\gamma^2}{2} n \left\| z_{[n]} - z_{[n]}^* \right\|_0 ,$$

where $D(z_{[n]}, z_{[n]}^*) = \left\{ (i, j) \mid i \neq j, \pi_{z_i, z_j}^* \neq \pi_{z'_i, z'_j}^* \right\}$, $\gamma > 0$ is the constant given by (A4), and $\left\| z_{[n]} - z_{[n]}^* \right\|_0 = \sum_{i=1}^n \mathbb{1}_{(z_i \neq z'_i)}$.

Proof of Proposition B.4.

Since one assumes (A1) holds true, π can be *permutation-invariant* (see Example 2). Then, 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

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

with $\sigma(z_{[n]}) = (\sigma(z_1), \sigma(z_2), \dots, \sigma(z_n))$. Furthermore, neither $\left| D(z_{[n]}, z_{[n]}^*) \right|$ nor $\left\| z_{[n]} - z_{[n]}^* \right\|_0$ will change if the same permutation is applied to the coordinates of vectors $z_{[n]}$ and $z_{[n]}^*$. Then, computing $\left| D(z_{[n]}, z_{[n]}^*) \right|$ can be made by reordering $z_{[n]}$ and $z_{[n]}^*$.

Assumption (A4) 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

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

Then it comes

$$z_{[n]} = (\sigma_1(1), \sigma_1(2), \dots, \sigma_1(Q), \sigma_2(1), \sigma_2(2), \dots, \sigma_2(Q), \dots, \sigma_{n_\gamma}(1), \dots, \sigma_{n_\gamma}(Q)), \quad (16)$$

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

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,

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

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

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

Therefore, lower bounding $\left| D(z_{[n]}, z_{[n]}^*) \right|$ 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 $\|z_{[n]} - z_{[n]}^*\|_0 = r$. 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 exist k, k' such that $B(k, k') + B(k', k) = 0$. (The $B(k)$ s will be lower bounded by 0.)

Then for every such k, k' , σ_k and $\sigma_{k'}$ are permutations. Indeed such k, k' lead to $\pi_{q,l} = \pi_{\sigma_k(q), \sigma_{k'}(l)} = \pi_{\sigma_{k'}(q), \sigma_k(l)}$, for every $q, l \in \{1, \dots, Q\}$. 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 (A1). 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 $\{1, \dots, Q\}$.

Furthermore for any such k, k' , $\sigma_k = \sigma_{k'} = \sigma$ and $\pi^{\sigma_k} = \pi$, where σ denotes a permutation of $\{1, \dots, Q\}$. Indeed if one assumes $\sigma_k \neq \sigma_{k'}$, 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 $z_{[n]}$ decreases by 2 and is no longer minimal as earlier assumed. This yields $\sigma_k = \sigma_{k'}$ and thus $\pi^{\sigma_k} = \pi$. Note that the existence of such a unique permutation σ results from the fact that 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 (16), which contradicts our assumption. 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 r of differences satisfies

$$r \leq m \times Q \quad \Leftrightarrow \quad \frac{m}{r} \geq \frac{1}{Q} . \quad (17)$$

The conclusion results from

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

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 (16). However with the same reasoning as before, 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

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

by use of (17) and $\gamma \leq 1/Q$ (see Assumption (A4)).

□

Appendix C: Proof of Proposition 3.5

Proof of Proposition 3.5. Let us first recall that

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

Then,

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

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

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

where $g = \{g_{i,j}\}_{1 \leq i \neq j \leq n}$. Note that on the parameter set \mathcal{P} defined by (7), $|\xi_{ij} g_{ij}| < +\infty$ a.s. for every $1 \leq i \neq j \leq n$.

The expected control will result from the use of Talagrand's inequality (Theorem C.1). For every $z_{[n]}$ and $\epsilon > 0$, let us introduce the set

$$\mathcal{P}(z_{[n]}) = \{ \pi \mid (z_{[n]}, \pi) \in \mathcal{P} \} \quad ,$$

and define the event

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

where Γ and Λ are constants respectively defined in Lemmas C.2 and C.3, and $\{x_n\}_n$ is a sequence of positive real numbers to be chosen later. Then Theorem C.1 implies for any $z_{[n]}$

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

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

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 expected result. \square

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

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

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 $x > 0$,

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

Proof. A proof can be found in [Massart \(2007\)](#) (p.170, Eq. (5.50)). \square

Lemma C.2. *With the same notation as Theorem C.1, Assumption (A2) 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 (A2)) 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

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

\square

Lemma C.3. *With the same notation as Proposition 3.5, for every $z_{[n]}$ such that $(z_{[n]}, \pi) \in \mathcal{P}$, there exists a constant $\Lambda = \Lambda(\zeta) > 0$ (Assumption (A2)) such that*

$$\mathbb{E} \left[\sup_{\mathcal{P}(z_{[n]})} \rho_n \left| \sum_{i \neq j} \left(X_{i,j} - \pi_{z_i^*, z_j^*}^* \right) g_{ij} \right| \middle| Z = z^* \right] \leq \Lambda [n(n-1)]^{-1/2} .$$

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

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

where the $X'_{i,j}$ s are independent random variables with the same distribution as the $X_{i,j}$ s. A symmetrization argument based on Rademacher variables $\{\epsilon_{i,j}\}_{1 \leq i \neq j \leq n}$ leads to

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

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

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

□

Appendix D: Theorem 3.6

D.1. Proof of Theorem 3.6

D.1.1. Notation

For any metric space (Θ, d) and any real-valued function $f : \Theta \rightarrow \mathbb{R}$, let us define $\|\cdot\|_{\Theta}$ by

$$\|f\|_{\Theta} := \sup_{\theta \in \Theta} |f(\theta)| .$$

Let also α^* and π^* be the true values of α and π in SBM (see Section 2.1), \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

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

Note that \bar{A}_{π} is not necessarily unique for every π . However our analysis only requires unicity of \bar{A}_{π^*} and $\bar{A}_{\pi^*} = I_Q$, which is proved in the following reasoning. Furthermore $\mathbb{M}(\pi^*) = \sum_{q,l} \alpha_q^* \alpha_l^* \mathbb{H}_{q,l}^*$, where $\mathbb{H}_{q,l}^* = \pi_{q,l}^* \log \pi_{q,l}^* + (1 - \pi_{q,l}^*) \log(1 - \pi_{q,l}^*)$.

D.1.2. Proof

First let us prove \bar{A}_{π^*} is unique and $\bar{A}_{\pi^*} = I_Q$. Let us assume $\bar{A}_{\pi^*} \neq I_Q$. By definition of \bar{A}_{π^*} , it results

$$\begin{aligned} 0 &\leq \mathbb{M}(\pi^*, \bar{A}_{\pi^*}) - \mathbb{M}(\pi^*, I_Q) \\ &= \sum_{q,l} \alpha_q^* \alpha_l^* \sum_{q'l'} \bar{a}_{q,q'}(\pi^*) \bar{a}_{l,l'}(\pi^*) [\pi_{q,l}^* \log \frac{\pi_{q'l'}^*}{\pi_{q,l}^*} + (1 - \pi_{q,l}^*) \log \frac{1 - \pi_{q'l'}^*}{1 - \pi_{q,l}^*}] \\ &= - \sum_{q,l} \alpha_q^* \alpha_l^* \sum_{q'l'} \bar{a}_{q,q'}(\pi^*) \bar{a}_{l,l'}(\pi^*) K(\pi_{q,l}^*, \pi_{q'l'}^*) \leq 0 . \end{aligned}$$

Therefore for every (q, q', l, l') , $\bar{a}_{q,q'}(\pi^*)\bar{a}_{l,l'}(\pi^*)K(\pi_{q,l}^*, \pi_{q',l'}^*) = 0$ by **(A3)**.

Since $\sum_{l'} \bar{a}_{l,l'}(\pi^*) = 1$ implies for every $1 \leq l \leq Q$, there exists $1 \leq l' \leq Q$ such that $\bar{a}_{l,l'}(\pi^*) > 0$, there exists $f : \{1, \dots, Q\} \rightarrow \{1, \dots, Q\}$ such that $\pi_{q,l}^* = \pi_{f(q),f(l)}^*$. Then

- f is a permutation of $\{1, \dots, Q\}$ is excluded since we are working up to label switching,
- otherwise there exist two indices q_1 and q_2 ($q_1 \neq q_2$) such that rows q_1 and q_2 of π^* are equal and so do the corresponding columns, which is excluded by **(A1)**,

which proves the unicity and that $\bar{A}_{\pi^*} = I_Q$.

Second, 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 coefficients of \bar{A}_π . Without further indication, $\bar{a}_{q,l}$ depends on the matrix π . Then,

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

Since $\{\pi \mid d(\pi, \pi^*) \geq \eta, \text{ (A1), (A2)}\}$ is a compact set, there exists $\pi^0 \neq \pi^*$ satisfying **(A1)** and **(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 by **(A3)** that for every (q, l, q', l') , $\bar{a}_{q,q'} \bar{a}_{l,l'} K(\pi_{q,l}^*, \pi_{q',l'}^0) = 0$. It implies that there exists $f : \{1, \dots, Q\} \rightarrow \{1, \dots, Q\}$ such that $\pi_{q,l}^* = \pi_{f(q),f(l)}^0$. The same reasoning as for the unicity of \bar{A}_{π^*} leads to

- if f is a permutation of $\{1, \dots, Q\}$: a contradiction arises since $\pi^0 \neq \pi^*$ up to label switching,
- otherwise there exist two indices q_1 and q_2 ($q_1 \neq q_2$) such that rows q_1 and q_2 of π are equal and so do the corresponding columns, which is excluded by **(A1)**.

Third, let us prove that $\sup_{\alpha, \pi} |M_n(\alpha, \pi) - \mathbb{M}(\pi)| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0$. Set

$$|M_n(\alpha, \pi) - \mathbb{M}(\pi)| \leq |M_n(\alpha, \pi) - \phi_n(\pi, \hat{z}_{[n]})| \quad (18)$$

$$+ |\phi_n(\pi, \hat{z}_{[n]}) - \Phi_n(\pi, \tilde{z}_{[n]})| \quad (19)$$

$$+ |\Phi_n(\pi, \tilde{z}_{[n]}) - \mathbb{M}(\pi)| . \quad (20)$$

These three terms are successively controlled in the following.

Upper bound of (18): 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 (19): Let us first introduce several quantities. Set $\Delta(\pi, \widehat{z}_{[n]}, \widetilde{z}_{[n]}) = |\phi_n(\pi, \widehat{z}_{[n]}) - \Phi_n(\pi, \widetilde{z}_{[n]})|$, $\Delta^+(\pi, \widehat{z}_{[n]}, \widetilde{z}_{[n]}) = \phi_n(\pi, \widehat{z}_{[n]}) - \Phi_n(\pi, \widetilde{z}_{[n]})$, and $\Delta^-(\pi, \widehat{z}_{[n]}, \widetilde{z}_{[n]}) = -\Delta^+(\pi, \widehat{z}_{[n]}, \widetilde{z}_{[n]})$. Then, it comes

$$\begin{aligned} P^* \left[\sup_{\pi} \Delta(\pi, \widehat{z}_{[n]}, \widetilde{z}_{[n]}) > \eta \right] &\leq P^* \left[\sup_{\pi} \left\{ \Delta^-(\pi, \widehat{z}_{[n]}, \widetilde{z}_{[n]}) \mathbb{1}_{\Delta^-(\pi, \widehat{z}_{[n]}, \widetilde{z}_{[n]}) > 0} \right\} > \eta \right] \\ &\quad + P^* \left[\sup_{\pi} \left\{ \Delta^+(\pi, \widehat{z}_{[n]}, \widetilde{z}_{[n]}) \mathbb{1}_{\Delta^+(\pi, \widehat{z}_{[n]}, \widetilde{z}_{[n]}) \geq 0} \right\} > \eta \right]. \end{aligned}$$

1. If $\Delta^-(\pi, \widehat{z}_{[n]}, \widetilde{z}_{[n]}) > 0$,

$$\begin{aligned} |\phi_n(\pi, \widehat{z}_{[n]}) - \Phi_n(\pi, \widetilde{z}_{[n]})| &= \Phi_n(\pi, \widetilde{z}_{[n]}) - \phi_n(\pi, \widehat{z}_{[n]}) \\ &\leq \Phi_n(\pi, \widetilde{z}_{[n]}) - \phi_n(\pi, \widetilde{z}_{[n]}), \end{aligned}$$

since $\phi_n(\pi, \widetilde{z}_{[n]}) \leq \phi_n(\pi, \widehat{z}_{[n]})$. Then, Proposition 3.5 leads to

$$\begin{aligned} \sup_{\pi} \left\{ \Delta^-(\pi, \widehat{z}_{[n]}, \widetilde{z}_{[n]}) \mathbb{1}_{\Delta^-(\pi, \widehat{z}_{[n]}, \widetilde{z}_{[n]}) > 0} \right\} &\leq \sup_{(z_{[n]}, \pi) \in \mathcal{P}} \Delta(\pi, z_{[n]}, z_{[n]}) \\ &\xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0. \end{aligned}$$

2. Otherwise $\Delta^+(\pi, \widehat{z}_{[n]}, \widetilde{z}_{[n]}) \geq 0$,

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

Distinguishing between settings where $(\widehat{z}_{[n]}, \pi) \in \mathcal{P}$ or not, it results

$$\begin{aligned} &P^* \left[\sup_{\pi} \left\{ \Delta^+(\pi, \widehat{z}_{[n]}, \widetilde{z}_{[n]}) \mathbb{1}_{\Delta^+(\pi, \widehat{z}_{[n]}, \widetilde{z}_{[n]}) \geq 0} \right\} > \eta \right] \\ &\leq P^* \left[\sup_{\pi, (\widehat{z}_{[n]}, \pi) \in \mathcal{P}} \left\{ \Delta^+(\pi, \widehat{z}_{[n]}, \widetilde{z}_{[n]}) \mathbb{1}_{\Delta^+(\pi, \widehat{z}_{[n]}, \widetilde{z}_{[n]}) \geq 0} \right\} > \eta \right] \\ &\quad + P^* \left[\sup_{\pi, (\widehat{z}_{[n]}, \pi) \notin \mathcal{P}} \left\{ \Delta^+(\pi, \widehat{z}_{[n]}, \widetilde{z}_{[n]}) \mathbb{1}_{\Delta^+(\pi, \widehat{z}_{[n]}, \widetilde{z}_{[n]}) \geq 0} \right\} > \eta \right]. \end{aligned}$$

If $(\widehat{z}_{[n]}, \pi) \in \mathcal{P}$:

$\phi_n(\pi, \widehat{z}_{[n]}) > -\infty$ and

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

by definition of $\tilde{z}_{[n]}$. According to Proposition 3.5, one gets

$$\begin{aligned} & \sup_{\pi, (\hat{z}_{[n]}, \pi) \in \mathcal{P}} \left\{ \Delta^+(\pi, \hat{z}_{[n]}, \tilde{z}_{[n]}) \mathbb{1}_{\Delta^+(\pi, \hat{z}_{[n]}, \tilde{z}_{[n]}) \geq 0} \right\} \\ & \leq \sup_{(z_{[n]}, \pi) \in \mathcal{P}} \Delta(\pi, z_{[n]}, z_{[n]}) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0 . \end{aligned}$$

Otherwise $(\hat{z}_{[n]}, \pi) \notin \mathcal{P}$:

$$\begin{aligned} & P^* \left[\sup_{\pi, (\hat{z}_{[n]}, \pi) \notin \mathcal{P}} \left\{ \Delta^+(\pi, \hat{z}_{[n]}, \tilde{z}_{[n]}) \mathbb{1}_{\Delta^+(\pi, \hat{z}_{[n]}, \tilde{z}_{[n]}) \geq 0} \right\} > \eta \right] \\ & \leq P^* \left[\exists \pi, \Phi_n(\pi, \hat{z}_{[n]}) = -\infty, \Delta^+(\pi, \hat{z}_{[n]}, \tilde{z}_{[n]}) > \eta \right] . \end{aligned}$$

Set a sequence $\{\epsilon_n\}_{n \in \mathbb{N}^*}$ such that $\epsilon_n \rightarrow 0$ and $n\epsilon_n \rightarrow +\infty$ as $n \rightarrow +\infty$. Then,

$$\begin{aligned} & P^* \left[\sup_{\pi, (\hat{z}_{[n]}, \pi) \notin \mathcal{P}} \left\{ \Delta^+(\pi, \hat{z}_{[n]}, \tilde{z}_{[n]}) \mathbb{1}_{\Delta^+(\pi, \hat{z}_{[n]}, \tilde{z}_{[n]}) \geq 0} \right\} > \eta \right] \\ & = P^* \left[\exists \pi, \Phi_n(\pi, \hat{z}_{[n]}) = -\infty, \Delta^+(\pi, \hat{z}_{[n]}, \tilde{z}_{[n]}) > \eta, N(\hat{z}_{[n]}, \pi) \leq \epsilon_n n(n-1) \right] \quad (21) \\ & + P^* \left[\exists \pi, \Phi_n(\pi, \hat{z}_{[n]}) = -\infty, \Delta^+(\pi, \hat{z}_{[n]}, \tilde{z}_{[n]}) > \eta, N(\hat{z}_{[n]}, \pi) > \epsilon_n n(n-1) \right] , \quad (22) \end{aligned}$$

where $N(\hat{z}_{[n]}, \pi) = \left| \left\{ (i, j) \mid i \neq j, \pi_{\hat{z}_i, \hat{z}_j} \in \{0, 1\} \text{ and } \pi_{\hat{z}_i, \hat{z}_j} \neq \pi_{z_i^*, z_j^*}^* \right\} \right|$.

The first term (21) in the right-hand side is dealt with by Proposition D.1:

$$\begin{aligned} & P^* \left[\exists \pi, \Phi_n(\pi, \hat{z}_{[n]}) = -\infty, \Delta^+(\pi, \hat{z}_{[n]}, \tilde{z}_{[n]}) > \eta, N(\hat{z}_{[n]}, \pi) \leq \epsilon_n n(n-1) \right] \\ & \leq P^* \left[\exists \pi, \Phi_n(\pi, \hat{z}_{[n]}) = -\infty, 2a_n + \Delta(\pi, \hat{z}_{[n]}^P, \hat{z}_{[n]}^P) > \eta, N(\hat{z}_{[n]}, \pi) \leq \epsilon_n n(n-1) \right] \\ & \leq P^* \left[2a_n + \sup_{(z_{[n]}, \pi) \in \mathcal{P}} \Delta(\pi, z_{[n]}, z_{[n]}) > \eta \right] \xrightarrow[n \rightarrow +\infty]{} 0 , \end{aligned}$$

following the proof of Proposition 3.5.

The second term (22) is upper bounded by noticing that

$$\begin{aligned} & \left\{ \phi_n(\pi, \hat{z}_{[n]}) > -\infty \right\} \cap \left\{ \Phi_n(\pi, \hat{z}_{[n]}) = -\infty \right\} \\ & = \left\{ \sum_{(i,j) \in \mathcal{M}_0} X_{i,j} = 0 \right\} \cap \left\{ \sum_{(i,j) \in \mathcal{M}_1} (1 - X_{i,j}) = 0 \right\} , \end{aligned}$$

where $\mathcal{M}_0 = \left\{ (i, j) \mid i \neq j, \pi_{\hat{z}_i, \hat{z}_j} = 0 \text{ and } \pi_{z_i^*, z_j^*}^* > 0 \right\}$ and $\mathcal{M}_1 = \left\{ (i, j) \mid i \neq j, \pi_{\hat{z}_i, \hat{z}_j} = 1 \text{ and } \pi_{z_i^*, z_j^*}^* < 1 \right\}$.

Thus,

$$\begin{aligned} & P^* \left[\exists \pi, \Phi_n(\pi, \widehat{z}_{[n]}) = -\infty, \Delta^+(\pi, \widehat{z}_{[n]}, \widetilde{z}_{[n]}) > \eta, N(\widehat{z}_{[n]}, \pi) > \epsilon_n n(n-1) \right] \\ & \leq P^* \left[\sum_{k=1}^{\epsilon_n n(n-1)} Y_k = 0 \right] = (1 - \xi)^{\epsilon_n n(n-1)} \xrightarrow{n \rightarrow +\infty} 0, \end{aligned}$$

where $\{Y_k\}_{1 \leq k \leq \epsilon_n n(n-1)}$ denote *i.i.d.* Bernoulli variables with parameter $\xi = \min_{(q,l), \pi_{q,l}^* \notin \{0,1\}} \left[\pi_{q,l}^* \wedge 1 - \pi_{q,l}^* \right]$, and $a \wedge b = \min(a, b)$.

Finally since no upper bound does depend on $z_{[n]}^*$, every convergence in probability with respect to P^* can be turned into a convergence with respect to \mathbb{P} .

Upper bound of (20): $\Phi_n(\pi, z_{[n]})$ can be expressed as:

$$\begin{aligned} & \Phi_n(\pi, z_{[n]}) \\ & = \sum_{qlq'l'} \frac{N_{qq'}(z_{[n]}) N_{ll'}(z_{[n]})}{n(n-1)} \left[\pi_{q,l}^* \log \pi_{q'l'} + (1 - \pi_{q,l}^*) \log(1 - \pi_{q'l'}) \right], \quad (23) \end{aligned}$$

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 (23) shows that $\Phi_n(\pi, z_{[n]})$ only depends on $z_{[n]}$ through the matrix \widetilde{A}_π . Therefore, one uses the notation $\Phi_n(\pi, A(z_{[n]}))$ 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. $\Phi_n(\pi, \widetilde{A}_\pi) \geq \mathbb{M}(\pi)$
 $\Rightarrow 0 \leq \Phi_n(\pi, \widetilde{A}_\pi) - \mathbb{M}(\pi) \leq \Phi_n(\pi, \widetilde{A}_\pi) - \mathbb{M}(\pi, \widetilde{A}_\pi)$,
2. $\Phi_n(\pi, \bar{A}_\pi) \leq \mathbb{M}(\pi)$
 $\Rightarrow 0 \leq \mathbb{M}(\pi) - \Phi_n(\pi, \bar{A}_\pi) \leq \mathbb{M}(\pi, \bar{A}_\pi) - \Phi_n(\pi, \bar{A}_\pi)$.

Then,

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

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

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

Since any $\pi_{q'l'} \in \{0, 1\}$ such that $\pi_{q,l}^* \neq \pi_{q'l'}$ is excluded, (A2) 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 **(A2)**.

Then, since $0 \leq a_{q,l} \leq 1$ for every (q,l) , the strong law of large numbers applied to each N_q^* entails that $\sup_{\pi} \left\{ |\Phi_n(\pi, \widehat{z}_{[n]}(\pi)) - \mathbb{M}(\pi)| \right\} \xrightarrow[n \rightarrow +\infty]{} 0$ $\mathbb{P} - a.s.$.

D.2. Proof of Proposition D.1

Proposition D.1 (Existence of a copy of $\widehat{z}_{[n]}$ in \mathcal{P}). *Let π be defined as in Section 2.1 and satisfying **(A2)**. Let us further assume that there exists a sequence $\{\epsilon_n\}_{n \in \mathbb{N}^*}$ such that $\epsilon_n \rightarrow 0$ and $n\epsilon_n \rightarrow +\infty$ as $n \rightarrow +\infty$, and*

$$\left| \left\{ (i, j) \mid i \neq j, \pi_{\widehat{z}_i, \widehat{z}_j} \in \{0, 1\} \text{ and } \pi_{\widehat{z}_i, \widehat{z}_j} \neq \pi_{z_i^*, z_j^*}^* \right\} \right| \leq \epsilon_n n(n-1), \quad a.s. .$$

Then, there exist $z_{[n]}^P \in \mathcal{P}$ and a real sequence $\{a_n\}_{\mathbb{N}^*}$ such that

$$0 \leq \phi_n(\pi, \widehat{z}_{[n]}) - \phi_n(\pi, z_{[n]}^P) \leq a_n, \text{ and } \left| \Phi_n(\pi, \widehat{z}_{[n]}) - \Phi_n(\pi, z_{[n]}^P) \right| \leq a_n, \quad a.s. ,$$

where $a_n \rightarrow 0$ as $n \rightarrow +\infty$, and a_n does neither depend on $z_{[n]}^*$ nor on π .

Proof of Proposition D.1. First let us introduce

$$L = \left\{ (q_1, q_2) \in \{1, \dots, Q\}^2 \mid N_{q_1, q_2} > n\sqrt{\epsilon_n} \right\} ,$$

where

$$N_{q_1, q_2} = |\{1 \leq i \leq n \mid \widehat{z}_i = q_1, z_i^* = q_2\}| .$$

For every $1 \leq i \leq n$, we define z_i^P in the following way:

1. $z_i^P = \widehat{z}_i$, if $(\widehat{z}_i, z_i^*) \in L$,
2. $z_i^P = c(z_i^*)$, otherwise,

where $1 \leq c(z_i^*) \leq Q$ is obtained by applying Lemma D.2 with $q_2 = z_i^*$.

Then it results that $(z_i^P, z_i^*) \in L$ for every $1 \leq i \leq n$.

Let us now introduce

$$\mathcal{N} = \left\{ (q, q', l, l') \in \{1, \dots, Q\}^4 \mid \pi_{q,l} \in \{0, 1\} \text{ and } \pi_{q',l'}^* \neq \pi_{q,l} \right\} .$$

Then for every couple (i, j) , $(z_i^P, z_i^*, z_j^P, z_j^*) \notin \mathcal{N}$ since $(z_i^P, z_i^*) \in L$ and $(z_j^P, z_j^*) \in L$ thanks to Lemma D.3. As a consequence, it comes $z_{[n]}^P \in \mathcal{P}$ since

$$\left\{ (i, j) \mid i \neq j, \pi_{z_i^P, z_j^P} \in \{0, 1\} \text{ and } \pi_{z_i^P, z_j^P} \neq \pi_{z_i^*, z_j^*}^* \right\} = \emptyset .$$

Finally, the conclusion results from **(A2)** by noticing that the number of changes between $\widehat{z}_{[n]}$ and $z_{[n]}^P$ is at most $Q^2 n \sqrt{\epsilon_n}$. □

Lemma D.2. Set $L = \left\{ (q_1, q_2) \in \{1, \dots, Q\}^2 \mid N_{q_1, q_2} > n\sqrt{\epsilon_n} \right\}$, where

$$N_{q_1, q_2} = |\{1 \leq i \leq n \mid \widehat{z}_i = q_1, z_i^* = q_2\}| .$$

With the notation and assumptions of Proposition D.1, if (A4) holds true then

$$\forall 1 \leq q_2 \leq Q, \quad \exists 1 \leq q_1 \leq Q, \quad (q_1, q_2) \in L .$$

Proof of Lemma D.2. Otherwise, there exists q_2 such that for every $1 \leq q_1 \leq Q$, $(q_1, q_2) \notin L$. Then,

$$|\{1 \leq i \leq n \mid z_i^* = q_2\}| = \sum_{q_1=1}^Q N_{q_1, q_2} \leq Qn\sqrt{\epsilon_n} ,$$

which contradicts (A4). \square

Lemma D.3. With the same notation and assumptions as Lemma D.2, let us introduce

$$\mathcal{N} = \left\{ (q, q', l, l') \in \{1, \dots, Q\}^4 \mid \pi_{q, l} \in \{0, 1\} \text{ and } \pi_{q', l'}^* \neq \pi_{q, l} \right\} .$$

Then,

$$(q, q', l, l') \in \mathcal{N} \quad \Rightarrow \quad (q, q') \notin L \text{ or } (l, l') \notin L .$$

Proof of Lemma D.3. If $(q, q') \in L$ and $(l, l') \in L$, then $N_{q, q'} N_{l, l'} > n^2 \epsilon_n$, which contradicts that

$$\left| \left\{ (i, j) \mid i \neq j, \pi_{\widehat{z}_i, \widehat{z}_j} \in \{0, 1\} \text{ and } \pi_{\widehat{z}_i, \widehat{z}_j} \neq \pi_{z_i^*, z_j^*}^* \right\} \right| \leq \epsilon_n n(n-1) .$$

\square

Appendix E: Proof of Proposition 3.8

E.1. Proof of Proposition 3.8

Preliminaries First in the same line as the proof of Theorem 3.1, the main quantity to deal with is

$$\begin{aligned} & \log \frac{\widehat{P}^{X_{[n]}}(z_{[n]})}{\widehat{P}^{X_{[n]}}(z_{[n]}^*)} \\ &= \sum_{i \neq j} \left\{ X_{i,j} \log \left(\frac{\widehat{\pi}_{z_i, z_j}}{\widehat{\pi}_{z_i^*, z_j^*}} \right) + (1 - X_{i,j}) \log \left(\frac{1 - \widehat{\pi}_{z_i, z_j}}{1 - \widehat{\pi}_{z_i^*, z_j^*}} \right) \right\} + \sum_i \log \frac{\widehat{\alpha}_{z_i}}{\widehat{\alpha}_{z_i^*}}, \end{aligned} \quad (24)$$

where $\widehat{P}^{X_{[n]}}(z_{[n]}) = \mathbb{P}_{\widehat{\alpha}, \widehat{\pi}}(Z_{[n]} = z_{[n]} | X_{[n]})$ is the same quantity as $P^{X_{[n]}}(z_{[n]}) = \mathbb{P}_{\alpha^*, \pi^*}(Z_{[n]} = z_{[n]} | X_{[n]})$ where (α^*, π^*) has been replaced by $(\widehat{\alpha}, \widehat{\pi})$, and $z_{[n]}$ and $z_{[n]}^*$ denote label vectors such that $\|z_{[n]} - z_{[n]}^*\|_0 = r$, with $1 \leq r \leq n$.

Second, let us assume that

$$\|\widehat{\pi} - \pi^*\|_\infty \leq \min \left[\zeta, \min_{(q,l), \pi_{q,l}^* \notin \{0,1\}} \left\{ \frac{\pi_{q,l}^*(1 - \pi_{q,l}^*)}{2} \right\} \right], \quad (25)$$

where ζ is given by (A2), which is fulfilled on the event

$$\Omega_n = \{\|\widehat{\pi} - \pi^*\|_\infty \leq v_n\} \quad (26)$$

for large enough values of n since $v_n = o(\sqrt{\log n}/n)$. Note that by assumption, $\mathbb{P}[\Omega_n^c] \xrightarrow[n \rightarrow \infty]{} 0$. It is also important to notice that the definition of $\widehat{\pi}$ implies that every $\widehat{\pi}_{q,l} \in \{0, 1\} \cup [\zeta, 1 - \zeta]$ (see (A2)), which leads on Ω_n to

$$\forall (q, l), \quad \pi_{q,l}^* \in \{0, 1\} \Rightarrow \widehat{\pi}_{q,l} = \pi_{q,l}^*. \quad (27)$$

Finally for a given vector $z_{[n]}$, let us introduce the following sets of couples (i, j) :

$$D^* = D^*(z_{[n]}) := \left\{ (i, j) \mid i \neq j, \pi_{z_i, z_j}^* \neq \pi_{z_i^*, z_j^*}^* \right\}, \quad (28)$$

$$\widehat{D} = \widehat{D}(z_{[n]}) := \left\{ (i, j) \mid i \neq j, \widehat{\pi}_{z_i, z_j} \neq \widehat{\pi}_{z_i^*, z_j^*} \right\}. \quad (29)$$

Proof First, the log-ratio (24) can be decomposed into the following terms

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

Second using the definition of D^* and \widehat{D} given by Eq. (28) and Eq. (29), and that $\bar{D}^* \cap \widehat{\bar{D}} \subset \{(i, j) \mid i \neq j, h(X_{i,j}; \widehat{\pi}, \pi^*, z_i, z_j, z_i^*, z_j^*) = 0\}$ where \bar{A} denotes the complement of any set A and

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

it results

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

Finally from the following equalities

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

and

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

the last sum can be further split into

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

This leads to

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

Note that (27) implies for every $1 \leq i \neq j \leq n$,

$$\pi_{z_i, z_j}^* \in \{0, 1\} \Rightarrow \log \left[1 + \frac{(\widehat{\pi}_{z_i, z_j} - \pi_{z_i, z_j}^*)(X_{i,j} - \pi_{z_i, z_j}^*)}{\pi_{z_i, z_j}^*(1 - \pi_{z_i, z_j}^*)} \right] = 0 .$$

In the sequel, the strategy consists in providing successive upper bounds for 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

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

Then for every $t \in \mathbb{R}$,

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

Upper bound of $T_{1,2}$: The same proof as that of Lemma B.3 shows there exists a constant $K(\pi^*) = K^* > 0$ such that

$$T_{1,2} (|D^*|)^{-1} \leq \max_{(q,l) \neq (q',l'), \pi_{q,l}^* \notin \{0,1\}} -k(\pi_{q,l}^*, \pi_{q',l'}^*) = -K^* < 0 ,$$

for large enough values of n , where $k(\pi_{q,l}^*, \pi_{q',l'}^*) = \pi_{q,l}^* \log\left(\frac{\pi_{q,l}^*}{\pi_{q',l'}^*}\right) + (1 - \pi_{q,l}^*) \log\left(\frac{1 - \pi_{q,l}^*}{1 - \pi_{q',l'}^*}\right)$ and $|D^*|$ denotes the cardinality of D^* . Thus,

$$P^* [T_1 > t] \leq P^* [T_{1,1} - |D^*| K^* > t] .$$

Upper bound of $T_{1,1}$:

$$\begin{aligned} & P^* [T_1 > t] \\ & \leq P^* \left[\sum_{(i,j) \in D^*} (X_{i,j} - \pi_{z_i^*, z_j^*}^*) \log \left(\frac{\pi_{z_i^*, z_j^*}^*}{\pi_{z_i^*, z_j^*}^*} \frac{1 - \pi_{z_i^*, z_j^*}^*}{1 - \pi_{z_i^*, z_j^*}^*} \right) > t + |D^*| K^* \right] . \end{aligned}$$

Hoeffding's inequality associated with (A2) provides a constant $C_\zeta > 0$ such that for every $t \in \mathbb{R}$

$$\begin{aligned} & P^* \left[\sum_{(i,j) \in D^*} (X_{i,j} - \pi_{z_i^*, z_j^*}^*) \log \left(\frac{\pi_{z_i^*, z_j^*}^*}{\pi_{z_i^*, z_j^*}^*} \frac{1 - \pi_{z_i^*, z_j^*}^*}{1 - \pi_{z_i^*, z_j^*}^*} \right) > t + |D^*| K^* \right] \\ & \leq \exp \left[- \frac{|D^*|^2 (K^*)^2 + 2t |D^*| K^*}{|D^*| C_\zeta} \right] = \exp \left[-2t \frac{K^*}{C_\zeta} \right] \cdot \exp \left[-|D^*| \frac{(K^*)^2}{C_\zeta} \right] . \end{aligned}$$

Upper bounding T_2

With $t > 0$ on the event $\{T_2 > t\}$, $\log(1+x) \leq x$ for every $x > -1$ leads to

$$0 < t < T_2 \leq \sum_{(i,j) \in \widehat{D} \cup D^*} \frac{(\widehat{\pi}_{z_i, z_j} - \pi_{z_i^*, z_j^*}^*)(X_{i,j} - \pi_{z_i^*, z_j^*}^*)}{\pi_{z_i^*, z_j^*}^* (1 - \pi_{z_i^*, z_j^*}^*)} .$$

Then with $N_{q',l'}^{q,l} = \sum_{(i,j) \in \widehat{D} \cup D^*} \mathbf{1}_{(z_i^* = q', z_j^* = l')} \mathbf{1}_{(z_i = q, z_j = l)}$, it comes

$$\begin{aligned} T_2 & \leq \sum_{q,l} \left| \frac{(\widehat{\pi}_{q,l} - \pi_{q,l}^*)}{\pi_{q,l}^* (1 - \pi_{q,l}^*)} \sum_{\substack{(i,j) \in \widehat{D} \cup D^* \\ z_i = q, z_j = l}} (X_{i,j} - \pi_{q,l}^*) \right| \\ & \leq \sum_{q,l} \left| \frac{(\widehat{\pi}_{q,l} - \pi_{q,l}^*)}{\pi_{q,l}^* (1 - \pi_{q,l}^*)} \right| \left| \sum_{q',l'} \sum_{\substack{(i,j) \in \widehat{D} \cup D^* \\ (z_i^*, z_j^*) = (q', l') \\ (z_i, z_j) = (q, l)}} (X_{i,j} - \pi_{q',l'}^*) \right| \\ & \quad + \sum_{q,l} \left| \frac{(\widehat{\pi}_{q,l} - \pi_{q,l}^*)}{\pi_{q,l}^* (1 - \pi_{q,l}^*)} \right| \left| \sum_{q',l'} N_{q',l'}^{q,l} (\pi_{q',l'}^* - \pi_{q,l}^*) \right| . \end{aligned}$$

Introducing the event Ω_n defined by (26) and using (A2), one gets for every $t > 0$

$$\begin{aligned} & P^*[\Omega_n \cap \{T_2 > t\}] \\ & \leq P^* \left[\zeta^2 \sum_{q,l} \left| \sum_{q',l'} \sum_{\substack{(i,j) \in \bar{D} \cup D^* \\ (z_i^*, z_j^*) = (q', l') \\ (z_i, z_j) = (q, l)}} (X_{i,j} - \pi_{q',l'}^*) \right| > t/(2v_n) \right] \\ & \quad + P^* \left[\zeta^2 \sum_{q,l} \left| \sum_{q',l'} N_{q',l'}^{q,l} (\pi_{q',l'}^* - \pi_{q,l}^*) \right| > t/(2v_n) \right]. \end{aligned}$$

For the first term, Hoeffding's inequality requires summing over a deterministic set of indices, which leads to

$$\begin{aligned} & P^* \left[\zeta^2 \sum_{q,l} \left| \sum_{q',l'} \sum_{\substack{(i,j) \in \bar{D} \cup D^* \\ (z_i^*, z_j^*) = (q', l') \\ (z_i, z_j) = (q, l)}} (X_{i,j} - \pi_{q',l'}^*) \right| > t/(2v_n) \right] \\ & = \sum_k \sum_{D, |D|=k} P^* \left[\zeta^2 \sum_{q,l} \left| \sum_{q',l'} \sum_{\substack{(i,j) \in D \cup D^* \\ (z_i^*, z_j^*) = (q', l') \\ (z_i, z_j) = (q, l)}} (X_{i,j} - \pi_{q',l'}^*) \right| > t/(2v_n) \right]. \end{aligned}$$

where the sum over k is computed for $\lceil \gamma/2nr \rceil \leq k \leq 2nr$ by Proposition B.4 and Lemma E.1.

For each set D such that $|D| = k$, a union bound and Hoeffding's inequality provide

$$\begin{aligned} & P^* \left[\zeta^2 \sum_{q,l} \left| \sum_{q',l'} \sum_{\substack{(i,j) \in D \cup D^* \\ (z_i^*, z_j^*) = (q', l') \\ (z_i, z_j) = (q, l)}} (X_{i,j} - \pi_{q',l'}^*) \right| > t/(2v_n) \right] \\ & \leq Q^2 \max_{q,l} P^* \left[\left| \sum_{q',l'} \sum_{\substack{(i,j) \in D \cup D^* \\ (z_i^*, z_j^*) = (q', l') \\ (z_i, z_j) = (q, l)}} (X_{i,j} - \pi_{q',l'}^*) \right| > t/[2v_n(\zeta Q)^2] \right] \\ & \leq Q^2 \exp \left[-\frac{2}{[2v_n(\zeta Q)^2]^2} \frac{t^2}{k + |D^*|} \right] = Q^2 \exp \left[-\frac{1}{2(\zeta Q)^4} \frac{t^2}{v_n^2(k + |D^*|)} \right]. \end{aligned}$$

Then,

$$\begin{aligned}
 & P^* \left[\zeta^2 \sum_{q,l} \left| \sum_{\substack{q',l' \\ \begin{matrix} (i,j) \in \widehat{D} \cup D^* \\ (z_i^*, z_j^*) = (q', l') \\ (z_i, z_j) = (q, l) \end{matrix}}} (X_{i,j} - \pi_{q',l'}^*) \right| > t/(2v_n) \right] \\
 & \leq Q^2 \sum_{k=\lceil \gamma/2 nr \rceil}^{2nr} (2nr)^k \exp \left[-\frac{1}{2(\zeta Q)^4} \frac{t^2}{v_n^2(k + |D^*|)} \right].
 \end{aligned}$$

For the second term, Lemma E.1 provides

$$\begin{aligned}
 & P^* \left[\zeta^2 \sum_{q,l} \left| \sum_{q',l'} N_{q',l'}^{q,l} (\pi_{q',l'}^* - \pi_{q,l}^*) \right| > t/(2v_n) \right] \\
 & \leq Q^2 \max_{q,l} P^* \left[\left| \sum_{q',l'} N_{q',l'}^{q,l} (\pi_{q',l'}^* - \pi_{q,l}^*) \right| > t/ [2v_n(\zeta Q)^2] \right] \\
 & \leq Q^2 P^* [4nr > t/ [2v_n(\zeta Q)^2]].
 \end{aligned}$$

Upper bounding T3

Let us first notice

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

Then,

$$\begin{aligned}
 & \sum_{(i,j) \in \widehat{D} \cup D^*} \log \left[1 + \frac{(\widehat{\pi}_{z_i^*, z_j^*} - \pi_{z_i^*, z_j^*}^*)(X_{i,j} - \pi_{z_i^*, z_j^*}^*)}{\pi_{z_i^*, z_j^*}^*(1 - \pi_{z_i^*, z_j^*}^*)} \right] \\
 & = \sum_{q,l} \sum_{\substack{(i,j) \in \widehat{D} \cup D^* \\ (z_i^*, z_j^*) = (q, l)}} (1 - X_{i,j}) \log \left[1 - \frac{(\widehat{\pi}_{q,l} - \pi_{q,l}^*)}{(1 - \pi_{q,l}^*)} \right] + X_{i,j} \log \left[1 + \frac{(\widehat{\pi}_{q,l} - \pi_{q,l}^*)}{\pi_{q,l}^*} \right].
 \end{aligned}$$

Centering the $X_{i,j}$ s, it comes

$$\begin{aligned}
 & \sum_{(i,j) \in \widehat{D} \cup D^*} \log \left[1 + \frac{(\widehat{\pi}_{z_i^*, z_j^*} - \pi_{z_i^*, z_j^*}^*)(X_{i,j} - \pi_{z_i^*, z_j^*}^*)}{\pi_{z_i^*, z_j^*}^*(1 - \pi_{z_i^*, z_j^*}^*)} \right] \\
 &= \sum_{q,l} \sum_{\substack{(i,j) \in \widehat{D} \cup D^* \\ (z_i^*, z_j^*) = (q,l)}} (\pi_{q,l}^* - X_{i,j}) \log \left[1 - \frac{(\widehat{\pi}_{q,l} - \pi_{q,l}^*)}{(1 - \pi_{q,l}^*)} \right] + (X_{i,j} - \pi_{q,l}^*) \log \left[1 + \frac{(\widehat{\pi}_{q,l} - \pi_{q,l}^*)}{\pi_{q,l}^*} \right] \\
 &+ \sum_{q,l} \sum_{\substack{(i,j) \in \widehat{D} \cup D^* \\ (z_i^*, z_j^*) = (q,l)}} (1 - \pi_{q,l}^*) \log \left[1 - \frac{(\widehat{\pi}_{q,l} - \pi_{q,l}^*)}{(1 - \pi_{q,l}^*)} \right] + \pi_{q,l}^* \log \left[1 + \frac{(\widehat{\pi}_{q,l} - \pi_{q,l}^*)}{\pi_{q,l}^*} \right],
 \end{aligned}$$

which leads to

$$\begin{aligned}
 T_3 &= \sum_{q,l} \left(\log \left[1 + \frac{(\widehat{\pi}_{q,l} - \pi_{q,l}^*)}{\pi_{q,l}^*} \right] - \log \left[1 - \frac{(\widehat{\pi}_{q,l} - \pi_{q,l}^*)}{(1 - \pi_{q,l}^*)} \right] \right) \sum_{\substack{(i,j) \in \widehat{D} \cup D^* \\ (z_i^*, z_j^*) = (q,l)}} (X_{i,j} - \pi_{q,l}^*) \\
 &+ \sum_{q,l} N_{q,l}^* \left[(1 - \pi_{q,l}^*) \log \left[1 - \frac{(\widehat{\pi}_{q,l} - \pi_{q,l}^*)}{(1 - \pi_{q,l}^*)} \right] + \pi_{q,l}^* \log \left[1 + \frac{(\widehat{\pi}_{q,l} - \pi_{q,l}^*)}{\pi_{q,l}^*} \right] \right],
 \end{aligned}$$

where $N_{q,l}^* = \sum_{(i,j) \in \widehat{D} \cup D^*} \mathbb{1}_{(z_i^*=q, z_j^*=l)}$.

Second on the event Ω_n , (25) and $|\log(1+x)| \leq 2|x|$ for every $x \in [-1/2, 1/2]$ entail

$$|T_3| \leq 4v_n \sum_{q,l} \left| \sum_{\substack{(i,j) \in \widehat{D} \cup D^* \\ (z_i^*, z_j^*) = (q,l)}} (X_{i,j} - \pi_{q,l}^*) \right| + 4v_n \sum_{q,l} N_{q,l}^*.$$

Then for every $t > 0$,

$$\begin{aligned}
 P^* [\Omega_n \cap \{|T_3| > t\}] &\leq P^* \left[4v_n \sum_{q,l} \left| \sum_{\substack{(i,j) \in \widehat{D} \cup D^* \\ (z_i^*, z_j^*) = (q,l)}} (X_{i,j} - \pi_{q,l}^*) \right| > t/2 \right] \\
 &+ P^* \left[4v_n \sum_{q,l} N_{q,l}^* > t/2 \right].
 \end{aligned}$$

Similarly to T_2 , partitioning and Hoeffding's inequality lead to

$$\begin{aligned}
 & P^* \left[4v_n \sum_{q,l} \left| \sum_{\substack{(i,j) \in \widehat{D} \cup D^* \\ (z_i^*, z_j^*) = (q,l)}} (X_{i,j} - \pi_{q,l}^*) \right| > t/2 \right] \\
 &\leq Q^2 \sum_{k=\lceil \gamma/2 nr \rceil}^{2nr} (2nr)^k \exp \left[-\frac{2}{8^2 Q^4 v_n^2} \frac{t^2}{(k + |D^*|)} \right],
 \end{aligned}$$

and Lemma E.1 provides

$$P^* \left[4v_n \sum_{q,l} N_{q,l}^* > t/2 \right] \leq P^* [v_n 4nr > t/8] .$$

Then,

$$P^* [\Omega_n \cap \{|T_3| > t\}] \leq Q^2 \sum_{k=\lceil \gamma/2 nr \rceil}^{2nr} (2nr)^k \exp \left[-\frac{2}{8^2 Q^4} \frac{t^2}{v_n^2 (k + |D^*|)} \right] + P^* [v_n 4nr > t/8] .$$

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

At the beginning the following steps are very close to those in the proof of Theorem B.4.

For every $\epsilon > 0$

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

Furthermore,

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

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

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

Upper bounding T_1 comes from Proposition B.4 and results in

$$\begin{aligned}
 P^* [T_1 > -7r \log n] & \leq \exp \left[r \log n \frac{14K^*}{C_\zeta} \right] \cdot \exp \left[-|D^*| \frac{(K^*)^2}{C_\zeta} \right] \\
 & \leq \exp \left[r \log n \frac{14K^*}{C_\zeta} \right] \cdot \exp \left[-nr \frac{\gamma(K^*)^2}{2C_\zeta} \right] .
 \end{aligned}$$

For T_2 , Lemma E.1 provides

$$\begin{aligned}
 P^* [\{|T_2| > r \log n\} \cap \Omega_n] & \leq Q^2 \sum_{k=\lceil \gamma/2 nr \rceil}^{2nr} (2nr)^k \exp \left[-\frac{1}{2(\zeta Q)^4} \frac{(r \log n)^2}{4nr v_n^2} \right] \\
 & \quad + Q^2 P^* [4nr > r \log n / [2v_n(\zeta Q)^2]] \\
 & \leq Q^2 \exp [8nr \log n] \cdot \exp \left[-\frac{1}{2(\zeta Q)^4} \frac{(r \log n)^2}{4nr v_n^2} \right] \\
 & \quad + Q^2 P^* \left[v_n > \frac{\log n}{8n(\zeta Q)^2} \right] .
 \end{aligned}$$

Similarly for T_3 , it results

$$\begin{aligned}
 P^* [\{|T_3| > r \log n\} \cap \Omega_n] & \leq Q^2 \sum_{k=\lceil \gamma/2 nr \rceil}^{2nr} (2nr)^k \exp \left[-\frac{2}{8^2 Q^4} \frac{(r \log n)^2}{4nr v_n^2} \right] \\
 & \quad + P^* [v_n 4nr > (r \log n)/8] \\
 & \leq Q^2 \exp [8nr \log n] \cdot \exp \left[-\frac{2}{8^2 Q^4} \frac{(r \log n)^2}{4nr v_n^2} \right] \\
 & \quad + P^* \left[v_n > \frac{\log n}{32n} \right] .
 \end{aligned}$$

From the previous bounds, one observes that requiring $v_n = o(\log n/n)$ makes

$P^* \left[v_n > \frac{\log n}{32n} \right]$ and $P^* \left[v_n > \frac{\log n}{8n(\zeta Q)^2} \right]$ vanish as n grows, which leads to

$$\begin{aligned} & P^* [\{T_1 + T_2 - T_3 > -5r \log n\} \cap \Omega_n] \\ & \leq Q^2 \exp [8nr \log n] \cdot \exp \left[-\frac{2}{8^2 Q^4} \frac{r(\log n)^2}{4nv_n^2} \right] \\ & + Q^2 \exp [8nr \log n] \cdot \exp \left[-\frac{1}{2(\zeta Q)^4} \frac{r(\log n)^2}{4nv_n^2} \right] \\ & + \exp \left[r \log n \frac{14K^*}{C_\zeta} \right] \cdot \exp \left[-nr \frac{\gamma(K^*)^2}{2C_\zeta} \right] \\ & \leq C_1 \left(\exp \left[8n \log n - C_2 \frac{(\log n)^2}{nv_n^2} \right] \right)^r \end{aligned}$$

for large enough values of n and constants $C_1, C_2 > 0$ only depending on Q, ζ, γ , and K^* but not of z^* .

Following the same line as in the proof of Theorem 3.1, for every $\epsilon > 0$ and large enough values of n , it comes

$$\begin{aligned} & P^* \left[\left\{ \sum_{[z_{[n]}] \neq [z_{[n]}^*]} \frac{\widehat{P}^{X_{[n]}}([z_{[n]}])}{\widehat{P}^{X_{[n]}}([z_{[n]}^*])} > \epsilon \right\} \cap \Omega_n \right] \\ & \leq \sum_{r=1}^n \binom{n}{r} (Q-1)^r C_1 \left(\exp \left[8n \log n - C_2 \frac{(\log n)^2}{nv_n^2} \right] \right)^r \\ & = C_1 \left([1 + (Q-1)u'_n]^n - 1 \right) , \end{aligned}$$

where $u'_n = \exp \left[8n \log n - C_2 \frac{(\log n)^2}{nv_n^2} \right]$. Thus requiring $v_n = o(\sqrt{\log n}/n)$, it comes

$$\begin{aligned} [1 + (Q-1)u'_n]^n & = \exp [n \log (1 + (Q-1)u'_n)] \\ & \leq \exp [(Q-1)nu'_n] \xrightarrow{n \rightarrow +\infty} 1 , \end{aligned}$$

which concludes the proof since no upper bound does depend on z^* .

E.2. Lemma E.1

Lemma E.1. *Let $\pi \in \mathcal{M}_Q(\mathbb{R})$ denote a matrix with coefficients $\pi_{q,l}$ belong to $[0, 1]$, and $z_{[n]}$ and $z_{[n]}^*$ be two label vectors such that $\sum_{i=1}^n \mathbb{1}_{z_i \neq z_i^*} = r$. Then,*

$$\left| \left\{ (i, j) \mid i \neq j, \pi_{z_i, z_j} \neq \pi_{z_i^*, z_j^*} \right\} \right| \leq 2nr .$$

Proof of Lemma E.1. Without loss of generality, one can assume the first r coordinates of $z_{[n]}$ are different from those of $z_{[n]}^*$. Then, any difference between π_{z_i, z_j} and $\pi_{z_i^*, z_j^*}$ can only occur if $(z_i, z_j) \neq (z_i^*, z_j^*)$. It results

$$\begin{aligned} \left| \left\{ (i, j) \mid i \neq j, \pi_{z_i, z_j} \neq \pi_{z_i^*, z_j^*} \right\} \right| &= \left| \left\{ (i, j) \mid i \neq j, \pi_{z_i, z_j} \neq \pi_{z_i^*, z_j^*}, i \leq r \right\} \right| \\ &\quad + \left| \left\{ (i, j) \mid i, \pi_{z_i^*, z_j} \neq \pi_{z_i^*, z_j^*}, i > r, j \leq r \right\} \right| \\ &\leq nr + (n-r)r \\ &\leq 2nr . \end{aligned}$$

□

Lemma E.2. *With the same notation as Proposition 3.8 and the assumptions of Theorem 3.9, the maximum likelihood estimator of α is given by*

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

Proof of Lemma E.2. Let us introduce some notation:

- $(z_{[n]}, \pi) \mapsto f_{X_{[n]}}(z_{[n]}, \pi) = \mathcal{L}_1(X_{[n]}; z_{[n]}, \pi)$,
- $(\alpha, \pi) \mapsto f_{X_{[n]}}(\alpha, \pi) = \mathcal{L}_2(X_{[n]}; \alpha, \pi)$,
- $\alpha \mapsto f_{Z_{[n]}}(\alpha) = \prod_{i=1}^n \alpha_{Z_i}$,
- $(\alpha, \pi) \mapsto f_{X_{[n]}, Z_{[n]}}(\alpha, \pi) = f_{X_{[n]}}(Z_{[n]}, \pi) f_{Z_{[n]}}(\alpha)$ denote the complete likelihood of (α, π) .

We start computing the derivative of $f_{X_{[n]}}(\alpha, \pi) + \lambda(\sum_q \alpha_q - 1)$ with respect to α_k , for $1 \leq k \leq Q$ and $\lambda \in \mathbb{R}$.

$$\frac{\partial \left[f_{X_{[n]}}(\alpha, \pi) + \lambda(\sum_q \alpha_q - 1) \right]}{\partial \alpha_k} = \sum_{z_{[n]}} \frac{N_k(z_{[n]})}{\alpha_k} f_{X_{[n]}}(z_{[n]}, \pi) f_{z_{[n]}}(\alpha) + \lambda ,$$

where $N_k(z_{[n]}) = \sum_{i=1}^n \mathbb{1}_{(z_i=k)}$. Multiplying by α_k and summing over k leads to

$$\lambda = -n f_{X_{[n]}}(\hat{\alpha}, \pi) ,$$

where $\hat{\alpha}$ denotes the optimum location of α (for which the derivative vanishes).

It results for every k

$$\begin{aligned} \hat{\alpha}_k &= \sum_{z_{[n]}} \frac{N_k(z_{[n]})}{n} \frac{f_{X_{[n]}}(z_{[n]}, \pi) f_{Z_{[n]}}(\hat{\alpha})}{f_{X_{[n]}}(\hat{\alpha}, \pi)} \\ &= \sum_{z_{[n]}} \frac{N_k(z_{[n]})}{n} \frac{f_{X_{[n]}, z_{[n]}}(\hat{\alpha}, \pi)}{f_{X_{[n]}}(\hat{\alpha}, \pi)} \\ &= \sum_{z_{[n]}} \frac{N_k(z_{[n]})}{n} \frac{f_{X_{[n]}, z_{[n]}}(\hat{\alpha}, \pi)}{f_{X_{[n]}}(\hat{\alpha}, \pi)} = \sum_{z_{[n]}} \frac{N_k(z_{[n]})}{n} f_{z_{[n]}}^{X_{[n]}}(\hat{\alpha}, \pi) , \end{aligned}$$

where $f_{z_{[n]}}^{X_{[n]}}(\hat{\alpha}, \pi) = \mathbb{P}_{\hat{\alpha}, \pi} [Z_{[n]} = z_{[n]} | X_{[n]}]$ (Section 3.1.2) denotes the *a posteriori* probability of $Z_{[n]} = z_{[n]}$ given $X_{[n]}$ with parameters $(\hat{\alpha}, \pi)$.

Finally, the result comes from

$$\begin{aligned} \hat{\alpha}_k &= \frac{1}{n} \sum_{i=1}^n \sum_{z_{[n]}} \mathbb{1}_{(z_i=k)} f_{z_{[n]}}^{X_{[n]}}(\hat{\alpha}, \pi) \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{z_{[n]}} \mathbb{1}_{(z_i=k)} \mathbb{P}_{\hat{\alpha}, \pi} [Z_{[n]} = z_{[n]} | X_{[n]}] \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{P}_{\hat{\alpha}, \pi} [Z_i = k | X_{[n]}] . \end{aligned}$$

Replacing π by the MLE $\hat{\pi}$ of π^* , the MLE of α^* satisfies for every k

$$\hat{\alpha}_k = \frac{1}{n} \sum_{i=1}^n \mathbb{P}_{\hat{\alpha}, \hat{\pi}} [Z_i = k | X_{[n]}] = \frac{1}{n} \sum_{i=1}^n \hat{P} (Z_i = k | X_{[n]}) .$$

□

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 results from the definition of \mathcal{J} given by Eq. (10).

The second one comes from $\hat{z}_{[n]}(\pi) = \text{Argmax}_{z_{[n]}} \mathcal{L}_1(X_{[n]}; z_{[n]}, \pi)$. Thus for every (α, π) ,

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

□

Lemma F.2. *Lemma F.1 and Assumption (A3) 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 $\mathcal{J}(X_{[n]}; \hat{\tau}_{[n]}, \alpha, \pi) = \mathcal{L}_1(X_{[n]}; \hat{z}_{[n]}, \pi) + \sum_{i=1}^n \log \alpha_{\hat{z}_i}$ (see Lemma F.3), 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 (A3) yields the conclusion. □

Lemma F.3. *With the same notation as Theorem 4.2, let $\widehat{z}_{[n]} = \widehat{z}_{[n]}(\pi) = \text{Argmin}_{z_{[n]}} \mathcal{L}_1(X_{[n]}; z_{[n]}, \pi)$. Then for every (α, π) ,*

$$\mathcal{J}(X_{[n]}; \widehat{z}_{[n]}, \alpha, \pi) = \mathcal{L}_1(X_{[n]}; \widehat{z}_{[n]}, \pi) + \sum_{i=1}^n \log \alpha_{\widehat{z}_i} .$$

Proof of Lemma F.3. First, let us recall Eq. (10)

$$\begin{aligned} \mathcal{J}(X_{[n]}; \tau_{[n]}, \alpha, \pi) &= \mathcal{L}_2(X_{[n]}; \alpha, \pi) - K(D_{\tau_{[n]}}; P^{X_{[n]}}) \\ &= \log [f(X_{[n]}; \alpha, \pi)] - K(D_{\tau_{[n]}}; P^{X_{[n]}}) , \end{aligned}$$

where $(\alpha, \pi) \mapsto f(X_{[n]}; \alpha, \pi)$ denotes the likelihood of (α, π) .

Second, Eq. (9) and simple calculations lead to

$$\begin{aligned} &K(D_{\tau_{[n]}}; P^{X_{[n]}}) \\ &= \sum_{z_{[n]}} D_{\tau_{[n]}}(z_{[n]}) \log D_{\tau_{[n]}}(z_{[n]}) - \sum_{z_{[n]}} D_{\tau_{[n]}}(z_{[n]}) \log \mathbb{P}(Z_{[n]} = z_{[n]} | X_{[n]}) \\ &= \sum_{i,q} \tau_{i,q} \log \tau_{i,q} - \sum_{z_{[n]}} D_{\tau_{[n]}}(z_{[n]}) \log \left(\frac{f(X_{[n]}, z_{[n]}; \alpha, \pi)}{f(X_{[n]}; \alpha, \pi)} \right) , \end{aligned}$$

where $(\alpha, \pi) \mapsto f(X_{[n]}, z_{[n]}; \alpha, \pi)$ denotes the complete-likelihood of (α, π) . Then,

$$\begin{aligned} &K(D_{\tau_{[n]}}; P^{X_{[n]}}) - \sum_{i,q} \tau_{i,q} \log \tau_{i,q} \\ &= - \sum_{z_{[n]}} D_{\tau_{[n]}}(z_{[n]}) \log f(X_{[n]}; z_{[n]}, \pi) \\ &\quad + \sum_{z_{[n]}} D_{\tau_{[n]}}(z_{[n]}) \log f(z_{[n]}; \alpha) + \log f(X_{[n]}; \alpha, \pi) , \end{aligned}$$

where $\alpha \mapsto f(z_{[n]}; \alpha) = \prod_{i=1}^n \left(\sum_{q=1}^Q \alpha_q^{z_i} \right)$. Hence,

$$\begin{aligned} &K(D_{\tau_{[n]}}; P^{X_{[n]}}) - \sum_{i,q} \tau_{i,q} \log \tau_{i,q} \\ &= - \sum_{i \neq j} \sum_{q,l} [X_{i,j} \log \pi_{q,l} + (1 - X_{i,j}) \log(1 - \pi_{q,l})] \tau_{i,q} \tau_{j,l} \\ &\quad + \sum_{i,q} \tau_{i,q} \log \alpha_q + \mathcal{L}_2(X_{[n]}; \alpha, \pi) . \end{aligned}$$

Therefore for every $\tau_{[n]}$,

$$\begin{aligned} &\mathcal{J}(X_{[n]}; \tau_{[n]}, \alpha, \pi) \\ &= \sum_{i \neq j} \sum_{q,l} [X_{i,j} \log \pi_{q,l} + (1 - X_{i,j}) \log(1 - \pi_{q,l})] \tau_{i,q} \tau_{j,l} - \sum_{i,q} \tau_{i,q} (\log \tau_{i,q} - \log \alpha_q) . \end{aligned}$$

With $\tau_{[n]} = \widehat{z}_{[n]}$, it comes $\mathcal{J}(X_{[n]}; \widehat{z}_{[n]}, \alpha, \pi) = \mathcal{L}_1(X_{[n]}; \widehat{z}_{[n]}, \pi) + \sum_{i=1}^n \log \alpha_{\widehat{z}_i}$, which concludes the proof. \square

Lemma F.4.

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

Proof of Lemma F.4.

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

\square

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. (2012). New consistent and asymptotically normal estimators for random graph mixture models. *JRSS. B*, **74**(1), 3–35.
- 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. (2009). A nonparametric view of network models and Newman-Girvan and other modularities. *PNAS*, pages 1–6.
- Bickel, P. J., Chen, A., and Levina, E. (2011). The method of moments and degree distributions for network models. *Ann. Statist.*, **39**(5), 2280–2301.
- 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. (2012). Stochastic blockmodels with growing number of classes. *Biometrika*, **In press**(arXiv:1011.4644v2).
- 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://stat.genopole.cnrs.fr/logiciels/mixnet>.
- 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. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.*, **39**(4), 1878–1915.
- Snijders, T. A. B. and Nowicki, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification*

tion, **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.