Bernoulli **21**(1), 2015, 537–573
DOI: 10.3150/13-BEJ579

Convergence of the groups posterior distribution in latent or stochastic block models

MAHENDRA MARIADASSOU[1] and CATHERINE MATIAS[2]

[1]*INRA, UR1077 Unité Mathématique, Informatique et Génome, 78350 Jouy-en-Josas, France. E-mail: mahendra.mariadassou@jouy.inra.fr*
[2]*Laboratoire Statistique et Génome, Université d'Évry Val d'Essonne, UMR CNRS 8071 – USC INRA, 23 bvd de France, 91 037 Évry, France. E-mail: catherine.matias@genopole.cnrs.fr*

We propose a unified framework for studying both latent and stochastic block models, which are used to cluster simultaneously rows and columns of a data matrix. In this new framework, we study the behaviour of the groups posterior distribution, given the data. We characterize whether it is possible to asymptotically recover the actual groups on the rows and columns of the matrix, relying on a consistent estimate of the parameter. In other words, we establish sufficient conditions for the groups posterior distribution to converge (as the size of the data increases) to a Dirac mass located at the actual (random) groups configuration. In particular, we highlight some cases where the model assumes symmetries in the matrix of connection probabilities that prevents recovering the original groups. We also discuss the validity of these results when the proportion of non-null entries in the data matrix converges to zero.

Keywords: biclustering; block clustering; block modelling; co-clustering; latent block model; posterior distribution; stochastic block model

1. Introduction

Cluster analysis is an important tool in a variety of scientific areas including pattern recognition, microarrays analysis, document classification and more generally data mining. In these contexts, one is interested in data recorded in a table or matrix, where for instance rows index objects and columns index features or variables. While the majority of clustering procedures aim at clustering either the objects or the variables, we focus here on procedures which consider the two sets simultaneously and organize the data into homogeneous blocks. More precisely, we are interested in probabilistic models called latent block models (LBMs), where both rows and columns are partitioned into latent groups [14].

Stochastic block models (SBMs, [18]) may be viewed as a particular case of LBMs where data consists in a random graph which is encoded in its adjacency matrix. An adjacency matrix is a square matrix where rows and columns are indexed by the same set of objects and an entry in the matrix describes the relation between two objects. For instance, binary random graphs are described by a binary matrix where entry (i, j) equals 1 if and only if there is an edge between nodes (i, j) in the graph. Similarly, weighted random graphs are encoded in square matrices where the entries describe the edges weights (the weight being 0 in case of no edge between the two nodes). In this context the partitions on rows and columns of the square matrix are further constrained to be identical.

To our knowledge and despite their similarities, LBMs and SBMs have never been explored from the same point of view. We aim at presenting a unified framework for studying both LBMs and SBMs. We are more precisely interested in the behaviour of the groups posterior distribution, given the data. Our goal is to characterize whether it is possible to asymptotically recover the actual groups on the rows and columns of the matrix, relying on a consistent estimate of the parameter. In other words, we establish sufficient conditions for the groups posterior distribution to converge (as the size of the data increases) to a Dirac mass located at the actual (random) groups configuration. In particular, we highlight some cases where the model assumes symmetries in the matrix of connection probabilities that prevents recovering the original groups (see Theorem 1 and following corollaries). Note that the asymptotic framework is particularly suited in this context as the datasets are often huge.

One of the first occurrences of LBMs appears in the pioneering work [17] under the name *three partitions*. LBMs were later developed as an intuitive extension of the finite mixture model, to allow for simultaneous clustering of objects and features. Many different names are used in the literature for such procedures, among which we mention block clustering, block modelling, biclustering, co-clustering and two-mode clustering. All of these procedures differ through the type of clusters they consider. LBMs induce a specific clustering on the data matrix, namely we *partition* the rows and columns of the data matrix and the data clusters are restricted to *cartesian products* of a row cluster and a column cluster. Frequentist parameter estimation procedures for LBMs have been proposed in [14,15] for binary data and in [16] for Poisson random variables. A Bayesian version of the model has been introduced in [10] for random variables belonging to the set [0, 1], combined with a Markov chain Monte Carlo (MCMC) procedure to estimate the model parameters. Moreover, model selection in a Bayesian setting is performed at the same time as parameter estimation in [28], that considers two different types of models: a Bernoulli LBM for binary data and a Gaussian one for continuous observations. All of these parameter estimation procedures also provide a clustering of the data, based on the groups posterior distribution computed at the estimated parameter value. In the following, *a posteriori estimation* of the groups refers to maximum a posteriori (MAP) procedure on the groups posterior distribution computed at some estimated parameter value. To our knowledge, there is no result in the literature about the quality of such clustering procedures nor about convergence of the groups posterior distribution in LBMs.

SBMs were (re)-discovered many different times in the literature, and introduced at first in social sciences to study relational data (see, for instance, [9,12,18,27]). In this context, the data consists in a random graph over a set of nodes, or equivalently in a square matrix (the adjacency matrix) whose entries characterize the relation between two nodes. The nodes are partitioned into latent groups so that the clustering of the rows and columns of the matrix is now constrained to be identical. Various parameter estimation procedures have been proposed in this context, from Bayesian strategies [23,27], to variational approximations of expectation maximization (EM) algorithm [9,21,24] or variational Bayes approaches [20], online procedures [29,30] and direct methods [4,7]. Note that most of these works are concerned with binary data and only some of the most recent of them deal with weighted random graphs [4,21].

In each of these procedures, a clustering of the graph nodes is performed according to the groups posterior distribution (computed at the estimated parameter value). The behaviour of this posterior distribution for binary SBMs is studied in [6]. These authors establish two different results. The first one (Theorem 3.1 in [6]) states that at the true parameter value, the groups posterior distribution converges to a Dirac mass at the actual value of groups configuration (controlling also the corresponding rate of convergence). This result is valid only at the true parameter value, while the above mentioned procedures rely on the groups posterior distribution at an estimated value of the parameter instead of the true one. Note also that this result establishes a convergence under the *conditional* distribution of the data, given the actual configuration on the groups. However, as this convergence is uniform with respect to the actual configuration, the result also holds under the unconditional distribution of the observations. The second result they obtain on the convergence of the groups posterior distribution (Proposition 3.8 in [6]) is valid at an estimated parameter value, provided this estimator converges at rate at least n^{-1} to the true value, where n is the number of nodes in the graph (number of rows and columns in the square data matrix). Note that this latter assumption is not harmless as it is not established that such an estimator exists, except in a particular setting [4]; see also [13] for empirical results. There are thus many differences between our result (Theorem 1 and following corollaries) and theirs: we provide a result for any parameter value in the neighborhood of the true value, we work with non-necessarily binary data and our work encompasses both SBMs and LBMs. We however mention that the main goal of these authors is different from ours and consists in establishing the consistency of maximum likelihood and variational estimators in SBMs.

We stress here that our result relies on the existence of *consistent* parameter estimates (without any constraint on the convergence rate). Such consistency results have been established for instance, in [4] in the specific context of affiliation (namely only two connections types are considered: intra-group and inter-group connections) for binary or weighted SBMs; in [6] for binary (possibly directed) SBMs and concerning the connectivity parameters (a result on the groups proportions requires an additional assumption whose validity is not yet established) and also by [5] for binary SBMs where the parameter estimates are derived from groups estimators that rely on specific consistent modularities. As already stressed in the above paragraph, the behaviour of the groups posterior distribution is not fully resolved in those contexts. Moreover, to our knowledge, consistency results have not been (theoretically) established in LBMs but we believe that our common framework enables to obtain results in LBMs similar as those obtained in SBMs.

Let us now discuss the articles [5,8] and [25] on the performance of clustering procedures for random graphs, as well as the very recent works [11,26] on the performance of co-clustering procedures. Those articles, which are of a different nature from ours, establish that under some conditions, the fraction of misclassified nodes (resulting from different algorithmic procedures) converges to zero as the number of nodes increases. These results, with the exception of [11], apply to binary graphs only, while we shall deal both with binary and weighted graphs; as well as real-valued array data. Moreover, they establish results on procedures that estimate parameters while clustering the data, while we are rather interested in MAP based procedures relying on any consistent parameter estimate. In [5], Bickel and Chen show that groups estimates based on the use of different modularities are consistent in the sense that with probability tending to one, these recover the original groups. Rohe and Yu [26] are concerned with a framework in which nodes of a graph belong to two groups: a *receiver* group and a *sender* group. This is a refinement of standard SBM, which assumes equal sender and receiver groups, and is motivated by the study of directed graphs. The results of [26] are very similar to those of [25] that apply on symmetric binary graphs: they propose a classification algorithm based on spectral clustering that achieves vanishing classification error rate. Flynn and Perry [11] share our framework with a few exceptions: they replace the profile likelihood (used for instance, in [5]) with a rate function allowing for model misspecification. Besides, they model sparsity with a scaling parameter acting directly on the mean interaction value instead of inflating the number of zero-valued interactions with a Bernoulli variable as we do (see Section 4.3). They essentially extend the results of [8] to weighted graphs. The focus on the mean interaction value allows for model misspecification but prevents the detection of groups that differ mostly in interaction variance. Indeed, the simulation study from [11] considers only groups varying in their mean interaction value, while we can detect groups varying through their variance for instance. We also mention that [8,25] and [26] are concerned with an asymptotic setting where the number of groups is allowed to grow with network size and the average network degree grows at least nearly linearly [25,26] or poly-logarithmically [8] in this size. In Section 5 of the present work, we explore the validity of our results in a similar framework, by assuming that the numbers of groups remain fixed while the connections probabilities between groups converge to zero. Finally and most importantly, note that *all these works but* [5] propose convergence results in a setup of *independent* random observations (and Bernoulli distributed, except for [11] that consider more general distributions), viewing the latent groups as parameters instead of random variables. On the contrary in our context, the observed random variables are non-independent. This makes a tremendous difference in the validity of the statements.

We also want to outline that many different generalization allowing for overlapping groups exist, both for LBMs and SBMs. We refer the interested reader to the works [10] for LBMs and [1,19] in the case of SBMs, as well as the references therein. However in this work, we restrict our attention to non-overlapping groups.

This work is organized as follows. Section 2 describes LBMs and SBMs and introduces some important concepts such as equivalent group configurations. Section 3 establishes general and sufficient conditions for the groups posterior probability to converge (with large probability) to a (mixture of) Dirac mass, located at (the set of configurations equivalent to) the actual random configuration. In particular, we discuss the cases where it is likely that groups estimation relying on maximum posterior probabilities might not converge. Section 4 illustrates our main result, providing a large number of examples where the above mentioned conditions are satisfied. Finally, in Section 5 we explore the validity of our results when the connections probabilities between groups converge to zero. This corresponds to datasets with an asymptotically decreasing density of non-null entries. Some technical proofs are postponed to the Appendix.

2. Model and notation

2.1. Model and assumptions

We observe a matrix $\mathbf{X}_{n,m} := \{X_{ij}\}_{1\leq i\leq n, 1\leq j\leq m}$ of random variables in some space set $\mathcal{X}$, whose distribution is specified through latent groups on the rows and columns of the matrix.

Let $Q \geq 1$ and $L \geq 1$ denote the number of latent groups respectively on the rows and columns of the matrix. Consider the probability distributions $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_Q)$ on $\mathcal{Q} = \{1, \dots, Q\}$ and

$\boldsymbol{\beta} = (\beta_1, \dots, \beta_L)$ on $\mathcal{L} = \{1, \dots, L\}$, such that

$$\forall q \in \mathcal{Q}, \forall l \in \mathcal{L}, \qquad \alpha_q, \beta_l > 0 \quad \text{and} \quad \sum_{q=1}^{Q} \alpha_q = 1, \qquad \sum_{l=1}^{L} \beta_l = 1.$$

Let $\mathbf{Z}_n := Z_1, \dots, Z_n$ be independent and identically distributed (i.i.d.) random variables, with distribution $\boldsymbol{\alpha}$ on $\mathcal{Q}$ and $\mathbf{W}_m := W_1, \dots, W_m$ i.i.d. random variables with distribution $\boldsymbol{\beta}$ on $\mathcal{L}$. Two different cases will be considered in this work:

Latent block model (LBM). In this case, the random variables $\{Z_i\}_{1 \le i \le n}$ and $\{W_j\}_{1 \le j \le m}$ are independent. We let $\mathcal{I} = \{1, \dots, n\} \times \{1, \dots, m\}$ and $\boldsymbol{\mu} = \boldsymbol{\alpha}^{\otimes n} \otimes \boldsymbol{\beta}^{\otimes m}$ the distribution of $(\mathbf{Z}_n, \mathbf{W}_m) := (Z_1, \dots, Z_n, W_1, \dots, W_m)$ and set $U_{ij} = (Z_i, W_j)$ for (i, j) in $\mathcal{I}$. The random vector $(\mathbf{Z}_n, \mathbf{W}_m)$ takes values in the set $\mathcal{U} := \mathcal{Q}^n \times \mathcal{L}^m$ whereas the $\{U_{ij} := (Z_i, W_j)\}_{(i,j) \in \mathcal{I}}$ are non-independent random variables taking values in the set $(\mathcal{Q} \times \mathcal{L})^{nm}$.

Stochastic block model (SBM). In this case, we have $n = m$, $\mathcal{Q} = \mathcal{L}$, $Z_i = W_i$ for all $1 \le i \le n$ and $\boldsymbol{\alpha} = \boldsymbol{\beta}$. We let $\mathcal{I} = \{1, \dots, n\}^2$, $\boldsymbol{\mu} = \boldsymbol{\alpha}^{\otimes n}$ the distribution of $\mathbf{Z}_n$ and set $U_{ij} = (Z_i, Z_j)$ for $(i, j) \in \mathcal{I}$. The random variables $\{U_{ij} := (Z_i, Z_j)\}_{(i,j) \in \mathcal{I}}$ are not independent and take values in the set

$$\mathcal{U} = \big\{ \big\{ (q_i, q_j) \big\}_{(i,j) \in \mathcal{I}}; \forall i \in \{1, \dots, n\}, q_i \in \mathcal{Q} \big\}.$$

This case corresponds to the observation of a random graph whose adjacency matrix is given by $\{X_{ij}\}_{1 \le i,j \le n}$. As particular cases, we may also consider graphs with no self-loops in which case $\mathcal{I} = \{1, \dots, n\}^2 \setminus \{(i, i); 1 \le i \le n\}$. We may also consider undirected random graphs, possibly with no self-loops, by imposing symmetric adjacency matrices $X_{ij} = X_{ji}$. In this latter case, $\mathcal{I} = \{1 \le i < j \le n\}$.

In the following, we refer to each of these two cases by indicating the symbols (LBM) and (SBM). Whenever possible, we give general formulas valid for the two cases, and which could be simplified appropriately in SBM. We introduce a matrix of connectivity parameters $\boldsymbol{\pi} = (\pi_{ql})_{(q,l) \in \mathcal{Q} \times \mathcal{L}}$ belonging to some set of matrices $\Pi_{\mathcal{Q}\mathcal{L}}$ whose coordinates π_{ql} belong to some set Π (note that $\Pi_{\mathcal{Q}\mathcal{L}}$ may be different from the product set Π^{QL}). Now, conditional on the latent variables $\{U_{ij} = (Z_i, W_j)\}_{(i,j) \in \mathcal{I}}$, the observed random variables $\{X_{ij}\}_{(i,j) \in \mathcal{I}}$ are assumed to be independent, with a parametric distribution on each entry depending on the corresponding rows and columns groups. More precisely, conditional on $Z_i = q$ and $W_j = l$, the random variable X_{ij} follows a distribution parameterized by π_{ql}. We let $f(\cdot; \pi_{ql})$ denote its density with respect to some underlying measure (either the counting or Lebesgue measure).

The model may be summarized as follows:

- $(\mathbf{Z}_n, \mathbf{W}_m)$ latent random variables in $\mathcal{U}$ with distribution given by $\boldsymbol{\mu}$,
- $\mathbf{X}_{n,m} = \{X_{ij}\}_{(i,j) \in \mathcal{I}}$ observations in $\mathcal{X}$,
- $\mathbb{P}(\mathbf{X}_{n,m} | \mathbf{Z}_n, \mathbf{W}_m) = \bigotimes_{(i,j) \in \mathcal{I}} \mathbb{P}(X_{ij} | Z_i, W_j)$, (1)
- $\forall (i, j) \in \mathcal{I}$ and $\forall (q, l) \in \mathcal{Q} \times \mathcal{L}$, we have $X_{ij} | (Z_i, W_j) = (q, l) \sim f(\cdot; \pi_{ql})$.

We consider the following parameter set

$$\Theta = \{\theta = (\boldsymbol{\mu}, \boldsymbol{\pi}); \boldsymbol{\pi} \in \Pi_{\mathcal{QL}} \text{ and } \forall (q,l) \in \mathcal{Q} \times \mathcal{L}, \alpha_q \geq \alpha_{\min} > 0, \beta_l \geq \beta_{\min} > 0\},$$

and define $\alpha_{\max} = \max\{\alpha_q; q \in \mathcal{Q}; \theta \in \Theta\}$ and similarly $\beta_{\max} = \max\{\beta_l; l \in \mathcal{L}; \theta \in \Theta\}$. We let $\mu_{\min} := \alpha_{\min} \wedge \beta_{\min}$ and $\mu_{\max} := \alpha_{\max} \vee \beta_{\max}$. Note that in SBM, $\mu_{\min}$ (resp. $\mu_{\max}$) reduces to $\alpha_{\min}$ (resp. $\alpha_{\max}$). We denote by $\mathbb{P}_\theta$ and $\mathbb{E}_\theta$ the probability distribution and expectation under parameter value θ. In the following, we assume that the observations $\mathbf{X}_{n,m}$ are drawn under the true parameter value $\theta^\star \in \Theta$. We let $\mathbb{P}_\star$ and $\mathbb{E}_\star$ respectively, denote probability and expectation under parameter value $\theta^\star$. We now introduce a necessary condition for the connectivity parameters to be identifiable from $\mathbb{P}_\theta$.

Assumption 1.

(i) *The parameter $\pi \in \Pi$ is identifiable from the distribution $f(\cdot;\pi)$, namely $f(\cdot;\pi) = f(\cdot;\pi') \Rightarrow \pi = \pi'$.*

(ii) *For all $q \neq q' \in \mathcal{Q}$, there exists some $l \in \mathcal{L}$ such that $\pi_{ql} \neq \pi_{q'l}$. Similarly, for all $l \neq l' \in \mathcal{L}$, there exists some $q \in \mathcal{Q}$ such that $\pi_{ql} \neq \pi_{ql'}$.*

Assumption 1 will be in force throughout this work. Note that it is a very natural assumption. In particular, (i) will be satisfied by any reasonable family of distributions and if (ii) is not satisfied, there exist for instance, two row groups $q \neq q'$ with the same behavior. These groups (and thus the corresponding parameters) may then not be distinguished relying on the marginal distribution of $\mathbb{P}_\theta$ on the observation space $\mathcal{X}^{\mathbb{N}}$. Note also that Assumption 1 is in general not sufficient to ensure identifiability of the parameters in LBM or SBM. Identifiability results for SBM have first been given in a particular case in [2] and then later more thoroughly discussed in [3] for undirected, binary or weighted random graphs. See also [6] for the case of directed and binary random graphs.

In the following, for any subset A we denote by either 1_A or $1\{A\}$ the indicator function of event A, by $|A|$ its cardinality and by $\bar{A}$ the complementary subset (in the ambient set).

2.2. Equivalent configurations

First of all, it is important to note that the classical label switching issue that arises in any latent variable model also takes place in LBMs and SBMs. As such, any permutation on the labels of the rows and columns groups will induce the same distribution on the data matrix. To be more specific, we let $\mathfrak{S}_Q$ (resp. $\mathfrak{S}_L$) be the set of permutations of $\mathcal{Q}$ (resp. $\mathcal{L}$). In the following, we define $\mathfrak{S}_{\mathcal{QL}}$ to be either the set $\mathfrak{S}_Q \times \mathfrak{S}_L$ (LBM) or the set $\{(s,s); s \in \mathfrak{S}_Q\}$ (SBM). We consider some $\sigma \in \mathfrak{S}_{\mathcal{QL}}$ and for any parameter value $\theta = (\boldsymbol{\mu}, \boldsymbol{\pi})$ we denote by $\sigma(\theta)$ the parameter induced by permuting the labels of rows and columns groups according to σ. Then label switching corresponds to the fact that

$$\mathbb{P}_\theta = \mathbb{P}_{\sigma(\theta)}. \tag{2}$$

Now in LBM and SBM, there exists an additional phenomenon, that is specific to these models and comes from the fact that the distribution of any random variable depends on two different latent ones. Let us explain this now. In a classical latent variable model where the distribution of individuals belonging to group q is characterised by the parameter π_q, identifiability conditions will require that $\pi_q \neq \pi_l$ for any two different groups $q \neq l$. Now, when considering two groups characterising the distribution of one random variable, it may happen that for instance $\pi_{ql} = \pi_{q'l}$ for two different groups $q \neq q'$. Indeed, the groups q, q' may be differentiated through their connectivity to other groups than the group l (see point (ii) in Assumption 1). As a consequence, if the parameter matrix $\boldsymbol{\pi}$ has some symmetries (which is often the case for model parsimony reasons), it may happen that some row and column groups can be permuted while the connectivity matrix $\boldsymbol{\pi}$ remains unchanged. Note that in this case, the global parameter $\theta = (\boldsymbol{\mu}, \boldsymbol{\pi})$ remains identifiable as soon as the groups proportions (characterised by $\boldsymbol{\mu}$) are different. More precisely, it may happen that for some $\sigma \in \mathfrak{S}_{\mathcal{QL}}$, we have $\boldsymbol{\pi} = \sigma(\boldsymbol{\pi})$ (a case that can never occur for simple latent variables models) and thus

$$\mathbb{P}_{\boldsymbol{\mu},\boldsymbol{\pi}} = \mathbb{P}_{\boldsymbol{\mu},\sigma(\boldsymbol{\pi})}. \tag{3}$$

Note the difference between (2) and (3). In particular, whenever $\boldsymbol{\mu} \neq \sigma(\boldsymbol{\mu})$ we have $\mathbb{P}_{\boldsymbol{\mu},\boldsymbol{\pi}} \neq \mathbb{P}_{\sigma(\boldsymbol{\mu}),\sigma(\boldsymbol{\pi})}$ and we are not facing an instance of label switching.

We now formalize the concept of *equivalent configurations* that will enable us to deal with possible symmetries in the parameter matrices $\boldsymbol{\pi}$. Note that from a practical perspective, these subtleties have little impact (in fact, the same kind of impact as the label switching). But these are necessary for stating our results rigorously.

For any $(s, t) \in \mathfrak{S}_{\mathcal{QL}}$, we let

$$\boldsymbol{\pi}^{s,t} := \left(\pi_{ql}^{s,t}\right)_{(q,l)\in\mathcal{Q}\times\mathcal{L}} := \left(\pi_{s(q)t(l)}\right)_{(q,l)\in\mathcal{Q}\times\mathcal{L}}.$$

Fix a subgroup $\mathfrak{S}$ of $\mathfrak{S}_{\mathcal{QL}}$ and a parameter set $\Pi_{\mathcal{QL}}$. Whenever for any pair of permutations $(s, t) \in \mathfrak{S}$ and any parameter $\boldsymbol{\pi} \in \Pi_{\mathcal{QL}}$ we have $\boldsymbol{\pi}^{s,t} = \boldsymbol{\pi}$, we say that the parameter set $\Pi_{\mathcal{QL}}$ is *invariant under the action of* $\mathfrak{S}$. In the following, we will consider parameter sets that are invariant under some subgroup $\mathfrak{S}$. This includes the case where $\mathfrak{S}$ is reduced to identity. We will moreover exclude from the parameter set $\Pi_{\mathcal{QL}}$ any point $\boldsymbol{\pi}$ admitting specific symmetries, namely such that there exists

$$(s, t) \in \mathfrak{S}_{\mathcal{QL}} \setminus \mathfrak{S} \quad \text{satisfying} \quad \boldsymbol{\pi}^{s,t} = \boldsymbol{\pi}.$$

Note that this corresponds to excluding a subset of null Lebesgue measure from the parameter set $\Pi_{\mathcal{QL}}$.

Assumption 2. *The parameter set $\Pi_{\mathcal{QL}}$ is invariant under the action of some (maximal) subgroup $\mathfrak{S}$ of $\mathfrak{S}_{\mathcal{QL}}$. Moreover, for any pair of permutations $(s, t) \in \mathfrak{S}_{\mathcal{QL}} \setminus \mathfrak{S}$ and any parameter $\boldsymbol{\pi} \in \Pi_{\mathcal{QL}}$, we assume that $\boldsymbol{\pi}^{s,t} \neq \boldsymbol{\pi}$.*

Example 1. In SBM, we consider $\mathfrak{S} = \{(Id, Id)\}$ where Id is the identity and let

$$\Pi_{\mathcal{QL}} = \left\{\boldsymbol{\pi} \in \Pi^{Q^2}; \forall s \in \mathfrak{S}_Q, s \neq Id, \text{ we have } \boldsymbol{\pi}^{s,s} \neq \boldsymbol{\pi}\right\}.$$

Example 2 (Affiliation SBM). In SBM, we consider $\mathfrak{S} = \{(s, s); s \in \mathfrak{S}_Q\}$ and let $\Pi_{\mathcal{QL}} = \{(\lambda - \nu)I_Q + \nu \mathbf{1}_Q^\intercal \mathbf{1}_Q; \lambda, \nu \in (0, 1), \lambda \neq \nu\}$.

In the above notation, I_Q is the identity matrix of size Q and $\mathbf{1}_Q$ is the size-Q vector filled with 1s. Affiliation SBM is a simple two-parameters submodel of SBM commonly used to detect communities with higher intra- than inter-groups connectivities. It imposes as much symmetry on elements of $\Pi_{\mathcal{QL}}$ as allowed by Assumption 1 and constitutes the only model where configuration equivalence (defined below) is confounded with label-switching.

In less constrained models and as soon as $\mathfrak{S}$ is not reduced to identity, each permutation in $\mathfrak{S}$ induces many different *equivalent* configurations. More precisely, for any $(s, t) \in \mathfrak{S}$ and any $\boldsymbol{\pi} \in \Pi_{\mathcal{QL}}$, we have

$$\mathbf{X}_{n,m} | \{\mathbf{Z}_n, \mathbf{W}_m\} \stackrel{d}{=} \mathbf{X}_{n,m} | \{s(\mathbf{Z}_n), t(\mathbf{W}_m)\}, \quad \text{under parameter value } \boldsymbol{\pi},$$

where $=^d$ means equality in distribution.

Remark 1. In SBM with affiliation structure (see Example 2), the whole group of permutations $\{(s, s); s \in \mathfrak{S}_Q\}$ leaves the parameter set $\Pi_{\mathcal{QL}}$ invariant. For more general models, let us denote by $[q, q']$ the transposition of q and q' in some set $\mathcal{Q}$. We consider $(s, t) = ([q, q'], [l, l']) \in \mathfrak{S}_{\mathcal{QL}}$. Then any $\boldsymbol{\pi} \in \Pi_{\mathcal{QL}}$ satisfies

$$\begin{aligned} &\forall i \in \mathcal{Q} \setminus \{q, q'\}, \qquad \pi_{il} = \pi_{il'}, \\ &\forall j \in \mathcal{L} \setminus \{l, l'\}, \qquad \pi_{qj} = \pi_{q'j}, \\ &\pi_{ql} = \pi_{q'l'} \quad \text{and} \quad \pi_{q'l} = \pi_{ql'}. \end{aligned}$$

In particular, for Assumption 1 to be satisfied while $([q, q'], [l, l'])$ belongs to $\mathfrak{S}$ that leaves $\Pi_{\mathcal{QL}}$ invariant, it is necessary that either $\pi_{ql} \neq \pi_{ql'}$ or $\pi_{q'l} \neq \pi_{q'l'}$ (and then both inequalities are satisfied).

Note that the parameter sets $\Pi_{\mathcal{QL}}$ that we consider are then in a one-to-one correspondence with the subgroups $\mathfrak{S}$. Note also that we have $|\mathfrak{S}| \leq Q!L!$ (LBM) or $|\mathfrak{S}| \leq Q!$ (SBM).

We now define equivalent configurations in $\mathcal{U}$.

Definition 1. *Consider a parameter set $\Pi_{\mathcal{QL}}$ invariant under the action of some subgroup $\mathfrak{S}$ of $\mathfrak{S}_{\mathcal{QL}}$ and fix a parameter value $\boldsymbol{\pi} \in \Pi_{\mathcal{QL}}$. Any two groups configurations $(\mathbf{z}_n, \mathbf{w}_m) := (z_1, \ldots, z_n, w_1, \ldots, w_m)$ and $(\mathbf{z}'_n, \mathbf{w}'_m) := (z'_1, \ldots, z'_n, w'_1, \ldots, w'_m)$ in $\mathcal{U}$ are called* equivalent *(a relation denoted by $(\mathbf{z}_n, \mathbf{w}_m) \sim (\mathbf{z}'_n, \mathbf{w}'_m)$) if and only if there exists $(s, t) \in \mathfrak{S}$ such that*

$$\left(s(\mathbf{z}'_n), t(\mathbf{w}'_m)\right) := \left(s(z'_1), \ldots, s(z'_n), t(w'_1), \ldots, t(w'_m)\right) = (\mathbf{z}_n, \mathbf{w}_m).$$

We let $\tilde{\mathcal{U}}$ denote the quotient of $\mathcal{U}$ by this equivalence relation. Note in particular that if $(\mathbf{z}_n, \mathbf{w}_m) \sim (\mathbf{z}'_n, \mathbf{w}'_m)$ then for any $\boldsymbol{\pi} \in \Pi_{\mathcal{QL}}$, we have $(\pi_{z_i w_j})_{(i,j) \in \mathcal{I}} = (\pi_{z'_i w'_j})_{(i,j) \in \mathcal{I}}$.

For any vector $u = (u_1, \dots, u_p) \in \mathbb{R}^p$, we let $\|u\|_0 := \sum_{i=1}^p 1\{u_i \neq 0\}$. The distance between two different configurations $(\mathbf{z}_n, \mathbf{w}_m) \in \tilde{\mathcal{U}}$ and $(\mathbf{z}'_n, \mathbf{w}'_m) \in \tilde{\mathcal{U}}$ is measured via the minimum $\|\cdot\|_0$ distance between any two representatives of these classes. We thus let

$$d\big((\mathbf{z}_n, \mathbf{w}_m), (\mathbf{z}'_n, \mathbf{w}'_m)\big) := \min\big\{ \big\| \mathbf{z}_n - s(\mathbf{z}'_n) \big\|_0 + \big\| \mathbf{w}_m - t(\mathbf{w}'_m) \big\|_0; (s,t) \in \mathfrak{S} \big\}. \tag{4}$$

Note that this distance is well-defined on the space $\tilde{\mathcal{U}}$. Note also that when $\mathfrak{S}$ is reduced to identity, the distance $d(\cdot,\cdot)$ is an ordinary ℓ_0 distance (up to a scale factor 2 in SBM).

2.3. Most likely configurations

Among the set of all (up to equivalence) configurations $\tilde{\mathcal{U}}$, we shall distinguish some which are well-behaved in the following sense. For any groups $q \in \mathcal{Q}$ and $l \in \mathcal{L}$, consider the events

$$A_q = \left\{ \omega \in \Omega; N_q\big(\mathbf{Z}_n(\omega)\big) := \sum_{i=1}^n 1\big\{Z_i(\omega) = q\big\} < n\mu_{\min}/2 \right\},$$

and

$$B_l = \left\{ \omega \in \Omega; N_l\big(\mathbf{W}_m(\omega)\big) := \sum_{j=1}^m 1\big\{W_j(\omega) = l\big\} < m\mu_{\min}/2 \right\}.$$

Since $N_q(\mathbf{Z}_n)$ and $N_l(\mathbf{W}_m)$ are sums of i.i.d. Bernoulli random variables with respective parameters $\alpha_q^\star$ and $\beta_l^\star$, satisfying $\alpha_q^\star \wedge \beta_l^\star \geq \mu_{\min}$, a standard Hoeffding's Inequality gives

$$\mathbb{P}_\star(A_q \cup B_l) \leq \exp\big[-n\big(\alpha_q^\star\big)^2/2\big] + \exp\big[-m\big(\beta_l^\star\big)^2/2\big] \leq 2\exp\big[-(n \wedge m)\mu_{\min}^2/2\big].$$

Taking an union bound, we obtain

$$\mathbb{P}_\star\bigg(\bigcup_{(q,l) \in \mathcal{Q}\times\mathcal{L}} (A_q \cup B_l) \bigg) \leq \begin{cases} 2QL\exp\big[-(n\wedge m)\mu_{\min}^2/2\big] & \text{(LBM)}, \\ 2Q\exp\big[-n\alpha_{\min}^2/2\big] & \text{(SBM)}. \end{cases}$$

Now, consider the event Ω_0 defined by

$$\begin{aligned} \Omega_0 &:= \big\{ \omega \in \Omega; \forall (q,l) \in \mathcal{Q}\times\mathcal{L}, N_q\big(\mathbf{Z}_n(\omega)\big) \geq n\mu_{\min}/2 \text{ and } N_l\big(\mathbf{W}_m(\omega)\big) \geq m\mu_{\min}/2 \big\} \\ &= \bigcap_{(q,l)\in\mathcal{Q}\times\mathcal{L}} (\bar{A}_q \cap \bar{B}_l), \end{aligned} \tag{5}$$

which has $\mathbb{P}_\star$-probability larger than $1 - 2QL\exp[-(n\wedge m)\mu_{\min}^2/2]$ (LBM) or larger than $1 - 2Q\exp[-n\alpha_{\min}^2/2]$ (SBM) and its counterpart $\mathcal{U}^0$ defined by

$$\mathcal{U}^0 = \big\{ (\mathbf{z}_n, \mathbf{w}_m) \in \mathcal{U}; \forall (q,l) \in \mathcal{Q}\times\mathcal{L}, N_q(\mathbf{z}_n) \geq n\mu_{\min}/2 \text{ and } N_l(\mathbf{w}_m) \geq m\mu_{\min}/2 \big\}, \tag{6}$$

where $N_q(\mathbf{z}_n) := \sum_{i=1}^n 1\{z_i = q\}$ and $N_l(\mathbf{w}_m)$ is defined similarly. We extend this notation up to equivalent configurations, by letting $\tilde{\mathcal{U}}^0$ be the set of configurations $(\mathbf{z}_n, \mathbf{w}_m) \in \tilde{\mathcal{U}}$ such that at least one (and then in fact all) representative in the class belongs to $\mathcal{U}^0$. Note that neither $N_q(\mathbf{z}_n)$ nor $N_l(\mathbf{w}_m)$ are properly defined on $\tilde{\mathcal{U}}$, as these quantities may take different values for equivalent configurations. However, as soon as one representative $(\mathbf{z}_n, \mathbf{w}_m)$ belongs to $\mathcal{U}^0$, we both get $N_q(\mathbf{z}'_n) \geq n\mu_{\min}/2$ and $N_l(\mathbf{w}'_m) \geq m\mu_{\min}/2$ for any $(\mathbf{z}'_n, \mathbf{w}'_m) \sim (\mathbf{z}_n, \mathbf{w}_m)$. In the following, some properties will only be valid on the set of configurations $\tilde{\mathcal{U}}^0$.

3. Groups posterior distribution

3.1. The groups posterior distribution

We provide a preliminary lemma on the expression of the groups posterior distribution.

Lemma 1. *For any $n, m \geq 1$ and any $\theta \in \Theta$, the groups posterior distribution writes for any $(\mathbf{z}_n, \mathbf{w}_m) \in \mathcal{U}$,*

$$p^\theta_{n,m}(\mathbf{z}_n, \mathbf{w}_m) := \mathbb{P}_\theta\big((\mathbf{Z}_n, \mathbf{W}_m) = (\mathbf{z}_n, \mathbf{w}_m) | \mathbf{X}_{n,m}\big)$$

$$\propto \begin{cases} \Big(\prod_{(i,j)\in\mathcal{I}} f(X_{ij}; \pi_{z_i w_j})\Big)\Big(\prod_{i=1}^n \alpha_{z_i}\Big)\Big(\prod_{j=1}^m \beta_{w_j}\Big) & (\mathit{LBM}), \\ \Big(\prod_{(i,j)\in\mathcal{I}} f(X_{ij}; \pi_{z_i z_j})\Big)\Big(\prod_{i=1}^n \alpha_{z_i}\Big) & (\mathit{SBM}), \end{cases} \tag{7}$$

where $\propto$ means equality up to a normalizing constant.

The proof of this lemma is straightforward and therefore omitted.

In the following, we will consider the main term in the log ratio $\log p^\theta_{n,m}(\mathbf{z}^\star_n, \mathbf{w}^\star_m) - \log p^\theta_{n,m}(\mathbf{z}_n, \mathbf{w}_m)$ for two different configurations $(\mathbf{z}^\star_n, \mathbf{w}^\star_m), (\mathbf{z}_n, \mathbf{w}_m) \in \mathcal{U}$. More precisely, we introduce

$$\forall \big(\mathbf{z}^\star_n, \mathbf{w}^\star_m\big), (\mathbf{z}_n, \mathbf{w}_m) \in \tilde{\mathcal{U}}, \qquad \delta^\pi\big(\mathbf{z}^\star_n, \mathbf{w}^\star_m, \mathbf{z}_n, \mathbf{w}_m\big) = \sum_{(i,j)\in\mathcal{I}} \log\bigg(\frac{f(X_{ij}; \pi_{z^\star_i w^\star_j})}{f(X_{ij}; \pi_{z_i w_j})}\bigg). \tag{8}$$

Note that this quantity is well-defined on $\tilde{\mathcal{U}} \times \tilde{\mathcal{U}}$. We also consider its expectation, under true parameter value $\theta^\star$ and conditional on the event $(\mathbf{Z}_n, \mathbf{W}_m) = (\mathbf{z}^\star_n, \mathbf{w}^\star_m)$; namely for any $(\mathbf{z}^\star_n, \mathbf{w}^\star_m)$ and $(\mathbf{z}_n, \mathbf{w}_m) \in \tilde{\mathcal{U}}$, we let

$$\Delta^\pi\big(\mathbf{z}^\star_n, \mathbf{w}^\star_m, \mathbf{z}_n, \mathbf{w}_m\big) = \sum_{(i,j)\in\mathcal{I}} \mathbb{E}_\star\bigg(\log\bigg(\frac{f(X_{ij}; \pi_{z^\star_i w^\star_j})}{f(X_{ij}; \pi_{z_i w_j})}\bigg)\Big|(\mathbf{Z}_n, \mathbf{W}_m) = \big(\mathbf{z}^\star_n, \mathbf{w}^\star_m\big)\bigg). \tag{9}$$

Probabilities and expectations conditional on $(\mathbf{Z}_n, \mathbf{W}_m) = (\mathbf{z}^\star_n, \mathbf{w}^\star_m)$ and under parameter value $\theta^\star$ will be denoted by $\mathbb{P}_\star^{\mathbf{z}^\star_n \mathbf{w}^\star_m}$ and $\mathbb{E}_\star^{\mathbf{z}^\star_n \mathbf{w}^\star_m}$, respectively.

3.2. Assumptions on the model

The results of this section are valid as long as the family of distributions $\{f(\cdot;\pi); \pi \in \Pi\}$ satisfies some properties. We thus formulate these as assumptions in this general section, and establish later that these assumptions are satisfied in each particular case to be considered.

The first of these assumptions is a (conditional on the configuration) concentration inequality on the random variable $\delta^{\boldsymbol{\pi}}(\mathbf{Z}_n, \mathbf{W}_m, \mathbf{z}_n, \mathbf{w}_m)$ around its conditional expectation. We only require it to be valid for configurations $(\mathbf{Z}_n, \mathbf{W}_m) = (\mathbf{z}^\star_n, \mathbf{w}^\star_m) \in \tilde{\mathcal{U}}^0$. Note that under conditional probability $\mathbb{P}_\star^{\mathbf{z}^\star_n \mathbf{w}^\star_m}$, the random variables $\{X_{ij}; (i,j) \in \mathcal{I}\}$ are independent.

Assumption 3 (Concentration inequality). *Fix $(\mathbf{z}^\star_n, \mathbf{w}^\star_m) \in \tilde{\mathcal{U}}^0$ and $(\mathbf{z}_n, \mathbf{w}_m) \in \tilde{\mathcal{U}}$ such that $(\mathbf{z}_n, \mathbf{w}_m) \nsim (\mathbf{z}^\star_n, \mathbf{w}^\star_m)$. There exists some positive function $\psi^\star : (0, +\infty) \to (0, +\infty]$ such that for any $\boldsymbol{\pi} \in \Pi_{\mathcal{QL}}$ and any $\varepsilon > 0$, we have*

$$\begin{aligned} &\mathbb{P}_\star^{\mathbf{z}^\star_n \mathbf{w}^\star_m}\Big(\big|\delta^{\boldsymbol{\pi}}\big(\mathbf{z}^\star_n, \mathbf{w}^\star_m, \mathbf{z}_n, \mathbf{w}_m\big) - \mathbb{E}_\star^{\mathbf{z}^\star_n \mathbf{w}^\star_m}\big(\delta^{\boldsymbol{\pi}}\big(\mathbf{z}^\star_n, \mathbf{w}^\star_m, \mathbf{z}_n, \mathbf{w}_m\big)\big)\big| \geq \varepsilon(mr_1 + nr_2)\Big) \\ &\quad \leq 2\exp\big[-\psi^\star(\varepsilon)(mr_1 + nr_2)\big], \end{aligned} \tag{10}$$

where the distance $d((\mathbf{z}^\star_n, \mathbf{w}^\star_m), (\mathbf{z}_n, \mathbf{w}_m))$ defined by (4) is attained for some permutations $(s,t) \in \mathfrak{S}$ and we set $r_1 := \|\mathbf{z}^\star_n - s(\mathbf{z}_n)\|_0$ and $r_2 := \|\mathbf{w}^\star_m - t(\mathbf{w}_m)\|_0$.

Remark 2. Assumption 3 is reasonable and is often obtained by an exponential control of the centered random variable

$$Y_{\pi,\pi'} = \log\left(\frac{f(X;\pi)}{f(X;\pi')}\right) - \mathbb{E}_\pi\left[\log\left(\frac{f(X;\pi)}{f(X;\pi')}\right)\right],$$

uniformly in $\pi, \pi' \in \Pi$, where $\mathbb{E}_\pi$ is the expectation under $f(\cdot,\pi)$. As shown in Section 4.1, as soon as

$$\psi_{\max}(\lambda) := \sup_{\pi,\pi' \in \Pi} \mathbb{E}_\pi\big(\exp(\lambda Y_{\pi,\pi'})\big)$$

is finite for λ in a small open interval $I \subset \mathbb{R}$ around 0, a Cramer–Chernoff bound shows that Inequality (10) is satisfied with

$$\psi^\star(\varepsilon) := \frac{\mu^2_{\min}}{8} \sup_{\lambda \in I}\big(\lambda\varepsilon - \psi_{\max}(\lambda)\big).$$

The second assumption needed is a bound on the Kullback–Leibler divergences for elements of the family $\{f(\cdot;\pi); \pi \in \Pi\}$. We let

$$D\big(\pi \parallel \pi'\big) := \int_{\mathcal{X}} \log\left(\frac{f(x;\pi)}{f(x;\pi')}\right) f(x;\pi)\,\mathrm{d}x. \tag{11}$$

Assumption 4 (Bounds on Kullbak–Leibler divergences). *We assume that*

$$\kappa_{\max} := \max\{D(\pi \parallel \pi'); \pi, \pi' \in \Pi\} < +\infty.$$

Note that $\kappa_{\max} < +\infty$ is automatically satisfied when the distributions in the family $\{f(\cdot;\pi); \pi \in \Pi\}$ form an exponential family with natural parameter π belonging to a compact set Π. In particular, this is not the case for Bernoulli distributions when we authorize some probabilities π to be 0 or 1, as the corresponding natural parameter then takes the values $-\infty$ and $+\infty$. In the following, we thus exclude for the binary case the possibility that classes may be almost never or almost surely connected. We also introduce

$$\kappa_{\min} = \kappa_{\min}(\boldsymbol{\pi}^\star) := \min\{D(\pi^\star_{ql} \parallel \pi^\star_{q'l'}); (q,l), (q',l') \in \mathcal{Q} \times \mathcal{L}, \pi^\star_{ql} \neq \pi^\star_{q'l'}\} > 0, \tag{12}$$

where positivity is a consequence of Assumption 1. The parameter $\kappa_{\min}$ measures how far apart the non-identical entries of $\boldsymbol{\pi}^\star$ are and is the main driver of the convergence rate of the posterior distribution. Note that the Kullback–Leibler divergence captures the differences between the distributions and not only their mean values. As we already mentioned in the Introduction, this is in contrast to results as in [11] and we may for instance recover groups that differ only in their variance.

The last assumption needed is a Lipschitz condition on an integrated version of the function $\pi \mapsto \log f(x;\pi)$.

Assumption 5. *There exists some positive constant L_0 such that for any $\boldsymbol{\pi}, \boldsymbol{\pi}' \in \Pi_{\mathcal{QL}}$ and any $(q,l), (q',l') \in \mathcal{Q} \times \mathcal{L}$, we have*

$$\left| \int_{\mathcal{X}} \log \frac{f(x;\pi_{ql})}{f(x;\pi'_{ql})} f(x;\pi_{q'l'})\,\mathrm{d}x \right| \leq L_0 \|\boldsymbol{\pi} - \boldsymbol{\pi}'\|_\infty.$$

Remark 3. As illustrated in Section 4.2, many exponential families satisfy Assumptions 3 to 5 as long as the *natural parameter* of that family (e.g., $\log(p)$ for Poisson distribution or $\log(p/(1-p))$ for the binomial) is restricted to a compact set. This includes but is not limited to Gaussian (location or scale model), Poisson, binary, binomial and multinomial distributions.

3.3. Convergence of the posterior distribution

We now establish some preliminary results. The first one gives the behavior of the conditional expectation $\Delta^{\boldsymbol{\pi}}$ defined by (9) with respect to the distance between the two configurations $(\mathbf{Z}_n, \mathbf{W}_m)$ and $(\mathbf{z}_n, \mathbf{w}_m)$.

Proposition 1 (Behavior of conditional expectation). *Under Assumptions 1, 2 and 4, the constant $C = 2\kappa_{\max} > 0$ is such that for any parameter value $\boldsymbol{\pi} \in \Pi_{\mathcal{QL}}$ and any configuration $(\mathbf{z}_n, \mathbf{w}_m) \in \tilde{\mathcal{U}}$, we have $\mathbb{P}_\star$-almost surely*

$$\mathbb{E}_\star^{\mathbf{Z}_n\mathbf{W}_m}\big(\delta^{\boldsymbol{\pi}}(\mathbf{Z}_n, \mathbf{W}_m, \mathbf{z}_n, \mathbf{w}_m)\big) \leq \frac{C}{2}(mr_1 + nr_2), \tag{13}$$

where the distance $d((\mathbf{Z}_n, \mathbf{W}_m), (\mathbf{z}_n, \mathbf{w}_m))$ is attained for some $(s,t) \in \mathfrak{S}$ and we set $r_1 := \|\mathbf{Z}_n - s(\mathbf{z}_n)\|_0$ and $r_2 := \|\mathbf{W}_m - t(\mathbf{w}_m)\|_0$.

Furthermore, under additional Assumption 5, the constant $c = \mu_{\min}^2\kappa_{\min}/16$ is such that on the set Ω_0 defined by (5) whose $\mathbb{P}_\star$-probability satisfies

$$\begin{cases} \mathbb{P}_\star(\Omega_0) \geq 1 - 2QL \times \exp\big[-(n \wedge m)\mu_{\min}^2/2\big] & (\textit{LBM}), \\ \mathbb{P}_\star(\Omega_0) \geq 1 - 2Q \times \exp\big[-n\alpha_{\min}^2/2\big] & (\textit{SBM}), \end{cases}$$

for any parameter value $\boldsymbol{\pi} \in \Pi_{\mathcal{QL}}$ and any sequence $(\mathbf{z}_n, \mathbf{w}_m) \in \tilde{\mathcal{U}}$, we have

$$\mathbb{E}_\star^{\mathbf{Z}_n\mathbf{W}_m}\big(\delta^{\boldsymbol{\pi}}(\mathbf{Z}_n, \mathbf{W}_m, \mathbf{z}_n, \mathbf{w}_m)\big) \geq 2\big(c - L_0\big\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\big\|_\infty\big)(mr_1 + nr_2). \tag{14}$$

Proof. Note that

$$\mathbb{E}_\star^{\mathbf{Z}_n\mathbf{W}_m}\big(\delta^{\boldsymbol{\pi}}(\mathbf{Z}_n, \mathbf{W}_m, \mathbf{z}_n, \mathbf{w}_m)\big) = \sum_{(\mathbf{z}_n^\star, \mathbf{w}_m^\star) \in \tilde{\mathcal{U}}} \mathbb{E}_\star^{\mathbf{z}_n^\star\mathbf{w}_m^\star}\big(\delta^{\boldsymbol{\pi}}\big(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m\big)\big) \times 1_{(\mathbf{Z}_n, \mathbf{W}_m) = (\mathbf{z}_n^\star, \mathbf{w}_m^\star)},$$

so that we can work on the set $\{(\mathbf{Z}_n, \mathbf{W}_m) = (\mathbf{z}_n^\star, \mathbf{w}_m^\star)\}$ for a fixed configuration $(\mathbf{z}_n^\star, \mathbf{w}_m^\star) \in \tilde{\mathcal{U}}$. Moreover, we can choose $(\mathbf{z}_n, \mathbf{w}_m) \in \tilde{\mathcal{U}}$ that realizes the distance $d((\mathbf{z}_n^\star, \mathbf{w}_m^\star), (\mathbf{z}_n, \mathbf{w}_m))$, namely such that $d((\mathbf{z}_n^\star, \mathbf{w}_m^\star), (\mathbf{z}_n, \mathbf{w}_m)) = \|\mathbf{z}_n^\star - \mathbf{z}_n\|_0 + \|\mathbf{w}_m^\star - \mathbf{w}_m\|_0 = r_1 + r_2$.

If $(\mathbf{z}_n, \mathbf{w}_m) = (\mathbf{z}_n^\star, \mathbf{w}_m^\star)$, namely $r_1 = r_2 = 0$, then we have $\delta^{\boldsymbol{\pi}}(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m) = 0$ and the lemma is proved. Otherwise, we may have r_1 or r_2 equal to zero but $r_1 + r_2 \geq 1$. Without loss of generality, we can assume that $\mathbf{z}_n^\star$ and $\mathbf{z}_n$ (respectively $\mathbf{w}_m^\star$ and $\mathbf{w}_m$) differ at the first r_1 (resp. r_2) indexes.

First, let us note that

$$\mathbb{E}_\star^{\mathbf{z}_n^\star\mathbf{w}_m^\star}\big(\delta^{\boldsymbol{\pi}}\big(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m\big)\big) = \sum_{(i,j) \in \tilde{\mathcal{I}}} \int_{\mathcal{X}} \log\left(\frac{f(x; \pi_{z_i^\star w_j^\star})}{f(x; \pi_{z_i w_j})}\right) f\big(x; \pi^\star_{z_i^\star w_j^\star}\big)\,\mathrm{d}x, \tag{15}$$

where $\tilde{\mathcal{I}} = \mathcal{I} \setminus \{(i,j); i > r_1 \text{ and } j > r_2\}$. This leads to

$$\mathbb{E}_\star^{\mathbf{z}_n^\star\mathbf{w}_m^\star}\big(\delta^{\boldsymbol{\pi}}\big(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m\big)\big) \leq (mr_1 + nr_2 - r_1r_2)\kappa_{\max} \leq \frac{C}{2}(mr_1 + nr_2),$$

with $C = 2\kappa_{\max}$, which establishes Inequality (13).

To prove Inequality (14), we write the decomposition

$$\begin{aligned} &\sum_{(i,j) \in \tilde{\mathcal{I}}} \int_{\mathcal{X}} \log\left(\frac{f(x; \pi_{z_i^\star w_j^\star})}{f(x; \pi_{z_i w_j})}\right) f\big(x; \pi^\star_{z_i^\star w_j^\star}\big)\,\mathrm{d}x \\ &\quad = \sum_{(i,j) \in \tilde{\mathcal{I}}} \bigg\{-D\big(\pi^\star_{z_i^\star w_j^\star} \,\|\, \pi_{z_i^\star w_j^\star}\big) + D\big(\pi^\star_{z_i^\star w_j^\star} \,\|\, \pi^\star_{z_i w_j}\big) \\ &\qquad\quad + \int_{\mathcal{X}} \log \frac{f(x; \pi^\star_{z_i w_j})}{f(x; \pi_{z_i w_j})} f\big(x; \pi^\star_{z_i^\star w_j^\star}\big)\,\mathrm{d}x\bigg\}. \end{aligned} \tag{16}$$

According to Assumption 5, the third term in the right-hand side of the above equation is lower-bounded by $-L_0\|\boldsymbol{\pi}-\boldsymbol{\pi}^\star\|_\infty(mr_1+nr_2-r_1r_2)$. The first term in this right-hand side is handled similarly as we have

$$\begin{aligned}0 &< \sum_{(i,j)\in\tilde{\mathcal{I}}} D\big(\pi^\star_{z_i^\star w_j^\star} \parallel \pi_{z_i^\star w_j^\star}\big) = \sum_{(i,j)\in\tilde{\mathcal{I}}} \int_{\mathcal{X}} \log \frac{f(x;\pi^\star_{z_i^\star w_j^\star})}{f(x;\pi_{z_i^\star w_j^\star})} f\big(x;\pi^\star_{z_i^\star w_j^\star}\big)\,\mathrm{d}x \\ &\le L_0\big\|\boldsymbol{\pi}-\boldsymbol{\pi}^\star\big\|_\infty(mr_1+nr_2-r_1r_2),\end{aligned}$$

where the second inequality is another application of Assumption 5.

The central term appearing in the right-hand side of decomposition (16) is handled relying on the next lemma, whose proof is postponed to the Appendix. It is a generalization to LBM of Proposition B.5 in [6] that considers SBM only. This lemma bounds from below the number of pairs (i,j) such that

$$\pi^\star_{z_i^\star w_j^\star} \neq \pi^\star_{z_i w_j}$$

and establishes that it is of order mr_1+nr_2. This is possible only for the configurations $(\mathbf{z}_n^\star,\mathbf{w}_m^\star)\in\tilde{\mathcal{U}}^0$ defined by (6). For the rest of the proof, we work on the set Ω_0, meaning that we assume $\{(\mathbf{Z}_n,\mathbf{W}_m)=(\mathbf{z}_n^\star,\mathbf{w}_m^\star)\in\tilde{\mathcal{U}}^0\}$.

Lemma 2 (Bound on the number of differences). *Under Assumptions 1 and 2, for any configurations $(\mathbf{z}_n,\mathbf{w}_m)\in\tilde{\mathcal{U}}$ and $(\mathbf{z}_n^\star,\mathbf{w}_m^\star)\in\tilde{\mathcal{U}}^0$, we have*

$$\operatorname{diff}\big(\mathbf{z}_n,\mathbf{w}_m,\mathbf{z}_n^\star,\mathbf{w}_m^\star\big) := \big|\big\{(i,j)\in\mathcal{I};\pi^\star_{z_i w_j}\neq\pi^\star_{z_i^\star w_j^\star}\big\}\big| \ge \frac{\mu_{\min}^2}{8}(mr_1+nr_2), \tag{17}$$

where the distance $d((\mathbf{z}_n,\mathbf{w}_m),(\mathbf{z}_n^\star,\mathbf{w}_m^\star))$ is attained for some permutations $(s,t)\in\mathfrak{S}$ and we set $r_1:=\|\mathbf{z}_n-s(\mathbf{z}_n^\star)\|_0$ and $r_2:=\|\mathbf{w}_m-t(\mathbf{w}_m^\star)\|_0$.

According to Assumption 4, if $\pi^\star_{z_i w_j}\neq\pi^\star_{z_i^\star w_j^\star}$, the divergence $D(\pi^\star_{z_i^\star w_j^\star}\parallel\pi^\star_{z_i w_j})$ is at least $\kappa_{\min}$. We thus get

$$\sum_{(i,j)\in\tilde{\mathcal{I}}} D\big(\pi^\star_{z_i^\star w_j^\star}\parallel\pi^\star_{z_i w_j}\big) \ge \frac{\mu_{\min}^2\kappa_{\min}}{8}(mr_1+nr_2).$$

Coming back to (16) and (15), we obtain

$$\begin{aligned}&\sum_{(i,j)\in\tilde{\mathcal{I}}} \int_{\mathcal{X}} \log\bigg(\frac{f(x;\pi_{z_i^\star w_j^\star})}{f(x;\pi_{z_i w_j})}\bigg) f\big(x;\pi^\star_{z_i^\star w_j^\star}\big)\,\mathrm{d}x \\ &\quad\ge \bigg(\frac{\mu_{\min}^2\kappa_{\min}}{8} - 2L_0\big\|\boldsymbol{\pi}-\boldsymbol{\pi}^\star\big\|_\infty\bigg)(mr_1+nr_2)\end{aligned}$$

and thus conclude

$$\mathbb{E}_{\star}^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\left(\delta^{\boldsymbol{\pi}}\left(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m\right)\right) \geq \left(\frac{\mu_{\min}^2 \kappa_{\min}}{8} - 2L_0\left\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\right\|_\infty\right)(mr_1 + nr_2).$$

By letting $c = \mu_{\min}^2 \kappa_{\min}/16$, we obtain exactly (14). □

In the following, we will consider asymptotic results where both n and m increase to infinity. The next assumption settles the relative rates of convergence of n and m in LBM. With no loss of generality, we assume in the following that $n \geq m$, view $m = m_n$ as a sequence depending on n and state the convergence results with respect to $n \to +\infty$. Note that the assumption is trivial for SBM.

Assumption 6 (Asymptotic setup). *The sequence $(m_n)_{n\geq 1}$ converges to infinity under the constraints $m_n \leq n$ and $(\log n)/m_n \to 0$.*

We now state the main theorem.

Theorem 1. *Under Assumptions 1 to 6, following the notation of Proposition 1, for any $\eta \in (0, c/(2L_0))$, there exists a family $\{\varepsilon_{n,m}\}_{n,m}$ of positive real numbers with $\sum_n \varepsilon_{n,m_n} < +\infty$, such that on a set Ω_1 whose $\mathbb{P}_\star$-probability is at least $1 - \varepsilon_{n,m}$ and for any $\theta = (\boldsymbol{\mu}, \boldsymbol{\pi}) \in \Theta$ satisfying $\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\|_\infty \leq \eta$, we have for any $(\mathbf{z}_n, \mathbf{w}_m) \in \mathcal{U}$ and any $(s, t) \in \mathfrak{S}$,*

$$\begin{aligned}
&\log \frac{p_{n,m}^\theta(s(\mathbf{Z}_n), t(\mathbf{W}_m))}{p_{n,m}^\theta(\mathbf{z}_n, \mathbf{w}_m)} \\
&\quad \geq \begin{cases} \left(c - 2L_0\left\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\right\|_\infty\right)(mr_1 + nr_2) - K\left(\left\|s(\mathbf{Z}_n) - \mathbf{z}_n\right\|_0 + \left\|t(\mathbf{W}_m) - \mathbf{w}_m\right\|_0\right) \\ \qquad (\textit{LBM}), \\ \left(c - 2L_0\left\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\right\|_\infty\right)2nr_1 - K\left\|s(\mathbf{Z}_n) - \mathbf{z}_n\right\|_0 \\ \qquad (\textit{SBM}), \end{cases}
\end{aligned} \tag{18}$$

and

$$\begin{aligned}
&\log \frac{p_{n,m}^\theta(s(\mathbf{Z}_n), t(\mathbf{W}_m))}{p_{n,m}^\theta(\mathbf{z}_n, \mathbf{w}_m)} \\
&\quad \leq \begin{cases} C(mr_1 + nr_2) + K\left(\left\|s(\mathbf{Z}_n) - \mathbf{z}_n\right\|_0 + \left\|t(\mathbf{W}_m) - \mathbf{w}_m\right\|_0\right) & (\textit{LBM}), \\ C2nr_1 + K\left\|s(\mathbf{Z}_n) - \mathbf{z}_n\right\|_0 & (\textit{SBM}), \end{cases}
\end{aligned} \tag{19}$$

where the distance $d((s(\mathbf{Z}_n), t(\mathbf{W}_m)), (\mathbf{z}_n, \mathbf{w}_m))$, which does not depend on (s, t), is attained for some permutation $(\tilde{s}, \tilde{t}) \in \mathfrak{S}$ and we set $r_1 := \|\mathbf{Z}_n - \tilde{s}(\mathbf{z}_n)\|_0$ and $r_2 := \|\mathbf{W}_m - \tilde{t}(\mathbf{w}_m)\|_0$ and $K = \log(\alpha_{\max}/\alpha_{\min}) \vee \log(\beta_{\max}/\beta_{\min})$.

Let us comment this result. Inequalities (18) and (19) provide a control of the concentration of the posterior distribution on the actual (random) configuration $(\mathbf{Z}_n, \mathbf{W}_m)$, *viewed as an equivalence class in $\tilde{\mathcal{U}}$*. The most important one is (18) that provides a lower bound on the posterior probability of any configuration equivalent to the actual configuration $(\mathbf{Z}_n, \mathbf{W}_m)$ compared to any other configuration $(\mathbf{z}_n, \mathbf{w}_m)$. In this inequality, two different distances appear between these configurations, namely the ℓ_0 distance and the distance $d(\cdot,\cdot)$ given by (4), on the set of actual configurations (so that $d(\cdot,\cdot)$ is linked with the parameter $\boldsymbol{\pi}$ and its symmetries). When the subgroup $\mathfrak{S}$ is reduced to identity (no symmetries allowed in $\boldsymbol{\pi}$), these two distances coincide and the statement substantially simplifies. Another case where it simplifies is when $K = 0$, corresponding to $\alpha_{\max} = \alpha_{\min}$ and $\beta_{\max} = \beta_{\min}$ or equivalently to uniform group proportions. These two particular cases are further expanded below in the first two corollaries. In general, the two different distances appear and play a different role in this inequality. In particular, consider Inequality (18) with for instance $s = Id = t$. It may be the case that a putative configuration $(\mathbf{z}_n, \mathbf{w}_m)$ is equivalent to the actual random one $(\mathbf{Z}_n, \mathbf{W}_m)$ in the sense of relation $\sim$, and thus their distance $d(\cdot,\cdot)$ is zero ($r_1 = r_2 = 0$ above), but their ℓ_0 distance is large. Then, the posterior distribution $p^{\theta}_{n,m}$ will not concentrate on $(\mathbf{z}_n, \mathbf{w}_m)$ due to the existence of different group proportions $\boldsymbol{\mu}$ that help distinguish between $(\mathbf{Z}_n, \mathbf{W}_m)$ and this equivalent configuration $(\mathbf{z}_n, \mathbf{w}_m)$. The extent to which the group proportions $\boldsymbol{\mu}$ are different is measured by $K = \log(\alpha_{\max}/\alpha_{\min}) \vee \log(\beta_{\max}/\beta_{\min})$. When this quantity is small compared to the term $c - 2L_0\eta$ (depending on $\boldsymbol{\pi}$, the connectivity part of the parameter) appearing in (18), the term $K(\|\mathbf{Z}_n - \mathbf{z}_n\|_0 + \|\mathbf{W}_m - \mathbf{w}_m\|_0)$ is negligible and the posterior distribution $p^{\theta}_{n,m}$ will not distinguish between the actual configuration and any equivalent one.

Before giving the proof of the theorem, we provide some corollaries that will help understand the importance of the previous result. The first two corollaries deal with special setups and the third one is an attempt to give a general understanding of the behaviour of the groups posterior distribution. All these results state that, under some appropriate assumptions, the posterior distribution $p^{\theta}_{n,m}$ concentrates on the actual random configuration $(\mathbf{Z}_n, \mathbf{W}_m)$, with large probability. We stress the fact that the results are valid for any parameter value θ (satisfying some additional assumption) and not only the true one $\theta^\star$. More precisely, the results are valid at any $\theta = (\boldsymbol{\mu}, \boldsymbol{\pi})$ such that $\boldsymbol{\pi}$ is close enough to the true value $\boldsymbol{\pi}^\star$.

Corollary 1 (Case $\mathfrak{S} = \{(Id, Id)\}$). *Under Assumptions 1 to 6 and when $\mathfrak{S} = \{(Id, Id)\}$, we obtain that on the set Ω_1 whose $\mathbb{P}_\star$-probability is at least $1 - \varepsilon_{n,m}$, for any parameter $\theta = (\boldsymbol{\mu}, \boldsymbol{\pi}) \in \Theta$ satisfying $\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\|_\infty \le \eta$ for small enough η, we have*

$$p^{\theta}_{n,m}(\mathbf{Z}_n, \mathbf{W}_m) \ge 1 - a_{n,m}\exp(a_{n,m}) \quad \textit{and} \quad p^{\theta}_{n,m}(\mathbf{Z}_n, \mathbf{W}_m) \le \big(1 + b_{n,m}\mathrm{e}^{b_{n,m}}\big)^{-1},$$

where

$$\begin{cases} a_{n,m} = \big(n\mathrm{e}^{-(c-2L_0\eta)m+K} + m\mathrm{e}^{-(c-2L_0\eta)n+K}\big); & \\ b_{n,m} = \big(n\mathrm{e}^{-Cm-K} + m\mathrm{e}^{-Cn-K}\big) & (LBM), \\ a_{n,n} = n\mathrm{e}^{-2n(c-2L_0\eta)+K}; \qquad b_{n,n} = n\mathrm{e}^{-2Cn-K} & (SBM), \end{cases} \tag{20}$$

all converge to 0 as $n \to +\infty$. As a consequence, relying on the maximum a posteriori (MAP) procedure, at a parameter value $\hat{\theta} = (\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\pi}})$ such that $\hat{\boldsymbol{\pi}}$ converges to the true parameter value $\boldsymbol{\pi}^\star$, namely

$$(\widehat{\mathbf{Z}}_n, \widehat{\mathbf{W}}_m) := \mathop{\arg\max}_{(\mathbf{z}_n, \mathbf{w}_m) \in \mathcal{U}} p^{\hat{\theta}}_{n,m}(\mathbf{z}_n, \mathbf{w}_m), \qquad \textit{where } \hat{\theta} = (\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\pi}}) \textit{ and } \hat{\boldsymbol{\pi}} \to \boldsymbol{\pi}^\star$$

the number of misclassified rows and/or columns on the set Ω_1

$$\sum_{i=1}^{n} 1\{\hat{Z}_i \neq Z_i\} + \sum_{j=1}^{m} 1\{\hat{W}_j \neq W_j\} \qquad (LBM) \quad or \quad \sum_{i=1}^{n} 1\{\hat{Z}_i \neq Z_i\} \qquad (SBM),$$

is exactly 0 *for large enough* n.

Corollary 2 (Case of uniform group proportions). *Under Assumptions 1 to 6 and when* $K = 0$, *we obtain that on the set* Ω_1, *for any parameter* $\theta = (\boldsymbol{\mu}, \boldsymbol{\pi}) \in \Theta$ *satisfying* $\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\|_\infty \leq \eta$ *for small enough* η, *we have*

$$p^{\theta}_{n,m}\big(\big\{(\mathbf{z}_n, \mathbf{w}_m) \in \mathcal{U}; (\mathbf{z}_n, \mathbf{w}_m) \sim (\mathbf{Z}_n, \mathbf{W}_m)\big\}\big) \geq 1 - |\mathfrak{S}| a_{n,m} \mathrm{e}^{a_{n,m}}$$

and

$$p^{\theta}_{n,m}\big(\big\{(\mathbf{z}_n, \mathbf{w}_m) \in \mathcal{U}; (\mathbf{z}_n, \mathbf{w}_m) \sim (\mathbf{Z}_n, \mathbf{W}_m)\big\}\big) \leq \big(1 + |\mathfrak{S}| b_{n,m} \mathrm{e}^{b_{n,m}}\big)^{-1},$$

where $a_{n,m}$ *and* $b_{n,m}$ *are defined through* (20) *with* $K = 0$ *and converge to* 0 *as* $n \to +\infty$. *Moreover,*

$$p^{\theta}_{n,m}(\mathbf{Z}_n, \mathbf{W}_m) = \frac{1}{|\mathfrak{S}|} p^{\theta}_{n,m}\big(\big\{(\mathbf{z}_n, \mathbf{w}_m) \in \mathcal{U}; (\mathbf{z}_n, \mathbf{w}_m) \sim (\mathbf{Z}_n, \mathbf{W}_m)\big\}\big).$$

Corollary 3 (General case). *Under Assumptions 1 to 6, we obtain that on the set* Ω_1, *for any parameter* $\theta = (\boldsymbol{\mu}, \boldsymbol{\pi}) \in \Theta$ *satisfying* $\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\|_\infty \leq \eta$ *for small enough* η, *we have*

$$p^{\theta}_{n,m}\big(\big\{(\mathbf{z}_n, \mathbf{w}_m) \in \mathcal{U}; (\mathbf{z}_n, \mathbf{w}_m) \sim (\mathbf{Z}_n, \mathbf{W}_m)\big\}\big) \geq 1 - |\mathfrak{S}| a_{n,m} \mathrm{e}^{a_{n,m}}$$

and

$$p^{\theta}_{n,m}\big(\big\{(\mathbf{z}_n, \mathbf{w}_m) \in \mathcal{U}; (\mathbf{z}_n, \mathbf{w}_m) \sim (\mathbf{Z}_n, \mathbf{W}_m)\big\}\big) \leq \big(1 + |\mathfrak{S}| b_{n,m} \mathrm{e}^{b_{n,m}}\big)^{-1},$$

where $a_{n,m}$ *and* $b_{n,m}$ *are defined through* (20) *and converge to* 0 *as* $n \to +\infty$.

Remark 4. Theorem 1 and Corollaries 1 to 3 are expressed in full generality but their results apply both to LBM and SBM with the notation adopted in Section 2. In particular, the expressions given in these statements simplify for SBM as $n = m$, $\mathbf{Z}_n = \mathbf{W}_m$, $s = t$ and $r_1 = r_2$.

Remark 5. Note that the convergence of the posterior distribution (to the set of configurations equivalent to the actual random one) happens at a rate determined by the constant

$$c - 2L_0\eta > 0.$$

Typically, the rate of this convergence is fast when $\boldsymbol{\pi}$ is not too different from $\boldsymbol{\pi}^\star$ (namely $\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\|_\infty$ and thus $L_0\eta$ small) while the connectivity parameters are sufficiently distinct (namely $\kappa_{\min}$ and thus c large).

When $\mathfrak{S} = \{(Id, Id)\}$, the actual configuration has no other equivalent one and the posterior distribution converges to it. When $K = 0$, group proportions are equal and do not discriminate between equivalent configurations. Therefore, all equivalent configurations (if any) are equally likely. In all other cases, the support of the posterior distribution converges to the set of configurations equivalent to the actual one, including the actual one. However, the latter may not be the most likely among those. Provided n and m are large enough, the most likely configuration is the configuration $(\mathbf{z}_n, \mathbf{w}_m)$ equivalent to $(\mathbf{Z}_n, \mathbf{W}_m)$ which maximizes the quantity

$$\sum_{i=1}^{n} \log \alpha_{z_i} + \sum_{j=1}^{m} \log \beta_{w_j} = \sum_{q=1}^{Q} N_q(\mathbf{z}_n) \log \alpha_q + \sum_{l=1}^{L} N_l(\mathbf{w}_m) \log \beta_l.$$

Also note that we control the number of errors made by a maximum a posteriori clustering procedure only in the case where $\mathfrak{S} = \{(Id, Id)\}$, namely when there are no symmetries in the set of matrices $\Pi_{\mathcal{QL}}$. In the other cases, this procedure is likely to select a configuration equivalent to the true one, but not equal to it. We stress again the fact that the equivalence relation is different from the label switching issue that can not be avoided in finite mixture models. Moreover, exactly as for the label switching issue, this phenomenon will not affect clustering performance.

Proof of Theorem 1. We shall exhibit the set Ω_1 on which Inequalities (18) and (19) are satisfied using LBM notation, the case of SBM easily follows.

First, note that we have

$$\log \frac{p_{n,m}^{\theta}(s(\mathbf{Z}_n), t(\mathbf{W}_m))}{p_{n,m}^{\theta}(\mathbf{z}_n, \mathbf{w}_m)} = \delta^{\boldsymbol{\pi}}\big(s(\mathbf{Z}_n), t(\mathbf{W}_m), \mathbf{z}_n, \mathbf{w}_m\big) + \sum_{i=1}^{n} \log\left(\frac{\alpha_{s(Z_i)}}{\alpha_{z_i}}\right) + \sum_{j=1}^{m} \log\left(\frac{\beta_{t(W_j)}}{\beta_{w_j}}\right).$$

Thus, by letting $K = \log(\alpha_{\max}/\alpha_{\min}) \vee \log(\beta_{\max}/\beta_{\min})$, Inequalities (18) and (19) are satisfied as soon as we have

$$\big(c - 2L_0\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\|_\infty\big)(mr_1 + nr_2) \le \delta^{\boldsymbol{\pi}}\big(s(\mathbf{Z}_n), t(\mathbf{W}_m), \mathbf{z}_n, \mathbf{w}_m\big) \le C(mr_1 + nr_2). \tag{21}$$

Note that the latter inequality is defined on the set of equivalent configurations $\tilde{\mathcal{U}}$ and we can thus replace $(s(\mathbf{Z}_n), t(\mathbf{W}_m))$ by $(\mathbf{Z}_n, \mathbf{W}_m)$. Let $(\mathbf{z}_n^\star, \mathbf{w}_m^\star)$ be a fixed configuration in $\tilde{\mathcal{U}}$, consider $(\mathbf{z}_n, \mathbf{w}_m) \in \tilde{\mathcal{U}}$. Whenever $(\mathbf{z}_n, \mathbf{w}_m) \sim (\mathbf{z}_n^\star, \mathbf{w}_m^\star)$, we have $r_1 + r_2 = 0$ and the previous inequality is automatically satisfied. Thus, we consider $(\mathbf{z}_n, \mathbf{w}_m) \in \tilde{\mathcal{U}}$ such that $(\mathbf{z}_n, \mathbf{w}_m) \neq (\mathbf{z}_n^\star, \mathbf{w}_m^\star)$ and let $r_1 := \|\mathbf{z}_n^\star - \tilde{s}(\mathbf{z}_n)\|_0$ and $r_2 := \|\mathbf{w}_m^\star - \tilde{t}(\mathbf{w}_m)\|_0$, where $(\tilde{s}, \tilde{t}) \in \mathfrak{S}$ realizes the distance $d((\mathbf{z}_n^\star, \mathbf{w}_m^\star), (\mathbf{z}_n, \mathbf{w}_m))$. We consider the event

$$\begin{aligned} A\big(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m\big) = {} & \big\{\delta^{\boldsymbol{\pi}}\big(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m\big) < \big(c - 2L_0\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\|_\infty\big)(mr_1 + nr_2)\big\} \\ & \cup \big\{\delta^{\boldsymbol{\pi}}\big(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m\big) > C(mr_1 + nr_2)\big\}, \end{aligned}$$

where the constants $c, C > 0$ have been previously introduced in Proposition 1. We also assume that $\boldsymbol{\pi}$ satisfies $c - 2L_0\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\|_\infty > 0$. According to this same proposition, as soon as the configuration $(\mathbf{z}_n^\star, \mathbf{w}_m^\star)$ is regular in the sense that it belongs to the set $\tilde{\mathcal{U}}^0$ defined through Equation

(6) and following lines, we obtain that on the set $\{(\mathbf{Z}_n, \mathbf{W}_m) = (\mathbf{z}_n^\star, \mathbf{w}_m^\star)\}$, we have

$$2\big(c - 2L_0\big\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\big\|_\infty\big)(mr_1 + nr_2) \le \mathbb{E}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\big(\delta^{\boldsymbol{\pi}}\big(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m\big)\big) \le \frac{C}{2}(mr_1 + nr_2).$$

We now control the probability of this event. Conditionally on $\{(\mathbf{Z}_n, \mathbf{W}_m) = (\mathbf{z}_n^\star, \mathbf{w}_m^\star)\}$, the event $A(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m)$ is included in the two-sided deviation of $\delta^{\boldsymbol{\pi}}(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m)$ from its conditional expectation $\Delta^{\boldsymbol{\pi}}(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m)$ at a distance at least

$$\begin{aligned}&\min\left\{\big(c - 2L_0\big\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\big\|_\infty\big)(mr_1 + nr_2), \frac{C}{2}(mr_1 + nr_2)\right\}\\ &\quad = \big(c - 2L_0\big\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\big\|_\infty\big)(mr_1 + nr_2) \ge (c - 2L_0\eta)(mr_1 + nr_2).\end{aligned}$$

In other words,

$$\begin{aligned}&A\big(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m\big) \cap \big\{(\mathbf{Z}_n, \mathbf{W}_m) = \big(\mathbf{z}_n^\star, \mathbf{w}_m^\star\big)\big\}\\ &\subset \Big(\big\{\big(\delta^{\boldsymbol{\pi}} - \Delta^{\boldsymbol{\pi}}\big)\big(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m\big) < -\big(c - 2L_0\big\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\big\|_\infty\big)(mr_1 + nr_2)\big\}\\ &\qquad \cup \left\{\big(\delta^{\boldsymbol{\pi}} - \Delta^{\boldsymbol{\pi}}\big)\big(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m\big) > \frac{C}{2}(mr_1 + nr_2)\right\}\Big)\\ &\subset \big\{\big|\big(\delta^{\boldsymbol{\pi}} - \Delta^{\boldsymbol{\pi}}\big)\big(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m\big)\big| > (c - 2L_0\eta)(mr_1 + nr_2)\big\},\end{aligned}$$

where the last inclusion comes from $(c - 2L_0\eta) \le C/2$.

Combining this sets' inclusions with Assumption 3 yields

$$\begin{aligned}&\mathbb{P}_\star\big(A\big(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m\big) \cap \big\{(\mathbf{Z}_n, \mathbf{W}_m) = \big(\mathbf{z}_n^\star, \mathbf{w}_m^\star\big)\big\}\big)\\ &\quad \le \mathbb{P}_\star\big((\mathbf{Z}_n, \mathbf{W}_m) = \big(\mathbf{z}_n^\star, \mathbf{w}_m^\star\big)\big)\\ &\qquad \times \mathbb{P}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\big(\big|\big(\delta^{\boldsymbol{\pi}} - \Delta^{\boldsymbol{\pi}}\big)\big(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m\big)\big| > (c - 2L_0\eta)(mr_1 + nr_2)\big)\\ &\quad \le 2\exp\big[-\psi^\star(c - 2L_0\eta)(mr_1 + nr_2)\big]\boldsymbol{\mu}\big(\mathbf{z}_n^\star, \mathbf{w}_m^\star\big).\end{aligned} \tag{22}$$

We now consider the set Ω_1 defined by

$$\begin{aligned}\Omega_1 &= \Omega_0 \cap \Big(\bigcap_{(\mathbf{z}_n, \mathbf{w}_m) \in \tilde{\mathcal{U}}} \overline{A(\mathbf{Z}_n, \mathbf{W}_m, \mathbf{z}_n, \mathbf{w}_m)}\Big)\\ &= \bigcup_{(\mathbf{z}_n^\star, \mathbf{w}_m^\star) \in \tilde{\mathcal{U}}^0} \bigcap_{(\mathbf{z}_n, \mathbf{w}_m) \in \tilde{\mathcal{U}}} \big(\overline{A\big(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m\big)} \cap \big\{(\mathbf{Z}_n, \mathbf{W}_m) = \big(\mathbf{z}_n^\star, \mathbf{w}_m^\star\big)\big\}\big).\end{aligned} \tag{23}$$

On the set Ω_1, Inequality (21) and thus Inequalities (18) and (19) are both satisfied. We let

$$\tilde{\mathcal{U}}^{\mathbf{z}_n^\star \mathbf{w}_m^\star} := \tilde{\mathcal{U}} \setminus \big\{\big(\mathbf{z}_n^\star, \mathbf{w}_m^\star\big)\big\} = \tilde{\mathcal{U}} \setminus \big\{\big(s\big(\mathbf{z}_n^\star\big), t\big(\mathbf{w}_m^\star\big)\big); (s, t) \in \mathfrak{S}\big\},$$

be the set of all configurations but those which are equivalent to $(\mathbf{z}_n^\star, \mathbf{w}_m^\star)$. Since for any $(s,t) \in \mathfrak{S}$, the event $A(\mathbf{z}_n^\star, \mathbf{w}_m^\star, s(\mathbf{z}_n^\star), t(\mathbf{w}_m^\star))$ has $\mathbb{P}_\star$-probability zero, we may write

$$\overline{\Omega}_1 = \overline{\Omega}_0 \cup \Bigg(\bigcup_{(\mathbf{z}_n^\star, \mathbf{w}_m^\star)\in \tilde{\mathcal{U}}^0} \bigcup_{(\mathbf{z}_n, \mathbf{w}_m)\in \tilde{\mathcal{U}}^{\mathbf{z}_n^\star \mathbf{w}_m^\star}} A\big(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m\big) \cap \big\{(\mathbf{Z}_n, \mathbf{W}_m) = \big(\mathbf{z}_n^\star, \mathbf{w}_m^\star\big)\big\} \Bigg).$$

We now partition the set of configurations $(\mathbf{z}_n, \mathbf{w}_m) \in \tilde{\mathcal{U}}^{\mathbf{z}_n^\star \mathbf{w}_m^\star}$ according to the distance of each point $(\mathbf{z}_n, \mathbf{w}_m)$ to $(\mathbf{z}_n^\star, \mathbf{w}_m^\star)$. We write the following disjoint union

$$\begin{aligned}
\tilde{\mathcal{U}}^{\mathbf{z}_n^\star \mathbf{w}_m^\star} &:= \bigsqcup_{r_1+r_2=1}^{n+m} \tilde{\mathcal{U}}^{\mathbf{z}_n^\star \mathbf{w}_m^\star}(r_1, r_2) \\
&:= \bigsqcup_{r_1+r_2=1}^{n+m} \big\{(\mathbf{z}_n, \mathbf{w}_m) \in \tilde{\mathcal{U}}^{\mathbf{z}_n^\star \mathbf{w}_m^\star}; d\big(\big(\mathbf{z}_n^\star, \mathbf{w}_m^\star\big); \\
&\qquad (\mathbf{z}_n, \mathbf{w}_m)\big) = \big\|\mathbf{z}_n^\star - s(\mathbf{z}_n)\big\|_0 + \big\|\mathbf{w}_m^\star - t(\mathbf{w}_m)\big\|_0 \text{ and} \\
&\qquad \big\|\mathbf{z}_n^\star - s(\mathbf{z}_n)\big\|_0 = r_1, \big\|\mathbf{w}_m^\star - t(\mathbf{w}_m)\big\|_0 = r_2\big\}.
\end{aligned} \tag{24}$$

Note that the above decomposition is not unique. Indeed, we may have that the distance $d((\mathbf{z}_n^\star, \mathbf{w}_m^\star); (\mathbf{z}_n, \mathbf{w}_m)) = r_1 + r_2 = r_1' + r_2'$ but $r_1 \neq r_1'$ and $r_2 \neq r_2'$. In such a case, we make an arbitrary choice between the couples (r_1, r_2) and (r_1', r_2') to represent the distance from $(\mathbf{z}_n, \mathbf{w}_m)$ to $(\mathbf{z}_n^\star, \mathbf{w}_m^\star)$. This decomposition leads to

$$\begin{aligned}
\mathbb{P}_\star(\overline{\Omega}_1) \leq \mathbb{P}_\star(\overline{\Omega}_0) + 2 &\sum_{(\mathbf{z}_n^\star, \mathbf{w}_m^\star)\in \tilde{\mathcal{U}}^0} \mu\big(\mathbf{z}_n^\star, \mathbf{w}_m^\star\big) \\
&\times \sum_{r_1+r_2=1}^{n+m} \big|\tilde{\mathcal{U}}^{\mathbf{z}_n^\star \mathbf{w}_m^\star}(r_1, r_2)\big| \exp\big[-\psi^\star(c - 2L_0\eta)(mr_1 + nr_2)\big].
\end{aligned}$$

Now, we use the bound

$$\big|\tilde{\mathcal{U}}^{\mathbf{z}_n^\star \mathbf{w}_m^\star}(r_1, r_2)\big| \leq |\mathfrak{S}| \binom{n}{r_1}\binom{m}{r_2}, \tag{25}$$

which leads to

$$\begin{aligned}
\mathbb{P}_\star(\overline{\Omega}_1) &\leq \mathbb{P}_\star(\overline{\Omega}_0) + 2 \sum_{r_1+r_2=1}^{n+m} |\mathfrak{S}| \binom{n}{r_1}\binom{m}{r_2} \exp\big[-\psi^\star(c - 2L_0\eta)(mr_1 + nr_2)\big] \\
&\leq \mathbb{P}_\star(\overline{\Omega}_0) + 2|\mathfrak{S}|\big[\big\{1 + \exp\big[-m\psi^\star(c - 2L_0\eta)\big]\big\}^n \big\{1 + \exp\big[-n\psi^\star(c - 2L_0\eta)\big]\big\}^m - 1\big].
\end{aligned}$$

We now rely on the following bound, valid for any $u, v > 0$,

$$(1+u)^n \times (1+v)^m - 1 \leq (nu + mv)\exp(nu + mv). \tag{26}$$

Combining the latter with the control of the probability of $\overline{\Omega}_0$ given in Proposition 1, we obtain

$$\mathbb{P}_\star(\overline{\Omega}_1) \leq 2QL \exp\big(-(n \wedge m)\mu_{\min}^2/2\big) + 2|\mathfrak{S}| d_{n,m} \exp(d_{n,m}),$$

where $d_{n,m} = [n \exp\{-\psi^\star(c - 2L_0\eta)m\} + m \exp\{-\psi^\star(c - 2L_0\eta)n\}]$.

Note that as soon as $(m_n)_{n\geq 1}$ is a sequence such that $m_n \to +\infty$ and $(\log n)/m_n \to 0$, we obtain that for any constant $a > 0$, the sequence $u_n = n \exp(-am_n)$ is negligible with respect to n^{-1-s}, for any $s > 0$, and thus $\sum_n u_n < +\infty$. In particular, the sequence

$$\varepsilon_{n,m} := 2QL \exp\big[-(n \wedge m)\mu_{\min}^2/2\big] + 2|\mathfrak{S}| d_{n,m} \exp(d_{n,m})$$

satisfies $\sum_n \varepsilon_{n,m_n} < +\infty$. As for SBM, it is easy to see that this expression reduces to

$$\varepsilon_{n,n} := 2Q \exp\big[-n\alpha_{\min}^2/2\big] + 2|\mathfrak{S}| d_n \exp(d_n)$$

with $d_n = n \exp\{-2\psi^\star(c - 2L_0\eta)n\}$ and which also satisfies $\sum_n \varepsilon_{n,n} < +\infty$. This concludes the proof. □

Proof of Corollaries 1, 2 and 3. The proof of these three corollaries relies on the same scheme that we shall now present in LBM notation. The proof is easily generalised to SBM. First, note that $\Omega_1 = \bigcup_{(\mathbf{z}_n^\star, \mathbf{w}_m^\star) \in \mathcal{U}^0} (\Omega_1 \cap \{(\mathbf{Z}_n, \mathbf{W}_m) = (\mathbf{z}_n^\star, \mathbf{w}_m^\star)\})$. Let us fix some configuration $(\mathbf{z}_n^\star, \mathbf{w}_m^\star)$ in $\mathcal{U}^0$. On the set $\Omega_1 \cap \{(\mathbf{Z}_n, \mathbf{W}_m) = (\mathbf{z}_n^\star, \mathbf{w}_m^\star)\}$, we have

$$\begin{aligned}
1 - p_{n,m}^\theta\big(\{(\mathbf{Z}_n, \mathbf{W}_m)\}\big) &\leq \frac{1 - p_{n,m}^\theta(\{(\mathbf{Z}_n, \mathbf{W}_m)\})}{p_{n,m}^\theta(\{(\mathbf{Z}_n, \mathbf{W}_m)\})} \\
&= \sum_{\substack{(\mathbf{z}_n, \mathbf{w}_m) \in \mathcal{U} \\ (\mathbf{z}_n, \mathbf{w}_m) \nsim (\mathbf{z}_n^\star, \mathbf{w}_m^\star)}} \exp\left(-\log \frac{p_{n,m}^\theta(\{(\mathbf{z}_n^\star, \mathbf{w}_m^\star)\})}{p_{n,m}^\theta(\mathbf{z}_n, \mathbf{w}_m)}\right),
\end{aligned}$$

where we abbreviate to $\{(\mathbf{Z}_n, \mathbf{W}_m)\}$ and $\{(\mathbf{z}_n^\star, \mathbf{w}_m^\star)\}$ the whole sets of configurations $\{(\mathbf{z}_n, \mathbf{w}_m) \sim (\mathbf{Z}_n, \mathbf{W}_m)\}$ and $\{(\mathbf{z}_n, \mathbf{w}_m) \sim (\mathbf{z}_n^\star, \mathbf{w}_m^\star)\}$, respectively. Let $(\mathbf{z}_n, \mathbf{w}_m) \nsim (\mathbf{z}_n^\star, \mathbf{w}_m^\star)$. There exists $(s, t) \in \mathfrak{S}$ such that $\|\mathbf{z}_n - s(\mathbf{z}_n^\star)\|_0 = r_1$ and $\|\mathbf{w}_m - t(\mathbf{w}_m^\star)\|_0 = r_2$. Using Inequality (18) and $\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\|_\infty \leq \eta$, we get

$$\log \frac{p_{n,m}^\theta(\{(\mathbf{z}_n^\star, \mathbf{w}_m^\star)\})}{p_{n,m}^\theta(\mathbf{z}_n, \mathbf{w}_m)} \geq \log \frac{p_{n,m}^\theta(s(\mathbf{z}_n^\star), t(\mathbf{w}_m^\star))}{p_{n,m}^\theta(\mathbf{z}_n, \mathbf{w}_m)} \geq (c - 2L_0\eta)(mr_1 + nr_2) + K(r_1 + r_2)$$

and therefore

$$1 - p_{n,m}^\theta\big(\{(\mathbf{Z}_n, \mathbf{W}_m)\}\big) \leq \sum_{\substack{(\mathbf{z}_n, \mathbf{w}_m) \in \mathcal{U} \\ (\mathbf{z}_n, \mathbf{w}_m) \nsim (\mathbf{z}_n^\star, \mathbf{w}_m^\star)}} \exp\big[-(c - 2L_0\eta)(mr_1 + nr_2) + K(r_1 + r_2)\big]. \quad (27)$$

When $\mathfrak{S} = \{(Id, Id)\}$, the set $\{(\mathbf{z}_n, \mathbf{w}_m) \sim (\mathbf{Z}_n, \mathbf{W}_m)\}$ reduces to a singleton and the previous bound becomes

$$1 - p^{\theta}_{n,m}(\mathbf{Z}_n, \mathbf{W}_m) \leq \sum_{\substack{(\mathbf{z}_n,\mathbf{w}_m)\in\mathcal{U} \\ (\mathbf{z}_n,\mathbf{w}_m)\nsim(\mathbf{z}^\star_n,\mathbf{w}^\star_m)}} \exp\bigl[-(c - 2L_0\eta)(mr_1 + nr_2) + K(r_1 + r_2)\bigr].$$

Using the decomposition (24) on the set $\tilde{\mathcal{U}}^{\mathbf{z}^\star_n \mathbf{w}^\star_m}$ and the bound (25) on the cardinality of each $\tilde{\mathcal{U}}^{\mathbf{z}^\star_n,\mathbf{w}^\star_m}(r_1, r_2)$, we get

$$\begin{aligned} 1 - p^{\theta}_{n,m}(\mathbf{Z}_n, \mathbf{W}_m) &\leq \sum_{r_1+r_2=1}^{n+m} \binom{n}{r_1}\binom{m}{r_2} \exp\bigl[-(c - 2L_0\eta)(mr_1 + nr_2) + K(r_1 + r_2)\bigr] \\ &= \bigl\{\bigl(1 + \exp(-mc_1 + K)\bigr)^n\bigl(1 + \exp(-nc_1 + K)\bigr)^m - 1\bigr\}, \end{aligned}$$

where $c_1 = c - 2L_0\eta$. Using again Inequality (26), we obtain

$$1 - p^{\theta}_{n,m}(\mathbf{Z}_n, \mathbf{W}_m) \leq a_{n,m} \exp(a_{n,m}),$$

where $a_{n,m} = (n\mathrm{e}^{-(c-2L_0\eta)m+K} + m\mathrm{e}^{-(c-2L_0\eta)n+K})$. In SBM, this quantity becomes $a_{n,n} = n\mathrm{e}^{-2(c-2L_0\eta)n+K}$.

The case where $K = 0$ is handled similarly and gives

$$1 - p^{\theta}_{n,m}\bigl(\bigl\{(\mathbf{Z}_n, \mathbf{W}_m)\bigr\}\bigr) \leq |\mathfrak{S}| a_{n,m} \exp(a_{n,m}),$$

with the same definition of $a_{n,m}$, replacing K with 0.

Moreover when $K = 0$, we have $\alpha_1 = \cdots = \alpha_Q$ and $\beta_1 = \cdots = \beta_L$ and it easy to check that

$$p^{\theta}_{n,m}(\mathbf{Z}_n, \mathbf{W}_m) = p^{\theta}_{n,m}\bigl(s(\mathbf{Z}_n), t(\mathbf{W}_m)\bigr)$$

for all $(s, t) \in \mathfrak{S}$.

Now, in the general case, we come back to (27). Using the decomposition (24) on the set $\tilde{\mathcal{U}}^{\mathbf{z}^\star_n \mathbf{w}^\star_m}$ and the bound (25) on the cardinality of each $\tilde{\mathcal{U}}^{\mathbf{z}^\star_n,\mathbf{w}^\star_m}(r_1, r_2)$, we get

$$\begin{aligned} 1 - p^{\theta}_{n,m}\bigl(\bigl\{(\mathbf{Z}_n, \mathbf{W}_m)\bigr\}\bigr) &\leq \sum_{r_1+r_2=1}^{n+m} |\mathfrak{S}| \binom{n}{r_1}\binom{m}{r_2} \exp\bigl[-(c - 2L_0\eta)(mr_1 + nr_2) + K(r_1 + r_2)\bigr] \\ &\leq |\mathfrak{S}| \bigl\{\bigl(1 + \exp(-mc_1 + K)\bigr)^n\bigl(1 + \exp(-nc_1 + K)\bigr)^m - 1\bigr\}, \end{aligned}$$

where $c_1 = c - 2L_0\eta$. Using again Inequality (26), we obtain

$$1 - p^{\theta}_{n,m}\bigl(\bigl\{(\mathbf{z}_n, \mathbf{w}_m) \sim (\mathbf{Z}_n, \mathbf{W}_m)\bigr\}\bigr) \leq |\mathfrak{S}| a_{n,m} \exp(a_{n,m}),$$

with same definition of $a_{n,m}$ as previously.

We now provide an upper bound for the posterior probability of the class $\{(\mathbf{z}_n, \mathbf{w}_m) \sim (\mathbf{Z}_n, \mathbf{W}_m)\}$, valid on the set Ω_1. Let us fix some configuration $(\mathbf{z}_n^\star, \mathbf{w}_m^\star)$ in $\mathcal{U}^0$. On the set $\Omega_1 \cap \{(\mathbf{Z}_n, \mathbf{W}_m) = (\mathbf{z}_n^\star, \mathbf{w}_m^\star)\}$, we have

$$\frac{1}{p_{n,m}^\theta(\{(\mathbf{Z}_n, \mathbf{W}_m)\})} = 1 + \sum_{(\mathbf{z}_n, \mathbf{w}_m) \nsim (\mathbf{Z}_n, \mathbf{W}_m)} \exp\biggl(-\log \frac{p_{n,m}^\theta(\{(\mathbf{Z}_n, \mathbf{W}_m)\})}{p_{n,m}^\theta(\mathbf{z}_n, \mathbf{w}_m)}\biggr)$$

and relying on Inequality (19), we get

$$\begin{aligned} p_{n,m}^\theta\bigl(\bigl\{(\mathbf{z}_n^\star, \mathbf{w}_m^\star)\bigr\}\bigr) &\leq \Biggl\{1 + \sum_{(\mathbf{z}_n, \mathbf{w}_m) \nsim (\mathbf{z}_n^\star, \mathbf{w}_m^\star)} \exp\biggl(-\log \frac{p_{n,m}^\theta(\{(\mathbf{z}_n^\star, \mathbf{w}_m^\star)\})}{p_{n,m}^\theta(\mathbf{z}_n, \mathbf{w}_m)}\biggr)\Biggr\}^{-1} \\ &\leq \Biggl\{1 + \sum_{(\mathbf{z}_n, \mathbf{w}_m) \nsim (\mathbf{z}_n^\star, \mathbf{w}_m^\star)} \exp\bigl(-C(mr_1 + nr_2) - K(r_1 + r_2)\bigr)\Biggr\}^{-1}. \end{aligned}$$

Following the same lines, we obtain the desired upper-bounds. □

4. Examples of application

The goal of this section is to derive the results of Theorem 1 and following corollaries in many different setups. The key ingredient for that lies in establishing the concentration of the ratio $\delta^{\boldsymbol{\pi}}$ around its conditional expectation $\Delta^{\boldsymbol{\pi}}$ (namely Assumption 3). As mentioned in Remarks 2 and 3, it is valid for many exponential families. We will first present the general proof for exponential families and then state the results for common exponential families.

4.1. Scheme of proof of concentration inequalities

One of the main issues for Theorem 1 to be valid is the existence of a concentration of the ratio $\delta^{\boldsymbol{\pi}}$ around its conditional expectation $\Delta^{\boldsymbol{\pi}}$, namely Assumption 3. This section presents the general methodology that will be employed.

The scheme of proof is as follows. Relying on the notation of Assumption 3 and using (9), we write

$$\begin{aligned} &\delta^{\boldsymbol{\pi}}\bigl(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m\bigr) - \Delta^{\boldsymbol{\pi}}\bigl(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m\bigr) \\ &\quad = \sum_{(i,j)\in\mathcal{I}} \log\biggl(\frac{f(X_{ij}; \pi_{z_i^\star w_j^\star})}{f(X_{ij}; \pi_{z_i w_j})}\biggr) - \mathbb{E}_\theta^{\mathbf{z}_n^\star \mathbf{w}_m^\star} \log\biggl(\frac{f(X_{ij}; \pi_{z_i^\star w_j^\star})}{f(X_{ij}; \pi_{z_i w_j})}\biggr) := \sum_{(i,j)\in\mathcal{I}} Y_{ij}. \end{aligned}$$

Conditional on $(\mathbf{Z}_n, \mathbf{W}_m) = (\mathbf{z}_n^\star, \mathbf{w}_m^\star)$, the random variables Y_{ij} are independent and centered. There are exactly $D := \mathrm{diff}(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m)$ such non-null variables and since $D \leq mr_1 + nr_2 -$

$r_1 r_2 \le m r_1 + n r_2$, we may write

$$\mathbb{P}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\big(\big|(\delta^{\boldsymbol{\pi}} - \Delta^{\boldsymbol{\pi}})(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m)\big| \ge \varepsilon(m r_1 + n r_2)\big) \le \mathbb{P}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\bigg(\bigg|\sum_{(i,j)\in\mathcal{I}} Y_{ij}\bigg| \ge \varepsilon D\bigg). \quad (28)$$

Thus, the problem boils down to establishing a concentration inequality for the sum $\sum Y_{ij}$ composed of D conditionally independent and centered random variables. As soon as we have the existence of a positive function $\psi_{\max}^\star$ such that for any $\epsilon > 0$,

$$\mathbb{P}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\bigg(\bigg|\sum_{(i,j)\in\mathcal{I}} Y_{ij}\bigg| \ge \varepsilon D\bigg) \le 2\exp\big\{-\psi_{\max}^\star(\varepsilon) D\big\}, \quad (29)$$

we can combine Lemma 2 and bound (28) to obtain

$$\begin{aligned}\mathbb{P}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\bigg(\bigg|\sum_{(i,j)\in\mathcal{I}} Y_{ij}\bigg| \ge \varepsilon(m r_1 + n r_2)\bigg) &\le 2\exp\big\{-\psi_{\max}^\star(\varepsilon)\mu_{\min}^2(m r_1 + n r_2)/8\big\} \\ &:= 2\exp\big\{-\psi^\star(\varepsilon)(m r_1 + n r_2)\big\},\end{aligned}$$

with $\psi^\star(\cdot) = \psi_{\max}^\star(\cdot)\mu_{\min}^2/8$. Note that Inequality (29) is often obtained through a Cramer–Chernoff bound in the following way. We let $\psi_{ij}(\lambda) := \log \mathbb{E}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}(\exp(\lambda Y_{ij}))$, for any $\lambda > 0$ such that this quantity is finite, let us say $\lambda \in I \subset \mathbb{R}$. Using a Cramer–Chernoff bound, we get for any $x > 0$,

$$\mathbb{P}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\big(|Y_{ij}| \ge x\big) \le 2\exp\Big\{-\sup_{\lambda\in I}\big(\lambda x - \psi_{ij}(\lambda)\big)\Big\}.$$

As soon as we can uniformly bound this quantity (uniformly with respect to i, j and also underlying $\boldsymbol{\pi}$), namely if we can write

$$\mathbb{P}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\big(|Y_{ij}| \ge x\big) \le 2\exp\Big\{-\sup_{\lambda\in I}\big(\lambda x - \psi_{\max}(\lambda)\big)\Big\},$$

with $\psi_{\max} := \sup_{\boldsymbol{\pi}\in\Pi_{\mathcal{QL}}} \max_{(i,j)\in\mathcal{I}} \psi_{ij}$, the conditional independence of the Y_{ij}'s gives that for any $\epsilon > 0$, and any $\lambda > 0$,

$$\mathbb{P}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\bigg(\bigg|\sum_{(i,j)\in\mathcal{I}} Y_{ij}\bigg| \ge \varepsilon D\bigg) \le 2\exp\big\{-\big(\lambda\epsilon D - D\psi_{\max}(\lambda)\big)\big\},$$

leading to

$$\mathbb{P}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\bigg(\bigg|\sum_{(i,j)\in\mathcal{I}} Y_{ij}\bigg| \ge \varepsilon D\bigg) \le 2\exp\Big\{-D\sup_{\lambda\in I}\big(\lambda\epsilon - \psi_{\max}(\lambda)\big)\Big\} \le 2\exp\big\{-D\psi_{\max}^\star(\epsilon)\big\},$$

where $\psi_{\max}^\star(\epsilon) := \sup_{\lambda\in I}(\lambda\epsilon - \psi_{\max}(\lambda))$. Note that since $\psi_{ij}(0) = 0$, we have $\psi_{\max}(0) = 0$ and $\psi_{\max}^\star$ is non-negative.

4.2. Examples from exponential families

We state here the rate functions $\psi^\star$ and validity assumptions of our main result under several models for the observations X_{ij}, all included in the exponential family framework.

Binary model

Let $X_{ij} \in \{0, 1\}$ and $f(\cdot; \pi)$ a Bernoulli distribution with parameter π. Assumptions 3 to 5 are satisfied if the paramater set is bounded away from 0 and 1, namely $\Pi \subset [a, 1-a]$ for some $a \in (0, 1/2)$. The corresponding rate function, given by Hoeffding's inequality is

$$\psi^\star(x) = x^2\mu_{\min}^2 / \big\{16\big[\log(1-a) - \log a\big]^2\big\}. \tag{30}$$

In the interesting special case where $\Pi \subset \xi[a, 1-a]$ with $\xi > 0$ small, Bernstein's inequality gives the sharper rate function

$$\psi^\star(x) = x^2\mu_{\min}^2 / \big\{64\xi\big[\log(1-a) - \log a\big]^2 + 32x\big[\log(1-a) - \log a\big]/3\big\}$$

which gives the following rate function for deviations of order ξ:

$$\begin{aligned} &\psi^\star(\xi x) = \xi\breve{\psi}^\star(x) \qquad \text{where} \\ &\breve{\psi}^\star(x) = x^2\mu_{\min}^2 / \big\{64\big[\log(1-a) - \log a\big]^2 + 32x\big[\log(1-a) - \log a\big]/3\big\}. \end{aligned} \tag{31}$$

This small deviation rate function is useful for sparse asymptotics, as studied in Section 5.

Binomial model

Let $X_{ij} \in \{0, \dots, p\}$ and $f(\cdot, \pi)$ a binomial distribution $\mathcal{B}(p, \pi)$. Assumptions 3 to 5 are satisfied if the paramater set is bounded away from 0 and 1, namely $\Pi \subset [a, 1-a]$ for some $a \in (0, 1/2)$. The corresponding rate function, given by Hoeffding's inequality is

$$\psi^\star(x) = x^2\mu_{\min}^2 / \big\{16p^2\big[\log(1-a) - \log a\big]^2\big\}.$$

Multinomial model

Let X_{ij} be discrete with p levels labelled 1 to p, parameter $\pi = (\pi(1), \dots, \pi(p))$ and $f(k, \pi) = \pi(k)$. Assumptions 3 to 5 are satisfied if the paramater set for $(\pi(k))_{1 \le k \le p}$ is bounded away from 0 and 1, namely $\Pi \subset [a, 1-a]^p$ for some $a \in (0, 1/2)$. The corresponding rate function, given by Hoeffding's inequality is

$$\psi^\star(x) = x^2\mu_{\min}^2 / \big\{8p\big[\log(1-a) - \log a\big]^2\big\}.$$

Poisson model

Let $X_{ij} \in \mathbb{N}$ and $f(\cdot, \pi)$ is a Poisson distribution with parameter π. Assumptions 3 to 5 are satisfied if the paramater set is bounded away from 0 and $+\infty$, namely $\Pi \subset [\pi_{\min}, \pi_{\max}] \subset (0, +\infty)$. The corresponding rate function, given by a Cramer–Chernoff bound for Poisson variable (see, for instance, [22]) is

$$\psi^\star(x) = \frac{1}{8}\mu_{\min}^2 \pi_{\max} h\left(\frac{x}{\pi_{\max}\log(\pi_{\max}/\pi_{\min})}\right),$$

where $\forall u \geq -1, h(u) = (1+u)\log(1+u) - u$.

Gaussian location model

We are interested here in Gaussian observations in the homoscedastic case. We assume that $X_{ij} \in \mathbb{R}$ and $f(\cdot, \pi)$ is a Gaussian distribution with mean value π and fixed variance σ^2, namely $f(x, \pi_{ij}) = c\exp\{-(x-\pi_{ij})^2/(2\sigma^2)\}$ where c is a normalizing constant. Assumptions 3 to 5 are satisfied if the paramater set is bounded away from $-\infty$ and $+\infty$, namely $\Pi \subset [\pi_{\min}, \pi_{\max}] \subset \mathbb{R}$. The corresponding rate function, given by a Cramer–Chernoff bound for Gaussian variables is

$$\psi^\star(x) = \frac{\mu_{\min}^2 \sigma^2 x^2}{16(\pi_{\max} - \pi_{\min})^2}.$$

Gaussian scale model

We are interested here in Gaussian observations with fixed mean and different variances. We assume that $X_{ij} \in \mathbb{R}$ and $f(\cdot, \pi)$ is a Gaussian distribution with fixed mean value m and variance $\pi \in (0, +\infty)$, namely $f(x, \pi_{ij}) = c(\pi_{ij})^{-1/2}\exp\{-(x-m)^2/(2\pi_{ij}^2)\}$ where c is a normalizing constant. Assumptions 3 to 5 are satisfied if the paramater set is bounded away from 0 and $+\infty$, namely $\Pi \subset [\pi_{\min}, \pi_{\max}] \subset (0, +\infty)$. The corresponding rate function, given by a Cramer–Chernoff bound for $\chi^2(1)$ random variables is

$$\psi^\star(x) = \frac{\mu_{\min}^2 \sigma^2 x}{8(\pi_{\max} - \pi_{\min})} + \frac{\mu_{\min}^2}{16}\log\left\{1 + \frac{2\pi_{\min} x}{\pi_{\max} - \pi_{\min}}\right\}.$$

4.3. Zero-inflated distributions

Here, we assume that X_{ij} follows a mixture of a Dirac mass at zero with another distribution (on $\mathbb{R}$ for instance). This situation is particularly relevant for modeling sparse matrices [4]. In this context, the former parameter π becomes now $(\pi, \gamma) \in (0, 1) \times \Gamma$ and we let

$$f(\cdot; \pi, \gamma) = \pi \tilde{f}(\cdot; \gamma) + (1-\pi)\delta_0(\cdot),$$

where δ_0 is the Dirac mass at 0. The set of parameter matrices $(\boldsymbol{\pi}, \boldsymbol{\gamma}) := ((\pi_{ql}), (\gamma_{ql}))_{q\in\mathcal{Q}, l\in\mathcal{L}}$ is denoted by $\Pi_{\mathcal{QL}} \times \Gamma_{\mathcal{QL}}$. For identifiability reasons, we also constrain the parametric family $\{\tilde{f}(\cdot; \gamma); \gamma \in \Gamma\}$ such that any distribution in this set admits a continuous cumulative distribution function (c.d.f.) at zero. Moreover, we shall assume that the distributions $\{\tilde{f}(\cdot;\gamma);\gamma\in\Gamma\}$ satisfy Assumption 4.

For instance, $\tilde{f}(\cdot;\gamma)$ may be absolutely continuous with respect to the Lebesgue measure. Another interesting case consists in considering the density (with respect to the counting measure) of the Poisson distribution, with parameter γ, but truncated at zero. Namely, for any $k\geq 1$, we let $\tilde{f}(k;\gamma)=\gamma^k/(k!)(\mathrm{e}^{\gamma}-1)^{-1}$. This leads to zero-inflated Poisson models and more generally, one could consider other zero-inflated counts models.

In the following, we will assume that the parameter set Π is included in $[a,1-a]$ for some $a\in(0,1/2)$ and that the family $\{\tilde{f}(\cdot;\gamma);\gamma\in\Gamma\}$ satisfies a concentration property on its likelihood ratio statistics as follows.

Assumption 7. *Fix* $(\mathbf{z}_n^\star,\mathbf{w}_m^\star)\in\tilde{\mathcal{U}}_0$ *and* $(\mathbf{z}_n,\mathbf{w}_m)$ *in* $\tilde{\mathcal{U}}$ *with* $(\mathbf{z}_n^\star,\mathbf{w}_m^\star)\neq(\mathbf{z}_n,\mathbf{w}_m)$. *Let* $\tilde{Y}_{ij}=\log[\tilde{f}(X_{ij};\gamma_{z_i^\star w_j^\star})/\tilde{f}(X_{ij};\gamma_{z_i w_j})]+c$, *where* c *is a centering constant. There exists a positive function* $\tilde{\psi}^\star_{\max}:(0,+\infty)\to(0,+\infty]$ *such that for any* $x>0$, *for any* $(i,j)\in\mathcal{I}$, *and any* $\boldsymbol{\gamma}\in\Gamma_{\mathcal{QL}}$,

$$\mathbb{P}_\star^{\mathbf{z}_n^\star\mathbf{w}_m^\star}\big(|\tilde{Y}_{ij}|\geq x|X_{ij}\neq 0\big)\leq 2\exp\Big\{-\sup_{\lambda\in I}\big(\lambda x-\tilde{\psi}_{\max}(\lambda)\big)\Big\}:=2\exp\big(-\tilde{\psi}^\star_{\max}(x)\big),$$

where $\tilde{\psi}_{\max}(\lambda)=\sup_{\boldsymbol{\gamma}\in\Gamma_{\mathcal{QL}}}\max_{(i,j)\in\mathcal{I}}\log\mathbb{E}_\star^{\mathbf{z}_n^\star\mathbf{w}_m^\star}(\exp(\lambda\tilde{Y}_{ij})|X_{ij}\neq 0)$ *exists for any* $\lambda\in I\subset(0,+\infty)$.

Under these assumptions, it is easy to see that Assumption 3 is satisfied, up to an extra factor 2, with

$$\psi^\star(x)=\big\{\mu^2_{\min}\tilde{\psi}^\star_{\max}(x/2)/8\big\}\wedge\psi^\star_{\mathrm{bin}}(x/2),$$

where $\psi^\star_{\mathrm{bin}}$ is a rate function for binary observations, defined for instance in Equations (30) or (31). Namely, using the same notation as in Assumption 3, we get

$$\mathbb{P}_\star^{\mathbf{z}_n^\star\mathbf{w}_m^\star}\big(\big|\big(\delta^{\boldsymbol{\pi}}-\Delta^{\boldsymbol{\pi}}\big)\big(\mathbf{z}_n^\star,\mathbf{w}_m^\star,\mathbf{z}_n,\mathbf{w}_m\big)\big|\geq\varepsilon\{mr_1+nr_2\}\big)\leq 4\exp\big[-\psi^\star(\varepsilon)\{mr_1+nr_2\}\big].$$

In order to ensure Assumption 5 on $f(\cdot;\pi,\gamma)$, we need the same hypothesis to be satisfied on the family $\{\tilde{f}(\cdot;\gamma);\gamma\in\Gamma\}$.

Assumption 8. *There exists some positive constant* $\tilde{L}_0$ *such that for any* $\boldsymbol{\gamma},\boldsymbol{\gamma}'\in\Gamma_{\mathcal{QL}}$ *and any* $(q,l),(q',l')\in\mathcal{Q}\times\mathcal{L}$, *we have*

$$\left|\int_{\mathcal{X}}\log\frac{\tilde{f}(x;\gamma_{ql})}{\tilde{f}(x;\gamma'_{ql})}\tilde{f}(x;\gamma_{q'l'})\,\mathrm{d}x\right|\leq\tilde{L}_0\|\boldsymbol{\gamma}-\boldsymbol{\gamma}'\|_\infty.$$

Note that we provided in the previous section many examples of families for which this assumption is satisfied. Then, the results of Section 3 apply.

5. Asymptotically decreasing connections density

In this section, we explore the limiting case where the numbers of groups Q and L remain constant while the connections probabilities between groups converge to 0. This framework is interesting as it models the case where groups sizes increase linearly with the number of row/column objects, while the mean number of connections (i.e., non-null observations in the data matrix) increases only sub-linearly, mimicking for example budget constraints in terms of global consumptions. More precisely, we will consider two different setups, the first one being built on the binary case developed in Section Binary model and the second one being built on the weighted case (also called zero-inflated model) from Section 4.3. As in the previous sections, we assume that $m \leq n$, view $m := m_n$ as a sequence depending on n and state the results with respect to $n \to +\infty$. We shall furthermore assume that the probability of connection (binary case) or the sparsity parameter (weighted case) $\pi_{ql,n}$ depends on n and writes $\pi_{ql,n} = \xi_n \pi_{ql}$ where $(\xi_n)_{n\geq 1}$ converges to zero and π_{ql} is a positive constant. The sequence $(\xi_n)_{n\in\mathbb{N}}$ controls the overall density of the block model and acts as a scaling factor while the parameters $(\pi_{ql})_{(q,l)\in\mathcal{Q}\times\mathcal{L}}$ reflect the *unscaled* connection probabilities from the different groups. This parametrization is analogous to the one studied in [5]. We shall now assume that the unscaled connection/sparsity probabilities are well-behaved, and introduce the new parameter sets denoted by Π_n and $\Pi_{\mathcal{QL},n}$ to account for the dependence on the data size (i.e., number of rows/columns).

Assumption 9. *The parameter sets* Π_n *and* $\Pi_{\mathcal{QL},n}$ *depend on the number of observations and we have*

$$\begin{aligned}
\Pi &\subset [a, 1-a] \qquad \textit{for some } a \in (0, 1/2),\\
\Pi_n &:= \xi_n \Pi = \{\xi_n \pi; \pi \in \Pi\},\\
\Pi_{\mathcal{QL}} &\subset \Pi^{QL},\\
\Pi_{\mathcal{QL},n} &:= \xi_n \Pi_{\mathcal{QL}} = \{\xi_n \boldsymbol{\pi}; \boldsymbol{\pi} \in \Pi_{\mathcal{QL}}\},
\end{aligned}$$

where $(\xi_n)_{n\geq 1}$ *is a sequence of values in* $(0, 1]$ *converging to* 0 *and such that*

$$\frac{\log n}{m_n \xi_n} \underset{n\to+\infty}{\to} 0.$$

5.1. Binary block models with a vanishing density

We let $X_{ij} \in \{0, 1\}$ and $f(\cdot; \pi)$ a Bernoulli distribution with parameter π. Here, the connectivity parameter $\boldsymbol{\pi}_n = (\pi_{ql,n})_{(q,l)\in\mathcal{Q}\times\mathcal{L}}$ depends on n and may be arbitrarily close to 0. Accordingly, the constant $\kappa_{\min}(\boldsymbol{\pi}_n)$ defined in (12) depends on n and is no longer bounded away from 0. We thus reconsider Assumptions 3, 5 and the definition of $\kappa_{\min}(\boldsymbol{\pi}_n)$ to exhibit the scaling in n of several key quantities in this setup. The proof of following lemma is postponed to the Appendix.

Lemma 3. *Fix two parameters $\boldsymbol{\pi}_n = \xi_n \boldsymbol{\pi}$ and $\boldsymbol{\pi}'_n = \xi_n \boldsymbol{\pi}'$ in the set $\Pi_{\mathcal{QL},n}$, where $\boldsymbol{\pi}, \boldsymbol{\pi}' \in \Pi_{\mathcal{QL}}$. Under Assumption 9, we have for all n and all $(q,l), (q',l') \in \mathcal{Q} \times \mathcal{L}$*

$$\kappa_{\min,n} := \kappa_{\min}\big(\boldsymbol{\pi}^\star_n\big) \geq \xi_n c_{\min}\big(\boldsymbol{\pi}^\star\big),$$

$$\left| \int_{\mathcal{X}} \log \frac{f(x;\pi_{ql,n})}{f(x;\pi'_{ql,n})} f(x;\pi_{q'l',n})\,\mathrm{d}x \right| \leq \frac{\xi_n \|\boldsymbol{\pi} - \boldsymbol{\pi}'\|_\infty}{a}, \tag{32}$$

$$\psi^\star_n(x) := \psi^\star(\xi_n x) \geq \xi_n \breve{\psi}^\star(x), \tag{33}$$

where

$$\begin{aligned} c_{\min} &:= c_{\min}\big(\boldsymbol{\pi}^\star\big) \\ &= \frac{1}{2}\left(\frac{a}{1-a}\right)^2 \min\left\{ \frac{(\pi^\star_{ql} - \pi^\star_{q'l'})^2}{\pi^\star_{ql}}; (q,l), \big(q',l'\big) \in \mathcal{Q} \times \mathcal{L}, \pi^\star_{ql} \neq \pi^\star_{q'l'} \right\} > 0, \\ \breve{\psi}^\star(x) &= \frac{x^2 \mu^2_{\min}}{64a\{\log(1-a) - \log a\}^2 + 32x\{\log(1-a) - \log a\}/3}. \end{aligned}$$

Corollary 4. *Under Assumption 1 on the unscaled parameter set $\Pi_{\mathcal{QL}}$ and Assumption 9, the conclusions of Theorem 1 and Corollaries 1 to 3 remain valid with the following modifications*

1. $c = \mu^2_{\min} c_{\min}/16$;
2. $L_0 = a^{-1}$;
3. $(c - 2L_0\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\|_\infty)$ *is replaced by* $\xi_n(c - 2L_0\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\|_\infty)$.

Remark 6. Note that Assumption 9 replaces Assumption 6 in this statement. The quantity m_n is replaced by $m_n \xi_n$ which plays the role of average number of connections and must grow faster than $\log n$. The scaling is consistent with results from [5] and [11].

Proof of Corollary 4. The proof is essentially the same as the one of Theorem 1. We will only highlight the differences and show how the scaling $\log n/(m_n \xi_n) \to 0$ is derived. First, Equation (14) from Proposition 1 now depends on n and should be

$$\mathbb{E}^{\mathbf{Z}_n \mathbf{W}_m}_\star \big(\delta^{\boldsymbol{\pi}_n}(\mathbf{Z}_n, \mathbf{W}_m, \mathbf{z}_n, \mathbf{w}_m)\big) \geq 2\xi_n\big(c' - L_0\big\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\big\|_\infty\big)(mr_1 + nr_2), \tag{34}$$

where the original $c = \mu^2_{\min}\kappa_{\min}/16$ has been changed to $c' = \mu^2_{\min} c_{\min}/16$. Next, the set $A(\mathbf{z}^\star_n, \mathbf{w}^\star_m, \mathbf{z}_n, \mathbf{w}_m)$ must be changed so that we consider two-sided deviations between $\delta^{\boldsymbol{\pi}_n}(\mathbf{Z}_n, \mathbf{W}_m, \mathbf{z}_n, \mathbf{w}_m)$ and its conditional expectation of order $\xi_n(c' - L_0\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\|_\infty)(mr_1 + nr_2)$ instead of the previous $(c - L_0\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\|_\infty)(mr_1 + nr_2)$. Equation (22) therefore turns to

$$\begin{aligned} &\mathbb{P}_\star\big(A\big(\mathbf{z}^\star_n, \mathbf{w}^\star_m, \mathbf{z}_n, \mathbf{w}_m\big) \cap \big\{(\mathbf{Z}_n, \mathbf{W}_m) = \big(\mathbf{z}^\star_n, \mathbf{w}^\star_m\big)\big\}\big) \\ &\quad \leq 2\exp\big[-\psi^\star_n\big(c' - 2L_0\eta\big)(mr_1 + nr_2)\big]\boldsymbol{\mu}\big(\mathbf{z}^\star_n, \mathbf{w}^\star_m\big) \\ &\quad \leq 2\exp\big[-\breve{\psi}^\star\big(c' - 2L_0\eta\big)\xi_n(mr_1 + nr_2)\big]\boldsymbol{\mu}\big(\mathbf{z}^\star_n, \mathbf{w}^\star_m\big). \end{aligned}$$

The set Ω_1 is still defined as in Equation (23) and on this set, Inequality (21) and thus both (18) and (19) are still satisfied. However, the upper bound on $\mathbb{P}_\star(\overline{\Omega_1})$ is modified as follows for LBM

$$\begin{aligned}\mathbb{P}_\star(\overline{\Omega}_1) &\leq \mathbb{P}_\star(\overline{\Omega}_0) \\ &\quad + 2|\mathfrak{S}|\big[\big\{1+\exp\big[-m\xi_n\breve{\psi}^\star\big(c'-2L_0\eta\big)\big]\big\}^n\big\{1+\exp\big[-n\xi_n\breve{\psi}^\star\big(c'-2L_0\eta\big)\big]\big\}^m - 1\big].\end{aligned}$$

Combining the latter with the control of the probability of $\overline{\Omega}_0$ given in Proposition 1, we obtain for LBM

$$\varepsilon_{n,m} := \mathbb{P}_\star(\overline{\Omega}_1) \leq 2QL\exp\big[-(n\wedge m)\mu_{\min}^2/2\big] + 2|\mathfrak{S}|d_{n,m}\exp(d_{n,m}),$$

where $d_{n,m} = [n\exp\{-m\xi_n\breve{\psi}^\star(c'-2L_0\eta)\} + m\exp\{-n\xi_n\breve{\psi}^\star(c'-2L_0\eta)\}]$. The condition required to make the $\varepsilon_{n,m}$ summable and conclude the proof is $\log n/(m\xi_n) \to 0$. This condition holds under Assumption 9. Note that for SBM, we get

$$\varepsilon_{n,n} := \mathbb{P}_\star(\overline{\Omega}_1) \leq 2Q\exp\big[-n\alpha_{\min}^2/2\big] + 2|\mathfrak{S}|d_n\exp(d_n),$$

with $d_n = n\exp\{-2n\xi_n\breve{\psi}^\star(c'-2L_0\eta)\}$ and which also satisfies $\sum_n \varepsilon_{n,n} < +\infty$. □

5.2. Weighted models with a vanishing density

We now consider the setup introduced in Section 4.3 as well as corresponding assumptions, except that we shall now assume that the sparsity parameters $\pi_{ql,n} := \xi_n\pi_{ql}$ may be arbitrarily close to zero (see Assumption 9). Note that the parameters $(\gamma_{ql})_{(q,l)\in\mathcal{Q}\times\mathcal{L}} \in \Gamma_{\mathcal{QL}}$ remain fixed. Flynn and Perry [11] adopt a sparse setup where the average entry value goes to 0. Our setup with inflated numbers of 0-valued entry is only a special instance of theirs but is in our opinion a realistic way to model sparse matrices.

In the next lemma, we provide the scaling of $\kappa_{\min}(\boldsymbol{\pi}_n, \gamma)$, or more accurately a lower bound thereof, and show that Assumption 8 is sufficient to guarantee the adequate scaling of the Lipschitz condition. We however need a stronger condition that in Assumption 7 to control deviations in this setup.

Assumption 10. *Fix $(\mathbf{z}_n^\star, \mathbf{w}_m^\star) \in \tilde{\mathcal{U}}_0$ and $(\mathbf{z}_n, \mathbf{w}_m)$ in $\tilde{\mathcal{U}}$ with $(\mathbf{z}_n^\star, \mathbf{w}_m^\star) \neq (\mathbf{z}_n, \mathbf{w}_m)$. Let $\tilde{Y}_{ij} = \log[\tilde{f}(X_{ij}; \gamma_{z_i^\star w_j^\star})/\tilde{f}(X_{ij}; \gamma_{z_i w_j})] + c$, where c is a centering constant. There exists some interval $I \subset (0, +\infty)$ such that the function $\tilde{\psi}^\star$ defined on $(0, +\infty)$ as*

$$\tilde{\psi}^\star(x) = \sup_{\lambda\in I}\Big(\lambda x - \sup_{\boldsymbol{\gamma}\in\Gamma_{\mathcal{QL}}}\max_{(i,j)\in\mathcal{I}}\mathbb{E}_\star^{\mathbf{z}_n^\star\mathbf{w}_m^\star}\big[\exp(\lambda\tilde{Y}_{ij})|X_{ij}\neq 0\big] + 1\Big) \tag{35}$$

exists and is positive on $(0, +\infty)$.

Remark 7. Note that $\tilde{\psi}^\star(x)$ is *not* the rate from the usual Cramer–Chernoff bound, which is in general $\sup_{\lambda>0}(\lambda x - \log\mathbb{E}[\exp(\lambda Z)])$ for a centered random variable Z. The part $\log\mathbb{E}[\exp(\lambda Z)]$

is replaced with the larger quantity $\exp(\lambda Z) - 1$ which induces a slower exponential decrease. The formula arises from a Taylor expansion of rate function $\psi^\star$ for deviations of order $\xi_n x$ to obtain a linear scaling of $\psi^\star(\xi_n x)$ with respect to ξ_n. Note also that the condition that $\tilde{\psi}^\star$ positive in Assumption 10 is stronger than and implies corresponding Assumption 7.

The proof of following lemma is postponed to the Appendix.

Lemma 4. *Fix two parameters $\boldsymbol{\pi}_n = \xi_n \boldsymbol{\pi}$ and $\boldsymbol{\pi}'_n = \xi_n \boldsymbol{\pi}'$ in the set $\Pi_{\mathcal{QL},n}$, where $\boldsymbol{\pi}, \boldsymbol{\pi}' \in \Pi_{\mathcal{QL}}$. Under Assumptions 8, 10 and using the notation of Section 4.3, we have for all n, all $(q,l), (q',l') \in \mathcal{Q} \times \mathcal{L}$ and all $\boldsymbol{\gamma}, \boldsymbol{\gamma}' \in \Gamma_{\mathcal{QL}}$,*

$$\kappa_{\min,n} := \kappa_{\min}\big(\xi_n \boldsymbol{\pi}^\star, \boldsymbol{\gamma}^\star\big) \geq \xi_n \big(c_{\min}\big(\boldsymbol{\pi}^\star\big) + a\kappa_{\min}\big(\boldsymbol{\gamma}^\star\big)\big), \tag{36}$$

$$\left| \int_{\mathcal{X}} \log \frac{f(x; \pi_{ql,n}, \gamma_{ql})}{f(x; \pi'_{ql,n}, \gamma'_{ql})} f(x; \pi_{q'l',n}, \gamma_{q'l'}) \, \mathrm{d}x \right| \leq \xi_n \left(\frac{\|\boldsymbol{\pi} - \boldsymbol{\pi}'\|_\infty}{a} + \tilde{L}_0 \big\| \boldsymbol{\gamma} - \boldsymbol{\gamma}' \big\|_\infty \right), \tag{37}$$

$$\psi_n^\star(x) := \psi^\star(\xi_n x) \geq \frac{\xi_n \mu_{\min}^2}{8} \left(\tilde{\psi}^\star\left(\frac{x}{2}\right) \wedge \breve{\psi}^\star\left(\frac{x}{2}\right) \right), \tag{38}$$

where

$$\kappa_{\min} := \kappa_{\min}\big(\boldsymbol{\gamma}^\star\big) = \min\big\{\tilde{D}\big(\gamma_{ql}^\star \parallel \gamma_{q'l'}^\star\big); (q,l), \big(q',l'\big) \in \mathcal{Q} \times \mathcal{L}, \gamma_{ql}^\star \neq \gamma_{q'l'}^\star\big\} > 0,$$

$$\tilde{D}\big(\gamma \parallel \gamma'\big) := \int_{\mathcal{X}} \log\left(\frac{\tilde{f}(x;\gamma)}{\tilde{f}(x;\gamma')}\right) \tilde{f}(x;\gamma) \, \mathrm{d}x, \qquad \forall \gamma, \gamma' \in \Gamma,$$

$$\begin{aligned} c_{\min} &:= c_{\min}\big(\boldsymbol{\pi}^\star\big) \\ &= \frac{1}{2}\left(\frac{a}{1-a}\right)^2 \min\left\{ \frac{(\pi_{ql}^\star - \pi_{q'l'}^\star)^2}{\pi_{ql}^\star}; (q,l), \big(q',l'\big) \in \mathcal{Q} \times \mathcal{L}, \pi_{ql}^\star \neq \pi_{q'l'}^\star \right\} > 0, \end{aligned}$$

$$\breve{\psi}^\star(x) = \frac{x^2}{8a[\log(1-a) - \log a]^2 + 4x[\log(1-a) - \log a]/3},$$

$$\tilde{\psi}^\star(x) = \sup_{\lambda \in I} \Big(\lambda x - \max_{(i,j) \in \mathcal{I}} \mathbb{E}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star} \big[\exp(\lambda \tilde{Y}_{ij}) | X_{ij} \neq 0 \big] + 1 \Big).$$

Corollary 5. *Under the assumptions from Section 4.3 and Assumption 10 replacing the weaker Assumption 7, Theorem 1 and Corollaries 1 to 3 remain valid with the following modifications*

1. $L_0 = a^{-1} + \tilde{L}_0$;
2. $c = \mu_{\min}^2 (c_{\min} + a\kappa_{\min})/16$;
3. $\boldsymbol{\pi}$ *is replaced by* $(\xi_n \boldsymbol{\pi}, \boldsymbol{\gamma})$;
4. $(c - 2L_0 \|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\|_\infty)$ *is replaced by* $\xi_n (c - 2L_0 \|(\boldsymbol{\pi}, \boldsymbol{\gamma}) - (\boldsymbol{\pi}^\star, \boldsymbol{\gamma}^\star)\|_\infty)$.

Proof. This result is proved following the proof of Theorem 1, exactly in the same way as we did for Corollary 4, with some changes in key quantities as listed in the corollary. □

Appendix: Technical proofs

Proof of Lemma 2. Let us recall that this proof is a generalization of the proof of Proposition B.5 in [6].

Since $(\mathbf{z}_n^\star, \mathbf{w}_m^\star) \in \tilde{\mathcal{U}}^0$, for any $q \in \mathcal{Q}$ and any $l \in \mathcal{L}$, the number of entries in $\mathbf{z}_n^\star$ (resp. in $\mathbf{w}_m^\star$) which take value q (resp. l) is at least $\lceil n\mu_{\min}/2 \rceil$ (resp. $\lceil m\mu_{\min}/2 \rceil$). Up to a reordering of the vectors $\mathbf{z}_n^\star$ and $\mathbf{w}_m^\star$, we may assume that the first $Q\lceil n\mu_{\min}/2 \rceil$ entries of $\mathbf{z}_n^\star$ and the first $L\lceil m\mu_{\min}/2 \rceil$ entries of $\mathbf{w}_m^\star$ are fixed, with

$$\begin{aligned} \mathbf{z}_n^\star &= \big(1, 2, \dots, Q, 1, 2, \dots, Q, \dots, 1, 2, \dots, Q, z^\star_{Q\lceil n\mu_{\min}/2 \rceil+1}, \dots, z_n^\star\big), \\ \mathbf{w}_m^\star &= \big(1, 2, \dots, L, 1, 2, \dots, L, \dots, 1, 2, \dots, L, w^\star_{L\lceil m\mu_{\min}/2 \rceil+1}, \dots, w_m^\star\big). \end{aligned} \tag{40}$$

Such ordering of the entries of $(\mathbf{z}_n^\star, \mathbf{w}_m^\star)$ induces a specific ordering of the entries of $(\mathbf{z}_n, \mathbf{w}_m)$. For each $k \in \{1, \dots, \lceil n\mu_{\min}/2 \rceil\}$ (resp. each $j \in \{1, \dots, \lceil m\mu_{\min}/2 \rceil\}$), we denote by s_k (resp. t_j) the application from $\mathcal{Q}$ to $\mathcal{Q}$ (resp. from $\mathcal{L}$ to $\mathcal{L}$) defined by

$$\forall q \in \mathcal{Q}, \qquad s_k\big(z^\star_{(k-1)Q+q}\big) = z_{(k-1)Q+q} \quad \text{and} \quad \forall l \in \mathcal{L}, \qquad t_j\big(w^\star_{(j-1)L+l}\big) = w_{(j-1)L+l}.$$

In other words, we write $\mathbf{z}_n$ and $\mathbf{w}_m$ in the form

$$\begin{aligned} \mathbf{z}_n = \big(&s_1(1), s_1(2), \dots, s_1(Q), s_2(1), \dots, s_2(Q), \dots, s_{\lceil n\mu_{\min}/2 \rceil}(1), \dots, s_{\lceil n\mu_{\min}/2 \rceil}(Q), \\ &z_{Q\lceil n\mu_{\min}/2 \rceil+1}, \dots, z_n\big) \\ \mathbf{w}_m = \big(&t_1(1), t_1(2), \dots, t_1(L), t_2(1), \dots, t_2(L), \dots, t_{\lceil m\mu_{\min}/2 \rceil}(1), \dots, t_{\lceil m\mu_{\min}/2 \rceil}(L), \\ &w_{L\lceil m\mu_{\min}/2 \rceil+1}, \dots, w_m\big). \end{aligned} \tag{41}$$

There are several possible orderings of $\mathbf{z}_n^\star$ (resp. $\mathbf{w}_m^\star$) in the form (40) and each one induces a different ordering of $\mathbf{z}_n$ (resp. $\mathbf{w}_m$) in the form (41). For example, for any $1 \le k, k' \le \lceil n\mu_{\min}/2 \rceil$ and any $q \in \mathcal{Q}$, we can exchange $z^\star_{(k-1)Q+q}$ and $z^\star_{(k'-1)Q+q}$ which are both equal to q and this induces a permutation between $s_k(q)$ and $s_{k'}(q)$ in $\mathbf{z}_n$. (Similarly for any $1 \le j, j' \le \lceil m\mu_{\min}/2 \rceil$ and any $l \in \mathcal{L}$, we can exchange $t_j(l)$ and $t_{j'}(l)$ in $\mathbf{w}_m$.) Also, for any $i > Q\lceil n\mu_{\min}/2 \rceil$, $z_i^\star$ is equal to some $q \in \mathcal{Q}$ and can be exchanged with $z^\star_{(k-1)Q+q}$ for any $1 \le k \le \lceil n\mu_{\min}/2 \rceil$. This induces a permutation between $s_k(z_i^\star)$ and z_i in $\mathbf{z}_n$. (Similarly, we can exchange $t_j(w_i^\star)$ and w_i in $\mathbf{w}_m$ for any $i > L\lceil m\mu_{\min}/2 \rceil$ and any $1 \le j \le \lceil m\mu_{\min}/2 \rceil$.) Note also that the orderings of $\mathbf{z}_n^\star$ and $\mathbf{w}_m^\star$ are independent. As already said, each s_k (resp. t_j) is a function from $\mathcal{Q}$ to $\mathcal{Q}$ (resp. from $\mathcal{L}$ to $\mathcal{L}$). We can therefore choose orderings of $\mathbf{z}_n^\star$ and $\mathbf{w}_m^\star$ which minimize the number (ranging from 0 to $\lceil n\mu_{\min}/2 \rceil$) of injective functions s as well as the number (ranging from 0 to $\lceil m\mu_{\min}/2 \rceil$) of injective functions t.

For $1 \le k \le \lceil n\mu_{\min}/2 \rceil$ and $1 \le j \le \lceil m\mu_{\min}/2 \rceil$, let

$$B_{kj} = \big|\big\{(q, l) \in \mathcal{Q} \times \mathcal{L}; \pi^\star_{ql} \ne \pi^\star_{s_k(q)t_j(l)}\big\}\big|.$$

We have of course $\operatorname{diff}(\mathbf{z}_n, \mathbf{w}_m, \mathbf{z}_n^\star, \mathbf{w}_m^\star) \ge \sum_{k=1}^{\lceil n\mu_{\min}/2 \rceil} \sum_{j=1}^{\lceil m\mu_{\min}/2 \rceil} B_{kj}$.

The simplest case is obtained when for any (k, j), we have $B_{k,j} \geq 1$ and then

$$\mathrm{diff}\big(\mathbf{z}_n, \mathbf{w}_m, \mathbf{z}_n^\star, \mathbf{w}_m^\star\big) \geq \left\lceil \frac{n\mu_{\min}}{2} \right\rceil \times \left\lceil \frac{m\mu_{\min}}{2} \right\rceil \geq \frac{\mu_{\min}^2}{8}(mr_1 + nr_2),$$

since both $r_1 \leq n$ and $r_2 \leq m$. In this case, the proof is finished.

Otherwise, there is at least one (k, j) such that $B_{kj} = 0$. In this case, we start by proving that at least one application among the $s_{k'}$ and at least one application among the $t_{j'}$ are permutations. Indeed, consider some (k, j) with $B_{kj} = 0$. Assume that $s_k(q) = s_k(q')$ for some $q \neq q'$. Then for all l, we have $\pi^\star_{ql} = \pi^\star_{s_k(q)t_j(l)} = \pi^\star_{s_k(q')t_j(l)} = \pi^\star_{q'l}$ which contradicts Assumption 1. The same holds if $t_j(l) = t_j(l')$ for some $l \neq l'$. Therefore if $B_{kj} = 0$, both s_k and t_j are injections and therefore permutations.

Now, we prove that all applications $s_{k'}$ which are permutations are in fact equal. Indeed, consider $k' \neq k$ such that $s_{k'}$ and s_k are injections. Assume there exists some q such that $s_k(q) \neq s_{k'}(q)$. Then exchanging $s_k(q)$ and $s_{k'}(q)$ in $\mathbf{z}_n$ decreases the number of injective applications s_i by 2, in contradiction with the minimality of the chosen ordering of coordinates in $\mathbf{z}_n^\star$. Therefore, $s_k = s_{k'}$. Thus, all injective $s_{k'}$ are equal to the same permutation $s \in \mathfrak{S}_Q$. Similarly, all injective $t_{j'}$ are equal to the same permutation $t \in \mathfrak{S}_L$. Since one of these pairs of permutations (s_k, t_j) is associated to the event $B_{kj} = 0$, this implies that $(\boldsymbol{\pi}^\star)^{s,t} = \boldsymbol{\pi}^\star$. Note also that according to Assumption 2, we necessarily have $(s, t) \in \mathfrak{S}$.

We now argue that as soon as there is at least one injective application s_k (which is thus equal to s), we must have $z_i = s(z_i^\star)$ for all $i \geq Q\lceil n\mu_{\min}/2\rceil + 1$. Otherwise, we could decrease by one the total number of injective $s_{k'}$ by permuting z_i and $s(z_i^\star)$, which contradicts the minimality of the number of injections. In the same way, if there is at least one injective application t_j (thus equal to t), we have $w_i = t(w_i^\star)$ for any $i \geq L\lceil m\mu_{\min}/2\rceil + 1$.

Let d_1 (resp. d_2) be the number (possibly equal to 0) of non-injective s_k (resp. t_j). It comes from the two previous points that we can in fact write

$$\mathbf{z}_n = \big(s_1(1), \dots, s_1(Q), \dots, s_{d_1}(1), \dots, s_{d_1}(Q), s\big(z^\star_{d_1 Q+1}\big), \dots, s\big(z_n^\star\big)\big),$$
$$\mathbf{w}_m = \big(t_1(1), \dots, t_1(L), \dots, t_{d_2}(1), \dots, t_{d_2}(L), t\big(w^\star_{d_2 L+1}\big), \dots, t\big(w_m^\star\big)\big),$$

where $(s, t) \in \mathfrak{S}$. Thus, we obtain that

$$r_1 = d\big(\mathbf{z}_n, \mathbf{z}_n^\star\big) \leq \big\|\mathbf{z}_n - s\big(\mathbf{z}_n^\star\big)\big\|_0 \leq d_1 Q,$$
$$r_2 = d\big(\mathbf{w}_m, \mathbf{w}_m^\star\big) \leq \big\|\mathbf{w}_m - t\big(\mathbf{w}_m^\star\big)\big\|_0 \leq d_2 L.$$

Finally, for each (k, j) such that either s_k or t_j is non-injective, we have $B_{kj} \geq 1$. Therefore,

$$\begin{aligned}
\mathrm{diff}\big(\mathbf{z}_n, \mathbf{w}_m, \mathbf{z}_n^\star, \mathbf{w}_m^\star\big) &\geq \sum_{k=1}^{\lceil n\mu_{\min}/2\rceil} \sum_{j=1}^{\lceil m\mu_{\min}/2\rceil} B_{kj} \\
&\geq d_1\lceil m\mu_{\min}/2\rceil + d_2\lceil n\mu_{\min}/2\rceil - d_1 d_2 \\
&\geq \frac{d_1\lceil m\mu_{\min}/2\rceil + d_2\lceil n\mu_{\min}/2\rceil}{2}
\end{aligned}$$

$$\geq \frac{r_1\lceil m\mu_{\min}/2\rceil + r_2\lceil n\mu_{\min}/2\rceil}{2Q}$$

$$\geq \frac{\mu_{\min}^2}{4}(mr_1 + nr_2),$$

where the last inequality comes from $\mu_{\min} \leq 1/Q$. This concludes the proof of the lemma. □

Proof of Lemma 3. For any $\pi, \pi' \in \Pi$ and any $\xi \in (0,1)$, the Kullback–Leibler divergence $D(\xi\pi \parallel \xi\pi')$ writes

$$\begin{aligned}D(\xi\pi \parallel \xi\pi') &= \xi\pi \log\frac{\pi}{\pi'} + (1-\xi\pi)\log\left(\frac{1-\xi\pi}{1-\xi\pi'}\right)\\ &= -\xi\pi\log\left(1+\frac{\pi'-\pi}{\pi}\right) - (1-\xi\pi)\log\left(1+\frac{\xi(\pi-\pi')}{1-\xi\pi}\right).\end{aligned}$$

Now, relying on the convexity inequality $\log(1+x) \leq x$ valid for $x > -1$ and on a Taylor series expansion of $\log(1+x)$, there exists some θ with $|\theta| \leq |\pi'-\pi|/\pi$ such that

$$\begin{aligned}D(\xi\pi \parallel \xi\pi') &\geq \xi(\pi-\pi') + \xi\frac{(\pi-\pi')^2}{2\pi}\frac{1}{(1+\theta)^2} - \xi(\pi-\pi')\\ &\geq \xi\frac{(\pi-\pi')^2}{2\pi}\left(\frac{a}{1-a}\right)^2.\end{aligned}$$

Coming back to the definition (12) of $\kappa_{\min}(\boldsymbol{\pi}_n^\star)$ yields

$$\begin{aligned}\kappa_{\min,n} &:= \kappa_{\min}(\boldsymbol{\pi}_n^\star) = \kappa_{\min}(\xi_n\boldsymbol{\pi}^\star)\\ &= \min\{D(\xi_n\pi_{ql}^\star \parallel \xi_n\pi_{q'l'}^\star); (q,l), (q',l') \in \mathcal{Q}\times\mathcal{L}, \pi_{ql}^\star \neq \pi_{q'l'}^\star\}\\ &\geq \xi_n c_{\min}(\boldsymbol{\pi}^\star), \qquad \text{for all } n.\end{aligned}$$

Note that $\kappa_{\min,n}$ scales with ξ_n only when Π is bounded away from 0 and 1. Otherwise a simple bound based on the comparison between Kullback–Leibler divergence and the total variation metric shows that $\kappa_{\min,n}$ scales with ξ_n^2.

A similar scaling can be found to replace Assumption 5. Indeed, for any $\pi, \pi', \pi'' \in \Pi$ and $\xi > 0$, we have in the binary case

$$\left|\int_{\mathcal{X}} \log\frac{f(x;\xi\pi)}{f(x;\xi\pi')} f(x;\xi\pi'')\,\mathrm{d}x\right| = \left|\xi\pi''\log\frac{\pi}{\pi'} + (1-\xi\pi'')\log\left(\frac{1-\xi\pi}{1-\xi\pi'}\right)\right| \leq \frac{\xi|\pi-\pi'|}{a}.$$

Therefore, for any $(q,l), (q',l') \in \mathcal{Q}\times\mathcal{L}$,

$$\left|\int_{\mathcal{X}} \log\frac{f(x;\pi_{ql,n})}{f(x;\pi'_{ql,n})} f(x;\pi_{q'l',n})\,\mathrm{d}x\right| \leq \frac{\xi_n\|\boldsymbol{\pi}-\boldsymbol{\pi}'\|_\infty}{a}.$$

Finally, the scaling for $\psi_n^\star(x)$ in Equation (33) results from Bernstein's inequality as in (31). $\square$

Proof of Lemma 4. For all $\pi, \pi', \pi'' \in \Pi$, $\gamma, \gamma', \gamma'' \in \Gamma$ and $\xi > 0$, we have

$$\begin{aligned}\int_{\mathcal{X}} \log \frac{f(x; \xi\pi, \gamma)}{f(x; \xi\pi', \gamma')} f\big(x; \xi\pi'', \gamma''\big)\,\mathrm{d}x = \xi\pi'' \log \frac{\pi}{\pi'} + \big(1 - \xi\pi''\big) \log \frac{1-\pi}{1-\pi'} \\ + \xi\pi'' \int_{\mathcal{X}} \log \frac{\tilde{f}(x; \gamma)}{\tilde{f}(x; \gamma')} \tilde{f}\big(x; \gamma''\big)\,\mathrm{d}x.\end{aligned} \tag{42}$$

When $(\pi'', \gamma'') = (\pi, \gamma)$, Equation (42) turns to

$$\begin{aligned} D\big((\xi\pi, \gamma) \parallel \big(\xi\pi', \gamma'\big)\big) &= D\big(\xi\pi \parallel \xi\pi'\big) + \xi\pi \tilde{D}\big(\gamma \parallel \gamma'\big) \\ &\geq \xi \frac{(\pi - \pi')^2}{2\pi} \left(\frac{a}{1-a}\right)^2 + \xi a \tilde{D}\big(\gamma \parallel \gamma'\big), \end{aligned}$$

from which we can deduce Inequality (36).

For general (π'', γ''), Equation (42) combined with Inequality (32) (that applies on the Dirac part of the distribution) and Assumption 8 gives

$$\left| \int_{\mathcal{X}} \log \frac{f(x; \xi\pi, \gamma)}{f(x; \xi\pi', \gamma')} f\big(x; \xi\pi'', \gamma''\big)\,\mathrm{d}x \right| \leq \xi \frac{|\pi - \pi'|}{a} + \xi \tilde{L}_0 \big|\gamma - \gamma'\big|,$$

from which we can deduce Inequality (37).

Finally, in this setup, $\breve{\psi}^\star$ arises by using a Bernstein's inequality instead of a Hoeffding's one to control deviations of the Dirac part in the distribution, see (31). Function $\tilde{\psi}^\star$ arises from the control of the deviations of the weighted part of the distribution. Indeed, recall that

$$\tilde{Y}_{ij} = \log\left(\frac{\tilde{f}(X_{ij}; \gamma_{z_i^\star w_j^\star})}{\tilde{f}(X_{ij}; \gamma_{z_i w_j})} \right) + c,$$

where c is a centering constant and let

$$\bar{Y}_{ij} := 1\{X_{ij} \neq 0\} \tilde{Y}_{ij}.$$

We get for $\lambda \in I$,

$$\begin{aligned} \mathbb{E}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\big[\mathrm{e}^{\lambda \bar{Y}_{ij}}\big] &= \mathbb{E}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\big[1\{X_{ij} \neq 0\}\mathrm{e}^{\lambda \tilde{Y}_{ij}} + \big(1 - 1\{X_{ij} \neq 0\}\big)\big] \\ &= 1 + \mathbb{E}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\big[1\{X_{ij} \neq 0\}\big(\mathrm{e}^{\lambda \tilde{Y}_{ij}} - 1\big)\big] \\ &= 1 + \xi_n \pi^\star_{z_i^\star w_j^\star}\big[\mathbb{E}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\big(\mathrm{e}^{\lambda \tilde{Y}_{ij}} | X_{i,j} \neq 0\big) - 1\big], \end{aligned}$$

from which we deduce

$$
\begin{aligned}
\mathbb{P}_{\star}^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\big(|\bar{Y}_{ij}| \geq \xi_n x\big) &\leq \inf_{\lambda \in I} \exp\big[\log\big(\mathbb{E}_{\star}^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\big[e^{\lambda \bar{Y}_{ij}}\big] - \lambda \xi_n x\big)\big] \\
&= \inf_{\lambda \in I} \exp\big[\log\big(1 + \xi_n \pi^\star_{z_i^\star w_j^\star}\big(\mathbb{E}_{\star}^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\big[e^{\lambda \tilde{Y}_{ij}} | X_{i,j} \neq 0\big] - 1\big)\big) - \lambda \xi_n x\big] \\
&\leq \exp\Big[\inf_{\lambda \in I} \xi_n \big(\pi^\star_{z_i^\star w_j^\star}\big(\mathbb{E}_{\star}^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\big[e^{\lambda \tilde{Y}_{ij}} | X_{i,j} \neq 0\big] - 1\big) - \lambda x\big)\Big] \\
&\leq \exp\Big[\xi_n \sup_{\lambda \in I}\Big(\lambda x + 1 - \max_{(i,,j) \in \mathcal{I}} \mathbb{E}_{\star}^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\big[e^{\lambda \tilde{Y}_{ij}} | X_{i,j} \neq 0\big]\Big)\Big],
\end{aligned}
$$

where we used that $\log(1 + x) \leq x$. The last inequality gives us the rate function $\tilde{\psi}^\star$. Equation (38) is constructed from $\tilde{\psi}^\star$ and $\breve{\psi}^\star$ as in Section 4.3. □

References

[1] Airoldi, E., Blei, D., Fienberg, S. and Xing, E. (2008). Mixed-membership stochastic blockmodels. *J. Mach. Learn. Res.* **9** 1981–2014.

[2] Allman, E.S., Matias, C. and Rhodes, J.A. (2009). Identifiability of parameters in latent structure models with many observed variables. *Ann. Statist.* **37** 3099–3132. MR2549554

[3] Allman, E.S., Matias, C. and Rhodes, J.A. (2011). Parameter identifiability in a class of random graph mixture models. *J. Statist. Plann. Inference* **141** 1719–1736. MR2763202

[4] Ambroise, C. and Matias, C. (2012). New consistent and asymptotically normal parameter estimates for random-graph mixture models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **74** 3–35. MR2885838

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

[6] Celisse, A., Daudin, J.-J. and Pierre, L. (2012). Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electron. J. Stat.* **6** 1847–1899. MR2988467

[7] Channarond, A., Daudin, J.-J. and Robin, S. (2012). Classification and estimation in the Stochastic Blockmodel based on the empirical degrees. *Electron. J. Stat.* **6** 2574–2601. MR3020277

[8] Choi, D.S., Wolfe, P.J. and Airoldi, E.M. (2012). Stochastic blockmodels with a growing number of classes. *Biometrika* **99** 273–284. MR2931253

[9] Daudin, J.-J., Picard, F. and Robin, S. (2008). A mixture model for random graphs. *Stat. Comput.* **18** 173–183. MR2390817

[10] DeSarbo, W.S., Fong, D.K.H., Liechty, J. and Saxton, M.K. (2004). A hierarchical Bayesian procedure for two-mode cluster analysis. *Psychometrika* **69** 547–572. MR2272464

[11] Flynn, C. and Perry, P. (2013). Consistent biclustering. Technical report, arXiv:1206.6927.

[12] Frank, O. and Harary, F. (1982). Cluster inference by using transitivity indices in empirical graphs. *J. Amer. Statist. Assoc.* **77** 835–840. MR0686407

[13] Gazal, S., Daudin, J.-J. and Robin, S. (2012). Accuracy of variational estimates for random graph mixture models. *J. Stat. Comput. Simul.* **82** 849–862. MR2929296

[14] Govaert, G. and Nadif, M. (2003). Clustering with block mixture models. *Pattern Recognition* **36** 463–473.

[15] Govaert, G. and Nadif, M. (2008). Block clustering with Bernoulli mixture models: Comparison of different approaches. *Comput. Statist. Data Anal.* **52** 3233–3245. MR2424788

[16] Govaert, G. and Nadif, M. (2010). Latent block model for contingency table. *Comm. Statist. Theory Methods* **39** 416–425. MR2745285

[17] Hartigan, J.A. (1972). Direct clustering of a data matrix. *J. Amer. Statist. Assoc.* **67** 123–129.

[18] Holland, P.W., Laskey, K.B. and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks* **5** 109–137. MR0718088

[19] Latouche, P., Birmelé, E. and Ambroise, C. (2011). Overlapping stochastic block models with application to the French political blogosphere. *Ann. Appl. Stat.* **5** 309–336. MR2810399

[20] Latouche, P., Birmelé, E. and Ambroise, C. (2012). Variational Bayesian inference and complexity control for stochastic block models. *Stat. Model.* **12** 93–115. MR2953099

[21] Mariadassou, M., Robin, S. and Vacher, C. (2010). Uncovering latent structure in valued graphs: A variational approach. *Ann. Appl. Stat.* **4** 715–742. MR2758646

[22] Massart, P. (2007). *Concentration Inequalities and Model Selection. Lecture Notes in Math.* **1896**. Berlin: Springer. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard. MR2319879

[23] Nowicki, K. and Snijders, T.A.B. (2001). Estimation and prediction for stochastic blockstructures. *J. Amer. Statist. Assoc.* **96** 1077–1087. MR1947255

[24] 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** 1–11.

[25] Rohe, K., Chatterjee, S. and Yu, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. MR2893856

[26] Rohe, K. and Yu, B. (2012). Co-clustering for directed graphs: the stochastic co-blockmodel and a spectral algorithm. Technical report, arXiv:1204.2296.

[27] Snijders, T.A.B. and Nowicki, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *J. Classification* **14** 75–100. MR1449742

[28] Wyse, J. and Friel, N. (2012). Block clustering with collapsed latent block models. *Stat. Comput.* **22** 415–428. MR2865026

[29] Zanghi, H., Ambroise, C. and Miele, V. (2008). Fast online graph clustering via Erdős Rényi mixture. *Pattern Recognition* **41** 3592–3599.

[30] Zanghi, H., Picard, F., Miele, V. and Ambroise, C. (2010). Strategies for online inference of model-based clustering in large and growing networks. *Ann. Appl. Stat.* **4** 687–714. MR2758645

Received June 2012 and revised November 2013