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

by

Mahendra Mariadassou, Catherine Matias

Research Report No. 38
June 2012

Statistics for Systems Biology Group
Jouy-en-Josas/Paris/Evry, France
http://www.ssbgroup.fr

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

Mahendra Mariadassou[1] and Catherine Matias[2]

July 23, 2012

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

Abstract

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. 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 from a correct recovering of the groups. We also discuss the validity of these results when the proportion of non-null entries in the data matrix converges to zero.

Keywords and phrases: Block clustering, block modelling, 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 (Govaert and Nadif, 2003).

Stochastic block models (SBMs, Holland et al., 1983) 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. 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 from a correct recovering of the 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 of Hartigan (1972) 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 Govaert and Nadif (2003, 2008) for binary data and Govaert and Nadif (2010) for Poisson random variables. A Bayesian version of the model has been introduced in DeSarbo et al. (2004) 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 Wyse and Friel (2012), who consider 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 (at the 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 Frank and Harary, 1982; Holland et al., 1983; Snijders and Nowicki, 1997; Daudin et al., 2008). 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 (Snijders and Nowicki, 1997; Nowicki and Snijders, 2001), to variational approximations of expectation maximization (EM) algorithm (Daudin et al., 2008; Mariadassou et al., 2010; Picard et al., 2009) or variational Bayes approaches (Latouche et al., 2012), online procedures (Zanghi et al., 2008, 2010) and direct methods (Ambroise and Matias, 2011). 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 (Ambroise and Matias, 2011; Mariadassou et al., 2010).

In each of these procedures, a clustering of the graph nodes is performed according to the groups posterior distribution (at the estimated parameter value). The behaviour of this posterior distribution for binary SBMs is studied in Celisse et al. (2011). These authors establish two different results. The first one (Theorem 3.1 in Celisse et al., 2011) 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 Celisse et al., 2011) 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 (Ambroise and Matias, 2011); see also Gazal et al. (2011) 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.

Next, to conclude with the literature concerning SBMs, we shall mention the works of Bickel and Chen (2009); Choi et al. (2012) and Rohe et al. (2011) on the performances of clustering procedures in random graphs. 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 only concern the case of binary graphs, while we shall deal both with binary and weighted graphs; as well as LBMs. Moreover, the works by Bickel and Chen (2009) and Rohe et al. (2011) are not based on a probabilistic model and only deal with community detection, that is to say finding a set of highly connected nodes, this task being more restrictive than block modeling. We also mention that Choi et al. (2012) and Rohe et al. (2011) both are concerned with an asymptotic setting where the number of groups is allowed to grow as the root of the network size and the average network degree grows at least nearly linearly (Rohe et al., 2011) or poly-logarithmically (Choi et al., 2012) 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 Choi et al. (2012) proposes convergence results in a setup of independent Bernoulli random variables (viewing the latent groups as parameters instead of random variables), while in our context, the observed random variables are not independent.

We also want to outline that many different generalizations allowing for overlapping groups exist, both for LBMs and SBMs. We refer the interested reader to the works of DeSarbo et al. (2004) for LBMs and Airoldi et al. (2008); Latouche et al. (2011) 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 connections. Some technical proofs are postponed to Appendix A.

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

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

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 $\{Z_i\}_{1\leq i\leq n}$ and $\{W_j\}_{1\leq j\leq m}$ 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-indepedent 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 $1 \leq i \leq n$ 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

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

This case corresponds to the observation of a random graph whose adjacency matrix is given by $\{X_{ij}\}_{1\leq i,j\leq n}$. As particular cases, we may also consider graphs with no self-loops in which case $\mathcal{I} = \{1, \dots, n\}^2 \setminus \{(i, i); 1 \leq i \leq n\}$. 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, $\mathcal{I} = \{1 \leq i < j \leq n\}$. In the following, some formulas are given in full generality, and one should take $\beta_l = 1$ for any l to obtain corresponding expressions 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 $\Pi_{\mathcal{QL}}$ whose coordinates π_{ql} belong to some set Π (note that $\Pi_{\mathcal{QL}}$ 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 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:

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

We consider the following parameter set

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

and define $\alpha_{\max} = \max\{\alpha_q; q \in \mathcal{Q}; \theta = (\boldsymbol{\mu}, \boldsymbol{\pi}) \in \Theta\}$ and similarly $\beta_{\max} = \max\{\beta_l; l \in \mathcal{L}; \theta = (\boldsymbol{\mu}, \boldsymbol{\pi}) \in \Theta\}$. We let $\mu_{\min} := \alpha_{\min} \wedge \beta_{\min}$ and $\mu_{\max} := \alpha_{\max} \vee \beta_{\max}$. We denote by $\mathbb{P}_\theta$ and $\mathbb{E}_\theta$ the probability distribution and expectation under parameter 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 Allman et al. (2009) and then later more thoroughly discussed in Allman et al. (2011) for undirected, binary or weighted random graphs. See also Celisse et al. (2011) 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

We introduce a concept that will enable us to deal with possible symmetries in the parameter matrices $\boldsymbol{\pi}$. For instance, the affiliation model (Frank and Harary, 1982; Ambroise and Matias, 2011) is a particular case of SBM where the parameter matrix $\boldsymbol{\pi}$ writes $\boldsymbol{\pi} = (\lambda - \nu) I_Q + \nu \mathbf{1}_Q^\intercal \mathbf{1}_Q$, with I_Q the $Q \times Q$ identity matrix and $\mathbf{1}_Q$ the Q-length vector of 1s. In other words, the model (that is motivated by parsimony) is characterized by only two different types of connections: inner-group connections all happen with the same probability λ, whereas outer-group connections happen with probability ν. In this case, for any permutation s of $\mathcal{Q}$, the permuted matrix $(\pi_{s(q)s(l)})_{1 \leq q,l \leq Q}$ is equal to the original $\boldsymbol{\pi}$. As a consequence, we have

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

where $=^d$ means equality in distribution and $s(\mathbf{Z}_n) := (s(Z_1), \ldots, s(Z_n))$. Thus, a posteriori estimation of the groups distinguishes the different configurations $\{s(\mathbf{Z}_n), s \in \mathfrak{S}_Q\}$ (where we let $\mathfrak{S}_Q$ be the set of permutations of $\mathcal{Q}$), if and only if they happen to have different probabilities of occurrences. In this latter case, a posteriori estimation will select among the set $\{s(\mathbf{Z}_n), s \in \mathfrak{S}_Q\}$ the configuration whose prior probability is higher.

More generally for LBMs or SBMs, the quality of a posteriori estimation of the row and column groups depends on whether there exist some permutations (apart from the trivial identity permutation) that leave the parameter matrix $\boldsymbol{\pi}$ invariant. If this is the case, then for instance a model with equal group proportions will recover with equal probability any of the configurations obtained by permuting the actual one. It should be stressed that this phenomenon is different from the classical label switching issue that arises in finite mixture models. LBMs and SBMs also experience the label switching issue: any permutations on the labels of the rows and columns groups will induce the same distribution on the data matrix but with rows and columns of $\boldsymbol{\pi}$ permuted accordingly. Here, we rather point out the fact that for some constrained models, there might exist permutations on the rows and columns groups that leave the connectivity parameter $\boldsymbol{\pi}$ invariant. As a consequence, when comparing the actual group configuration and its permuted version, a posteriori distribution does not rely on the data anymore. Indeed, the difference between those posterior probabilities is equal to the difference between their prior probabilities.

We let $\mathfrak{S}_Q$ and $\mathfrak{S}_L$ be the sets of permutations of $\mathcal{Q}$ and $\mathcal{L}$ respectively. For any $(s,t) \in \mathfrak{S}_Q \times \mathfrak{S}_L$, we let

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

Fix a subgroup $\mathfrak{S}$ of $\mathfrak{S}_Q \times \mathfrak{S}_L$ 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 $\mathfrak{S}$ of $\mathfrak{S}_Q \times \mathfrak{S}_L$. This includes the case where $\mathfrak{S}$ is reduced to identity pair and the parameter set is the unconstrained product set Π^{QL}. We will moreover exclude from the parameter set $\Pi_{\mathcal{QL}}$ any point $\boldsymbol{\pi}$ admitting specific symmetries, namely such that there exists some pair $(s,t) \in (\mathfrak{S}_Q \times \mathfrak{S}_L)\setminus\mathfrak{S}$ satisfying $\boldsymbol{\pi}^{s,t} = \boldsymbol{\pi}$. 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 $\mathfrak{S}_Q \times \mathfrak{S}_L$. Moreover, for any pair of permutations $(s,t) \in (\mathfrak{S}_Q \times \mathfrak{S}_L) \setminus \mathfrak{S}$ 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 $\mathfrak{S} = \{(Id, Id)\}$ and let*

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

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

Whenever $\mathfrak{S}$ is not reduced to the identity singleton pair, each parameter value $\boldsymbol{\pi} \in \Pi_{\mathcal{QL}}$ induces many different *equivalent* configurations. More precisely, for any $(s,t) \in \mathfrak{S}$ and any $\boldsymbol{\pi} \in \Pi_{\mathcal{QL}}$, we have

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

which means that the difference between the posterior distributions $\mathbb{P}_\theta(\{\mathbf{Z}_n, \mathbf{W}_m\}|\mathbf{X}_{n,m}) - \mathbb{P}_\theta(\{s(\mathbf{Z}_n), t(\mathbf{W}_m)\}|\mathbf{X}_{n,m})$ does not depend on the data $\mathbf{X}_{n,m}$.

Remark 1. *As already said, in SBM with affiliation structure, the group of permutations (s,s) with $s \in \mathfrak{S}_Q$ leaves the parameter set $\Pi_{\mathcal{QL}}$ invariant. For more general models, let us consider $(s,t) = ([q,q'],[l,l']) \in \mathfrak{S}_Q \times \mathfrak{S}_L$ where $[q,q']$ is the transposition of q and q' in $\mathcal{Q}$ and $[l,l']$ is the transposition of l and l' in $\mathcal{L}$. Then any $\boldsymbol{\pi} \in \Pi_{\mathcal{QL}}$ satisfies*

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

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 both $\pi_{ql} \neq \pi_{ql'}$ and $\pi_{q'l} \neq \pi_{q'l'}$.

Note that the parameter sets $\Pi_{\mathcal{QL}}$ that we consider are then in a one-to-one correspondence with the subgroups $\mathfrak{S}$ of $\mathfrak{S}_Q \times \mathfrak{S}_L$. Note also that we have $|\mathfrak{S}| \leq Q!L!$ in general and $|\mathfrak{S}| \leq Q!$ in the particular 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 $\mathfrak{S}_Q \times \mathfrak{S}_L$ and fix a parameter value $\boldsymbol{\pi} \in \Pi_{\mathcal{QL}}$. Any two groups configurations $(\mathbf{z}_n, \mathbf{w}_m) := (z_1, \dots, z_n, w_1, \dots, w_m)$ and $(\mathbf{z'}_n, \mathbf{w'}_m) := (z'_1, \dots, z'_n, w'_1, \dots, w'_m)$ 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*

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

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

$$d((\mathbf{z}_n, \mathbf{w}_m), (\mathbf{z'}_n, \mathbf{w'}_m)) := \min\{\|\mathbf{z}_n - s(\mathbf{z'}_n)\|_0 + \|\mathbf{w}_m - t(\mathbf{w'}_m)\|_0; (s,t) \in \mathfrak{S}\}. \quad (2)$$

Note that this distance is well-defined on the space $\tilde{\mathcal{U}}$. Note also that when $\mathfrak{S}$ is reduced to the identity pair, the distance $d(,)$ is an ordinary ℓ_0 distance.

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

$$A_q = \Big\{\omega \in \Omega; N_q(\mathbf{Z}_n(\omega)) := \sum_{i=1}^n 1\{Z_i(\omega) = q\} < n\mu_{\min}/2\Big\},$$
$$\text{and} \quad B_l = \Big\{\omega \in \Omega; N_l(\mathbf{W}_m(\omega)) := \sum_{j=1}^m 1\{W_j(\omega) = l\} < m\mu_{\min}/2\Big\}.$$

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

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

Taking an union bound, we obtain $\mathbb{P}_\star(\cup_{(q,l)\in\mathcal{Q}\times\mathcal{L}}(A_q \cup B_l)) \leq 2QL\exp[-(n \wedge m)\mu_{\min}^2/2]$. Now, consider the event Ω_0 defined by

$$\begin{aligned}\Omega_0 &:= \{\omega \in \Omega; \forall (q,l) \in \mathcal{Q}\times\mathcal{L}, N_q(\mathbf{Z}_n(\omega)) \geq n\mu_{\min}/2 \text{ and } N_l(\mathbf{W}_m(\omega)) \geq m\mu_{\min}/2\} \\ &= \cap_{(q,l)\in\mathcal{Q}\times\mathcal{L}}(\bar{A}_q \cap \bar{B}_l), \end{aligned} \quad (3)$$

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

$$\mathcal{U}^0 = \{(\mathbf{z}_n, \mathbf{w}_m) \in \mathcal{U}; \forall (q,l) \in \mathcal{Q}\times\mathcal{L}, N_q(\mathbf{z}_n) \geq n\mu_{\min}/2 \text{ and } N_l(\mathbf{w}_m) \geq m\mu_{\min}/2\}, \quad (4)$$

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 $(\mathbf{z}_n, \mathbf{w}_m) \in \mathcal{U}$,*

$$\begin{aligned} p^{\theta}_{n,m}(\mathbf{z}_n, \mathbf{w}_m) &:= \mathbb{P}_\theta((\mathbf{Z}_n, \mathbf{W}_m) = (\mathbf{z}_n, \mathbf{w}_m) | \mathbf{X}_{n,m}) \\ &\propto \Big(\prod_{(i,j)\in\mathcal{I}} f(X_{ij}; \pi_{z_i w_j}) \Big) \Big(\prod_{i=1}^{n} \alpha_{z_i} \Big) \Big(\prod_{j=1}^{m} \beta_{w_j} \Big), \end{aligned} \tag{5}$$

where $\propto$ means equality up to a normalizing constant and where we let $\beta_l = 1$ in SBM.

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

$$\forall (\mathbf{z}^\star_n, \mathbf{w}^\star_m), (\mathbf{z}_n, \mathbf{w}_m) \in \tilde{\mathcal{U}}, \quad \delta^{\boldsymbol{\pi}}(\mathbf{z}^\star_n, \mathbf{w}^\star_m, \mathbf{z}_n, \mathbf{w}_m) = \sum_{(i,j)\in\mathcal{I}} \log \left(\frac{f(X_{ij}; \pi_{z^\star_i w^\star_j})}{f(X_{ij}; \pi_{z_i w_j})} \right). \tag{6}$$

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

$$\Delta^{\boldsymbol{\pi}}(\mathbf{z}^\star_n, \mathbf{w}^\star_m, \mathbf{z}_n, \mathbf{w}_m) = \sum_{(i,j)\in\mathcal{I}} \mathbb{E}_\star \left(\log \Big(\frac{f(X_{ij}; \pi_{z^\star_i w^\star_j})}{f(X_{ij}; \pi_{z_i w_j})} \Big) \Big| (\mathbf{Z}_n, \mathbf{W}_m) = (\mathbf{z}^\star_n, \mathbf{w}^\star_m) \right). \tag{7}$$

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 $\mathbb{P}^{\mathbf{z}^\star_n \mathbf{w}^\star_m}_\star$ and $\mathbb{E}^{\mathbf{z}^\star_n \mathbf{w}^\star_m}_\star$, 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 $\mathbb{P}^{\mathbf{z}^\star_n \mathbf{w}^\star_m}_\star$, the random variables $\{X_{ij}; (i,j) \in \mathcal{I}\}$ are independent.

Assumption 3. *(Concentration inequality). Fix* $(\mathbf{z}_n^\star, \mathbf{w}_m^\star) \in \tilde{\mathcal{U}}^0$ *and* $(\mathbf{z}_n, \mathbf{w}_m) \in \tilde{\mathcal{U}}$ *such that* $(\mathbf{z}_n, \mathbf{w}_m) \not\sim (\mathbf{z}_n^\star, \mathbf{w}_m^\star)$*. There exists some positive function* $\psi^\star : (0, +\infty) \to (0; +\infty]$ *such that for any* $\varepsilon > 0$*, we have*

$$\mathbb{P}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\left(\left|\delta^{\boldsymbol{\pi}}(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m) - \mathbb{E}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\left(\delta^{\boldsymbol{\pi}}(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m)\right)\right| \geq \varepsilon(mr_1 + nr_2)\right) \leq 2\exp[-\psi^\star(\varepsilon)(mr_1 + nr_2)], \quad (8)$$

where the distance $d((\mathbf{z}_n^\star, \mathbf{w}_m^\star), (\mathbf{z}_n, \mathbf{w}_m))$ *defined by* (2) *is attained for some pair of permutations* $(s, t) \in \mathfrak{S}$ *and we set* $r_1 := \|\mathbf{z}_n^\star - s(\mathbf{z}_n)\|_0$ *and* $r_2 := \|\mathbf{w}_m^\star - t(\mathbf{w}_m)\|_0$.

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

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

uniformly in $\pi, \pi' \in \Pi$*. As shown in Section 4, as soon as*

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

is finite for λ *in a small open interval* $I \subset \mathbb{R}$ *around* 0*, a Cramer-Chernoff bound shows that Equation (8) is satisfied with*

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

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

$$D(\pi\|\pi') := \int_{\mathcal{X}} \log\left(\frac{f(x;\pi)}{f(x;\pi')}\right) f(x;\pi)dx. \quad (9)$$

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

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

Note that $\kappa_{\max} < +\infty$ is automatically satisfied when the distributions in the family $\{f(\cdot;\pi);\ \pi \in \Pi\}$ have same support. In particular, this is not the case for Bernoulli distributions when we authorize some probabilities π to be 0 or 1. In the following, we thus exclude the possibility that classes may be almost never or almost surely connected. We also introduce

$$\kappa_{\min} = \kappa_{\min}(\boldsymbol{\pi}^\star) := \min\{D(\pi_{ql}^\star\|\pi_{q'l'}^\star); (q,l), (q',l') \in \mathcal{Q}\times\mathcal{L}, \pi_{ql}^\star \neq \pi_{q'l'}^\star\} > 0, \quad (10)$$

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.

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*

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

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 (7) 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 sequence* $(\mathbf{z}_n, \mathbf{w}_m) \in \tilde{\mathcal{U}}$*, we have* $\mathbb{P}_\star$*-almost surely*

$$\mathbb{E}_\star^{\mathbf{Z}_n\mathbf{W}_m}\Big(\delta^{\boldsymbol{\pi}}(\mathbf{Z}_n, \mathbf{W}_m, \mathbf{z}_n, \mathbf{w}_m)\Big) \leq \frac{C}{2}(mr_1 + nr_2), \tag{11}$$

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 $c = \mu_{\min}^2\kappa_{\min}/16 \in (0, C/4)$ *is such that on the set* Ω_0 *defined by* (3) *whose* $\mathbb{P}_\star$*-probability satisfies* $\mathbb{P}_\star(\Omega_0) \geq 1 - 2QL \times \exp[-(n \wedge m)\mu_{\min}^2/2]$*, 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*

$$\mathbb{E}_\star^{\mathbf{Z}_n\mathbf{W}_m}\Big(\delta^{\boldsymbol{\pi}}(\mathbf{Z}_n, \mathbf{W}_m, \mathbf{z}_n, \mathbf{w}_m)\Big) \geq 2(c - L_0\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\|_\infty)(mr_1 + nr_2). \tag{12}$$

Proof. Note that

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

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 (w.l.o.g.), we can assume that $\mathbf{z}_n^\star, \mathbf{z}_n$ (respectively $\mathbf{w}_m^\star, \mathbf{w}_m$) differ at the first r_1 (resp. r_2) indexes.

First, let us note that

$$\mathbb{E}_\star^{\mathbf{z}_n^\star\mathbf{w}_m^\star}\Big(\delta^{\boldsymbol{\pi}}(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m)\Big) = \sum_{(i,j) \in \tilde{\mathcal{I}}} \int_{\mathcal{X}} \log\Big(\frac{f(x; \pi_{z_i^\star w_j^\star})}{f(x; \pi_{z_i w_j})}\Big) f(x; \pi^\star_{z_i^\star w_j^\star}) dx, \tag{13}$$

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

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

with $C = 2\kappa_{\max}$, which establishes Inequality (11).

To prove Inequality (12), we write the decomposition

$$\sum_{(i,j) \in \tilde{\mathcal{I}}} \int_{\mathcal{X}} \log\Big(\frac{f(x; \pi_{z_i^\star w_j^\star})}{f(x; \pi_{z_i w_j})}\Big) f(x; \pi^\star_{z_i^\star w_j^\star}) dx = \sum_{(i,j) \in \tilde{\mathcal{I}}} \Big\{ - D(\pi^\star_{z_i^\star w_j^\star} \| \pi_{z_i^\star w_j^\star}) + D(\pi^\star_{z_i^\star w_j^\star} \| \pi^\star_{z_i w_j}) + \int_{\mathcal{X}} \log\frac{f(x; \pi^\star_{z_i w_j})}{f(x; \pi_{z_i w_j})} f(x; \pi^\star_{z_i^\star w_j^\star}) dx \Big\}. \tag{14}$$

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

$$0 < \sum_{(i,j)\in\tilde{\mathcal{I}}} D(\pi^\star_{z_i^\star w_j^\star}\|\pi_{z_i^\star w_j^\star}) = \sum_{(i,j)\in\tilde{\mathcal{I}}} \int_{\mathcal{X}} \log \frac{f(x;\pi^\star_{z_i^\star w_j^\star})}{f(x;\pi_{z_i^\star w_j^\star})} f(x;\pi^\star_{z_i^\star w_j^\star})dx$$
$$\leq L_0\|\boldsymbol{\pi}-\boldsymbol{\pi}^\star\|_\infty(mr_1+nr_2-r_1r_2),$$

where the second inequality is another application of Assumption 5.

The central term appearing in the right-hand side of decomposition (14) is handled relying on the next lemma, whose proof is postponed to Appendix A. It is a generalization to LBM of Proposition B.5 in Celisse et al. (2011) 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 (4). 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*

$$\textit{diff}(\mathbf{z}_n,\mathbf{w}_m,\mathbf{z}_n^\star,\mathbf{w}_m^\star) := \left|\{(i,j)\in\mathcal{I};\pi^\star_{z_i w_j}\neq\pi^\star_{z_i^\star w_j^\star}\}\right| \geq \frac{\mu_{\min}^2}{8}(mr_1+nr_2), \tag{15}$$

where the distance $d((\mathbf{z}_n,\mathbf{w}_m),(\mathbf{z}_n^\star,\mathbf{w}_m^\star))$ *is attained for some pair of 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 $D(\pi^\star_{z_i^\star w_j^\star}\|\pi^\star_{z_i w_j})$ is at least $\kappa_{\min}$. We thus get

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

Coming back to (14) and (13), we obtain

$$\sum_{(i,j)\in\tilde{\mathcal{I}}} \int_{\mathcal{X}} \log\Big(\frac{f(x;\pi_{z_i^\star w_j^\star})}{f(x;\pi_{z_i w_j})}\Big) f(x;\pi^\star_{z_i^\star w_j^\star})dx \geq \Big(\frac{\mu_{\min}^2\kappa_{\min}}{8}-2L_0\|\boldsymbol{\pi}-\boldsymbol{\pi}^\star\|_\infty\Big)(mr_1+nr_2)$$

and thus conclude

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

By letting $c=\mu_{\min}^2\kappa_{\min}/16$ we obtain exactly (12). We moreover remark that $2c<C/2$. □

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. 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 $n \to +\infty$.

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 $(s, t) \in \mathfrak{S}$*

$$
\begin{aligned}
&(c - 2L_0\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\|_\infty)(mr_1 + nr_2) - K(\|s(\mathbf{Z}_n) - \mathbf{z}_n\|_0 + \|t(\mathbf{W}_m) - \mathbf{w}_m\|_0) \\
&\leq \log \frac{p^\theta_{n,m}(s(\mathbf{Z}_n), t(\mathbf{W}_m))}{p^\theta_{n,m}(\mathbf{z}_n, \mathbf{w}_m)} \leq C(mr_1 + nr_2) + K(\|s(\mathbf{Z}_n) - \mathbf{z}_n\|_0 + \|t(\mathbf{W}_m) - \mathbf{w}_m\|_0), \quad (16)
\end{aligned}
$$

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 invariant $(\tilde{s}, \tilde{t}) \in \mathfrak{S}$ 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. Inequality (16) provides 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 $\tilde{\mathcal{U}}$*. Its most important part is its left-hand side 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 $d(,)$ given by (2), on the set of actual configurations (so that $d(,)$ is linked with the parameter $\boldsymbol{\pi}$ and its symmetries). When the subgroup $\mathfrak{S}$ is reduced to the identity pair (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 (16) 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 $d(,)$ 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 the left-hand side of (16), 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 condition, 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 $\theta^{\star}$.

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 $\|\boldsymbol{\pi} - \boldsymbol{\pi}^{\star}\|_{\infty} \leq \eta$, we have*

$$p^{\theta}_{n,m}(\mathbf{Z}_n, \mathbf{W}_m) \geq 1 - a_{n,m} \exp(a_{n,m}) \quad \text{and} \quad p^{\theta}_{n,m}(\mathbf{Z}_n, \mathbf{W}_m) \leq (1 + b_{n,m} e^{b_{n,m}})^{-1}, \tag{17}$$

where $a_{n,m} = (ne^{-(c-2L_0\eta)m+K} + me^{-(c-2L_0\eta)n+K})$ and $b_{n,m} = (ne^{-Cm-K} + me^{-Cn-K})$ both 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

$$(\widehat{\mathbf{Z}_n}, \widehat{\mathbf{W}_m}) := \underset{(\mathbf{z}_n, \mathbf{w}_m)}{\mathrm{argmax}}\, p^{\hat{\theta}}_{n,m}(\mathbf{z}_n, \mathbf{w}_m), \quad \text{where } \hat{\theta} = (\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\pi}}) \text{ and } \hat{\boldsymbol{\pi}} \to \boldsymbol{\pi}^{\star},$$

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

$$\sum_{i=1}^{n} 1\{\hat{Z}_i \neq Z_i\} + \sum_{j=1}^{m} 1\{\hat{W}_j \neq W_j\} \text{ for LBMs and } \sum_{i=1}^{n} 1\{\hat{Z}_i \neq Z_i\} \text{ for SBMs},$$

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 $\|\boldsymbol{\pi} - \boldsymbol{\pi}^{\star}\|_{\infty} \leq \eta$, we have*

$$\begin{aligned} p^{\theta}_{n,m}\left(\{(\mathbf{z}_n, \mathbf{w}_m) \in \mathcal{U}; (\mathbf{z}_n, \mathbf{w}_m) \sim (\mathbf{Z}_n, \mathbf{W}_m)\}\right) &\geq 1 - |\mathfrak{S}| a_{n,m} e^{a_{n,m}} \qquad (18)\\ \text{and} \quad p^{\theta}_{n,m}\left(\{(\mathbf{z}_n, \mathbf{w}_m) \in \mathcal{U}; (\mathbf{z}_n, \mathbf{w}_m) \sim (\mathbf{Z}_n, \mathbf{W}_m)\}\right) &\leq (1 + |\mathfrak{S}| b_{n,m} e^{b_{n,m}})^{-1}, \end{aligned}$$

where $a_{n,m} = (ne^{-m(c-2L_0\eta)} + me^{-n(c-2L_0\eta)})$ and $b_{n,m} = (ne^{-mC} + me^{-nC})$ both converge to 0 as $n \to +\infty$. Moreover

$$p^{\theta}_{n,m}\left(\mathbf{Z}_n, \mathbf{W}_m\right) = \frac{1}{|\mathfrak{S}|} p^{\theta}_{n,m}\left(\{(\mathbf{z}_n, \mathbf{w}_m) \in \mathcal{U}; (\mathbf{z}_n, \mathbf{w}_m) \sim (\mathbf{Z}_n, \mathbf{W}_m)\}\right). \tag{19}$$

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 $\|\boldsymbol{\pi} - \boldsymbol{\pi}^{\star}\|_{\infty} \leq \eta$, we have*

$$\begin{aligned} p^{\theta}_{n,m}\left(\{(\mathbf{z}_n, \mathbf{w}_m) \in \mathcal{U}; (\mathbf{z}_n, \mathbf{w}_m) \sim (\mathbf{Z}_n, \mathbf{W}_m)\}\right) &\geq 1 - |\mathfrak{S}| a_{n,m} e^{a_{n,m}} \qquad (20)\\ \text{and} \quad p^{\theta}_{n,m}\left(\{(\mathbf{z}_n, \mathbf{w}_m) \in \mathcal{U}; (\mathbf{z}_n, \mathbf{w}_m) \sim (\mathbf{Z}_n, \mathbf{W}_m)\}\right) &\leq (1 + |\mathfrak{S}| b_{n,m} e^{b_{n,m}})^{-1}, \end{aligned}$$

where $a_{n,m} = (ne^{-m(c-2L_0\eta)+K} + me^{-n(c-2L_0\eta)+K})$ and $b_{n,m} = (ne^{-mC-K} + me^{-nC-K})$ both converge to 0 as $n \to +\infty$.

Remark 3. *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. When $\mathfrak{S} \neq \{(Id, Id)\}$ and $K > 0$, 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

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

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.

Proof of Theorem 1. We shall exhibit the set Ω_1 on which Inequality (16) is satisfied. First note that we have

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

Thus, by letting $K = \log(\alpha_{\max}/\alpha_{\min}) \vee \log(\beta_{\max}/\beta_{\min})$, Inequality (16) is satisfied as soon as we have

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

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 $(\mathbf{z}^\star_n, \mathbf{w}^\star_m)$ be a fixed configuration in $\tilde{\mathcal{U}}$, consider $(\mathbf{z}_n, \mathbf{w}_m) \in \tilde{\mathcal{U}}$. Whenever $(\mathbf{z}_n, \mathbf{w}_m) \sim (\mathbf{z}^\star_n, \mathbf{w}^\star_m)$, 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 $(\mathbf{z}_n, \mathbf{w}_m) \neq (\mathbf{z}^\star_n, \mathbf{w}^\star_m)$ and let $r_1 := \|\mathbf{z}^\star_n - \tilde{s}(\mathbf{z}_n)\|_0$ and $r_2 := \|\mathbf{w}^\star_m - \tilde{t}(\mathbf{w}_m)\|_0$, where $(\tilde{s}, \tilde{t}) \in \mathfrak{S}$ realizes the distance $d((\mathbf{z}^\star_n, \mathbf{w}^\star_m), (\mathbf{z}_n, \mathbf{w}_m))$. We consider the event

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

where the constants $c, C > 0$ have been previously introduced in Proposition 1 and satisfy $0 < 2c < C/2$. 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 (4) 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

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

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

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

In other words,

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

Combining this sets' inclusions with Assumption 3 yields

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

We now consider the set Ω_1 defined by

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

On the set Ω_1, Inequality (21) and thus Inequality(16) are both satisfied. We let

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

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

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

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

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

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

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

Now, we use the bound

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

which leads to

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

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

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

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

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

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

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

satisfies $\sum_n \varepsilon_{n,m_n} < +\infty$. 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. First note that $\Omega_1 = \cup_{(\mathbf{z}_n^\star, \mathbf{w}_m^\star)\in\mathcal{U}^0}(\Omega_1 \cap \{(\mathbf{Z}_n, \mathbf{W}_m) = (\mathbf{z}_n^\star, \mathbf{w}_m^\star)\})$. 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

$$\begin{aligned} 1 - p_{n,m}^\theta\left(\{(\mathbf{Z}_n, \mathbf{W}_m)\}\right) &\le \frac{1 - p_{n,m}^\theta\left(\{(\mathbf{Z}_n, \mathbf{W}_m)\}\right)}{p_{n,m}^\theta\left(\{(\mathbf{Z}_n, \mathbf{W}_m)\}\right)} \\ &= \sum_{\substack{(\mathbf{z}_n,\mathbf{w}_m)\in\mathcal{U} \\ (\mathbf{z}_n,\mathbf{w}_m)\not\sim(\mathbf{z}_n^\star,\mathbf{w}_m^\star)}} \exp\Big(- \log \frac{p_{n,m}^\theta\left(\{(\mathbf{z}_n^\star, \mathbf{w}_m^\star)\}\right)}{p_{n,m}^\theta(\mathbf{z}_n, \mathbf{w}_m)}\Big), \end{aligned}$$

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 $(\mathbf{z}_n, \mathbf{w}_m) \not\sim (\mathbf{z}_n^\star, \mathbf{w}_m^\star)$. 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 the left-hand side of Inequality (16) and $\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\|_\infty \le \eta$, we get

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

and therefore

$$1 - p_{n,m}^\theta\left(\{(\mathbf{Z}_n, \mathbf{W}_m)\}\right) \le \sum_{\substack{(\mathbf{z}_n,\mathbf{w}_m)\in\mathcal{U} \\ (\mathbf{z}_n,\mathbf{w}_m)\not\sim(\mathbf{z}_n^\star,\mathbf{w}_m^\star)}} \exp[-(c - 2L_0\eta)(mr_1 + nr_2) + K(r_1 + r_2)]. \quad (27)$$

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

$$1 - p_{n,m}^\theta(\mathbf{Z}_n, \mathbf{W}_m) \le \sum_{\substack{(\mathbf{z}_n,\mathbf{w}_m)\in\mathcal{U} \\ (\mathbf{z}_n,\mathbf{w}_m)\not\sim(\mathbf{z}_n^\star,\mathbf{w}_m^\star)}} \exp[-(c - 2L_0\eta)(mr_1 + nr_2) + K(r_1 + r_2)].$$

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

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

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

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

where $a_{n,m} = (ne^{-(c-2L_0\eta)m+K} + me^{-(c-2L_0\eta)n+K})$.

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

$$1 - p_{n,m}^\theta\left(\{(\mathbf{Z}_n, \mathbf{W}_m)\}\right) \le |\mathfrak{S}| a_{n,m} \exp(a_{n,m}),$$

where $a_{n,m} = (ne^{-(c-2L_0\eta)m} + me^{-(c-2L_0\eta)n})$. Moreover when $K = 0$, we have $\alpha_1 = \cdots = \alpha_Q$ and $\beta_1 = \cdots = \beta_L$ and it easy to check that

$$p^\theta_{n,m}(\mathbf{Z}_n, \mathbf{W}_m) = p^\theta_{n,m}(s(\mathbf{Z}_n), t(\mathbf{W}_m))$$

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

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

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

$$1 - p^\theta_{n,m}\left(\{(\mathbf{z}_n, \mathbf{w}_m) \sim (\mathbf{Z}_n, \mathbf{W}_m)\}\right) \leq |\mathfrak{S}| a_{n,m} \exp(a_{n,m}),$$

where $a_{n,m} = n\exp(-mc_1 + K) + m\exp(-nc_1 + K)$.

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 $(\mathbf{z}^\star_n, \mathbf{w}^\star_m)$ in $\mathcal{U}^0$. On the set $\Omega_1 \cap \{(\mathbf{Z}_n, \mathbf{W}_m) = (\mathbf{z}^\star_n, \mathbf{w}^\star_m)\}$, we have

$$\frac{1}{p^\theta_{n,m}\left(\{(\mathbf{Z}_n, \mathbf{W}_m)\}\right)} = 1 + \sum_{(\mathbf{z}_n,\mathbf{w}_m)\not\sim(\mathbf{Z}_n,\mathbf{W}_m)} \exp\Big(-\log\frac{p^\theta_{n,m}\left(\{(\mathbf{Z}_n, \mathbf{W}_m)\}\right)}{p^\theta_{n,m}(\mathbf{z}_n, \mathbf{w}_m)}\Big)$$

and relying on the right-hand side of Inequality (16), we get

$$\begin{aligned}
p^\theta_{n,m}\left(\{(\mathbf{z}^\star_n, \mathbf{w}^\star_m)\}\right) &\leq \Big\{1 + \sum_{(\mathbf{z}_n,\mathbf{w}_m)\not\sim(\mathbf{z}^\star_n,\mathbf{w}^\star_m)} \exp\Big(-\log\frac{p^\theta_{n,m}\left(\{(\mathbf{z}^\star_n, \mathbf{w}^\star_m)\}\right)}{p^\theta_{n,m}(\mathbf{z}_n, \mathbf{w}_m)}\Big)\Big\}^{-1} \\
&\leq \Big\{1 + \sum_{(\mathbf{z}_n,\mathbf{w}_m)\not\sim(\mathbf{z}^\star_n,\mathbf{w}^\star_m)} \exp\Big(-C(mr_1+nr_2) - K(r_1+r_2)\Big)\Big\}^{-1}.
\end{aligned}$$

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). The general scheme of proof is first presented, different setups are then explicitly explored.

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 (7), we write

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

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 $D := \operatorname{diff}(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m)$ such non null variables and since $D \leq mr_1 + nr_2 - r_1r_2 \leq mr_1 + nr_2$, we may write

$$\mathbb{P}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\left(|(\delta^{\boldsymbol{\pi}} - \Delta^{\boldsymbol{\pi}})(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m)| \geq \varepsilon(mr_1 + nr_2)\right) \leq \mathbb{P}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\left(\Big|\sum_{(i,j)\in\mathcal{I}} Y_{ij}\Big| \geq \varepsilon D\right). \tag{28}$$

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 $\psi^\star_{\max}$ such that for any $\epsilon > 0$,

$$\mathbb{P}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\left(\Big|\sum_{(i,j)\in\mathcal{I}} Y_{ij}\Big| \geq \varepsilon D\right) \leq 2\exp\{-\psi^\star_{\max}(\varepsilon) D\}, \tag{29}$$

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

$$\begin{aligned}\mathbb{P}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\left(\Big|\sum_{(i,j)\in\mathcal{I}} Y_{ij}\Big| \geq \varepsilon(mr_1 + nr_2)\right) &\leq 2\exp\Big\{-\psi^\star_{\max}(\varepsilon)\mu^2_{\min}(mr_1 + nr_2)/8\Big\} \\ &:= 2\exp\Big\{-\psi^\star(\varepsilon)(mr_1 + nr_2)\Big\},\end{aligned}$$

with $\psi^\star(\cdot) = \psi^\star_{\max}(\cdot)\mu^2_{\min}/8$. 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 $x > 0$,

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

As soon as we can uniformly bound this quantity, namely if we can write

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

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

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

leading to

$$\mathbb{P}_{\star}^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\left(\Big| \sum_{(i,j)\in\mathcal{I}} Y_{ij} \Big| \geq \varepsilon D \right) \leq 2\exp\{-D \sup_{\lambda\in I}(\lambda\epsilon - \psi_{\max}(\lambda))\} \leq 2\exp\Big\{-D\psi^\star_{\max}(\epsilon)\Big\},$$

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

4.2 Binary observations

In this section, we assume that $X_{ij} \in \{0,1\}$ and $f(\cdot;\pi)$ is a Bernoulli distribution with parameter π. In this case, point $i)$ in Assumption 1 is automatically satisfied.

We first state a condition on the parameters $\pi \in \Pi$ that ensures both that Assumptions 3 and 4 are satisfied. Note that this constraint, although rather general, does not cover the cases where some probabilities π_{ql} may either be 0 or 1.

Assumption 7. *The parameter set Π is included in $[a, 1-a]$ for some $a \in (0, 1/2)$.*

Lemma 3. *Under Assumption 7, we obtain that Assumption 3 is satisfied with $\psi^\star(x) = x^2\mu^2_{\min}/[16(\log(1-a) - \log a)^2]$.*

Proof. According to Equation (5) and Assumption 7 ensuring $\pi_{ql} \neq 0, 1$, the log ratio of the posterior probabilities $\delta^{\boldsymbol{\pi}}$ as well as its conditional expectation $\Delta^{\boldsymbol{\pi}}$ are always finite. Here, we have

$$Y_{ij} = X_{ij}\log\left(\frac{\pi_{z_i^\star w_j^\star}}{\pi_{z_i w_j}}\right) + (1 - X_{ij})\log\left(\frac{1-\pi_{z_i^\star w_j^\star}}{1-\pi_{z_i w_j}}\right) + c,$$

where c is a centering constant. Conditional on $(\mathbf{Z}_n, \mathbf{W}_m) = (\mathbf{z}_n^\star, \mathbf{w}_m^\star)$, the random variables Y_{ij} are independent, centered and bounded by $2[\log(1-a) - \log a]$. A simple Hoeffding's Inequality then yields (29) with $\psi^\star_{\max}(x) = x^2/[2\{\log(1-a)-\log a\}^2]$. This gives the desired the result. □

Corollary 4. *Consider the model defined by* (1) *where $f(\cdot;\pi)$ is a Bernoulli distribution with parameter $\pi \in \Pi$. Under ii) of Assumption 1 and Assumptions 6, 7, the conclusions of Theorem 1 and Corollaries 1 to 3 are valid.*

Proof. According to Lemma 3, it suffices to prove that Assumptions 4 and 5 are valid. But as we assume $\pi_{ql} \neq 0, 1$, the Bernoulli distributions are supported exactly on $\{0,1\}$ and the requirement $\kappa_{\max} < +\infty$ is satisfied. Moreover, as $\pi_{ql} \in [a, 1-a]$, Assumption 5 is satisfied with $L_0 = 1/a$. □

4.3 Binomial observations

In this section, we assume that $X_{ij} \in \{0, \ldots, p\}$ and $f(\cdot;\pi)$ is a Binomial distribution $\mathcal{B}(p,\pi)$. In this case, point $i)$ in Assumption 1 is satisfied. We shall also make Assumption 7, so that Assumptions 3 and 4 are also satisfied.

Lemma 4. *Under Assumption 7, we obtain that Assumption 3 is satisfied with $\psi^\star(x) = x^2\mu^2_{\min}/[16p^2(\log(1-a) - \log a)^2]$.*

Proof. According to Equation (5) and Assumption 7 ensuring $\pi_{ql} \neq 0, 1$, the log ratio of the posterior probabilities $\delta^{\boldsymbol{\pi}}$ as well as its conditional expectation $\Delta^{\boldsymbol{\pi}}$ are always finite. Here, we have

$$Y_{ij} = \sum_{k=0}^{p} 1\{X_{ij} = k\} \left\{ k \log \left(\frac{\pi_{z_i^\star w_j^\star}}{\pi_{z_i w_j}} \right) + (p-k) \log \left(\frac{1 - \pi_{z_i^\star w_j^\star}}{1 - \pi_{z_i w_j}} \right) \right\} + c,$$

where c is a centering constant. Now, the Y_{ij}'s are bounded by $2p\{\log(1-a) - \log a\}$ and the same proof as in Lemma 3 applies. □

The proof of the following corollary follows the same lines as for Corollary 4 and is omitted.

Corollary 5. *Consider the model defined by* (1) *where* $f(\cdot;\pi)$ *is a Binomial distribution* $\mathcal{B}(p,\pi)$ *with parameter* $\pi \in \Pi$. *Under ii) of Assumption 1 and Assumptions 6, 7, the conclusions of Theorem 1 and Corollaries 1 to 3 are valid.*

4.4 Discrete observations

In this section, we assume that $X_{ij} \in \{1, \ldots, p\}$ and $f(\cdot;\pi)$ is a discrete distribution with parameter $\pi = (\pi(1), \ldots, \pi(p))$ and $f(k;\pi) = \pi(k)$ for any $1 \leq k \leq p$. In this case, point *i)* in Assumption 1 is automatically satisfied. We state a condition on the parameters $\pi \in \Pi$ that ensures both Assumptions 3 and 4 are also satisfied.

Assumption 8. *The parameter set* Π *is included in* $[a, 1-a]^p$ *for some* $a \in (0, 1/2)$.

Lemma 5. *Under Assumption 8, we obtain that Assumption 3 is satisfied with* $\psi^\star(x) = x^2 \mu_{\min}^2 / \{8p[\log(1-a) - \log a]^2\}$.

Proof. According to Equation (5) and Assumption 8 ensuring $\pi_{ql}(k) \neq 0, 1$, the log ratio of the posterior probabilities $\delta^{\boldsymbol{\pi}}$ as well as its conditional expectation $\Delta^{\boldsymbol{\pi}}$ are always finite. Here, we have

$$Y_{ij} = \sum_{k=1}^{p} 1\{X_{ij} = k\} \log \left(\frac{\pi_{z_i^\star w_j^\star}(k)}{\pi_{z_i w_j}(k)} \right) + c,$$

where c is a centering constant. Conditional on $(\mathbf{Z}_n, \mathbf{W}_m) = (\mathbf{z}_n^\star, \mathbf{w}_m^\star)$, the random variables Y_{ij} are independent, centered and bounded by $p[\log(1-a) - \log a]$. A simple Hoeffding's Inequality then yields (29) with $\psi_{\max}^\star(x) = x^2 / \{p[\log(1-a) - \log a]^2\}$. This gives the desired the result. □

Corollary 6. *Consider the model defined by* (1) *where* $f(\cdot;\pi)$ *is a discrete distribution with parameter* $\pi \in \Pi$. *Under ii) of Assumption 1 and Assumptions 6, 8, the conclusions of Theorem 1 and Corollaries 1 to 3 are valid.*

Proof. According to Lemma 5, it suffices to prove that Assumptions 4 and 5 are valid. But as we assume $\pi_{ql}(k) \neq 0, 1$ for all k, the discrete distributions are supported exactly on $\{1, \ldots, p\}$ and the requirement $\kappa_{\max} < +\infty$ is satisfied. Moreover, Assumption 5 is satisfied with $L_0 = 1/a$. □

4.5 Poisson observations

In this section, we assume that $X_{ij} \in \mathbb{N}$ and $f(\cdot;\pi)$ is a Poisson distribution with parameter $\pi \in \Pi$. In this case, point *i)* in Assumption 1 is automatically satisfied. We state a condition on the parameter $\pi \in \Pi$ that ensures both Assumptions 3 and 4 are also satisfied.

Assumption 9. *The parameter set* Π *is included in* $[\pi_{\min}, \pi_{\max}] \subset (0; +\infty)$.

Lemma 6. *Under Assumption 9, we obtain that Assumption 3 is satisfied with* $\psi^\star(x) = \mu_{\min}^2 \pi_{\max} h(x/(\pi_{\max}\log(\pi_{\max}/\pi_{\min})))/8$ *and* $h(u) = (1+u)\log(1+u) - u$*, for all* $u \geq -1$.

Proof. According to Equation (5) and Assumption 9 ensuring $\pi_{ql} > 0$, the log ratio of the posterior probabilities $\delta^{\boldsymbol{\pi}}$ as well as its conditional expectation $\Delta^{\boldsymbol{\pi}}$ are always finite. Here, we have

$$Y_{ij} = \log\left(\frac{\pi_{z_i^\star w_j^\star}}{\pi_{z_i w_j}}\right)(X_{ij} - \pi_{z_i^\star w_j^\star}).$$

Conditional on $(\mathbf{Z}_n, \mathbf{W}_m) = (\mathbf{z}_n^\star, \mathbf{w}_m^\star)$, the random variables Y_{ij} are independent, centered and up to a scale factor, these are Poisson random variables. We let

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

and write for any $x > 0$, a Cramer-Chernoff bound for a Poisson variable

$$\begin{aligned}\mathbb{P}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}(|Y_{ij}| \geq x) &\leq \mathbb{P}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\left(|X_{ij} - \pi_{z_i^\star w_j^\star}| \geq \frac{x}{\log(\pi_{\max}/\pi_{\min})}\right) \\ &\leq 2\exp\left\{-\pi_{z_i^\star w_j^\star} h\left(\frac{x}{\pi_{z_i^\star w_j^\star}\log(\pi_{\max}/\pi_{\min})}\right)\right\},\end{aligned}$$

(see for instance Massart, 2007). Since for any $u > 0$, we have $\pi \mapsto -\pi h(u/\pi)$ is increasing on $(0, +\infty)$, we obtain

$$\mathbb{P}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}(|Y_{ij}| \geq x) \leq 2\exp\left\{-\pi_{\max} h\left(\frac{x}{\pi_{\max}\log(\pi_{\max}/\pi_{\min})}\right)\right\}.$$

Let $D = \mathrm{diff}(\mathbf{z}_n, \mathbf{w}_m, \mathbf{z}'_n, \mathbf{w}'_m)$. The conditional independence of the Y_{ij}'s combined with the previous Cramer-Chernoff bound yields

$$\mathbb{P}_\theta^{\mathbf{z}_n, \mathbf{w}_m}\left(\Big|\sum_{(i,j)\in\mathcal{I}} Y_{ij}\Big| \geq \epsilon D\right) \leq 2\exp\left\{-D\pi_{\max} h\left(\frac{\epsilon}{\pi_{\max}\log(\pi_{\max}/\pi_{\min})}\right)\right\},$$

which concludes the proof. □

Corollary 7. *Consider the model defined by* (1) *where* $f(\cdot;\pi)$ *is a Poisson distribution with parameter* $\pi \in \Pi$*. Under ii) of Assumption 1 and Assumptions 6, 9, the conclusions of Theorem 1 and Corollaries 1 to 3 are valid.*

Proof. According Lemma 6, it suffices to prove that Assumption 4 and 5 are valid. But as we assume $\pi_{ql} > 0$, the Bernoulli distributions are supported exactly on $\mathbb{N}$ and the requirement $\kappa_{\max} < +\infty$ is satisfied. Moreover, Assumption 5 is satisfied with $L_0 = \pi_{\max}/\pi_{\min}$. □

4.6 Gaussian location model

In this section, we are interested in Gaussian observations in the homoscedastic case. We thus 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, we have $f(x;\pi_{ij}) = c\exp\{-(x-\pi_{ij})^2/(2\sigma^2)\}$, where c is a normalizing constant. Note that point *i)* in Assumption 1 is satisfied. We also require bounded values for $\pi \in \Pi$ for concentration inequalities to be uniformly satisfied, namely a variant of Assumption 9 as we do not impose positivity on the parameters.

Assumption 10. *The parameter set* Π *is included in* $[\pi_{\min}, \pi_{\max}] \subset \mathbb{R}$.

Lemma 7. *Under Assumption 10, we obtain that Assumption 3 is satisfied with* $\psi^\star(x) = x^2\sigma^2\mu_{\min}^2/(16(\pi_{\max}-\pi_{\min})^2)$.

Proof. Since the distributions $f(\cdot;\pi)$ are absolutely continuous with respect to the Lebesgue measure, the log ratio of the posterior probabilities $\delta^{\boldsymbol{\pi}}$ as well as its conditional expectation $\Delta^{\boldsymbol{\pi}}$ are always finite. Here, we have

$$Y_{ij} = \frac{1}{\sigma^2}(\pi_{z_i w_j} - \pi_{z_i^\star w_j^\star})(X_{ij} - \pi_{z_i^\star w_j^\star}).$$

Thus, Y_{ij} is Gaussian centered with variance $(\pi_{z_i w_j} - \pi_{z_i^\star w_j^\star})^2/\sigma^2$. A Cramer-Chernoff bound for Gaussian variables gives, for any $x > 0$ (see for instance Massart, 2007),

$$\mathbb{P}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}(|Y_{ij}| \geq x) \leq 2\exp\left(-\frac{\sigma^2 x^2}{2(\pi_{z_i w_j} - \pi_{z_i^\star w_j^\star})^2}\right) \leq 2\exp\left(-\frac{\sigma^2 x^2}{2(\pi_{\max}-\pi_{\min})^2}\right).$$

Combining this with the independence of the Y_{ij}'s and letting $D = \mathrm{diff}(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m)$, we obtain that for any $\epsilon > 0$,

$$\mathbb{P}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\left(|\sum_{(i,j)\in\mathcal{I}} Y_{ij}| \geq \epsilon D\right) \leq 2\exp\left(-\frac{D\sigma^2\epsilon^2}{2(\pi_{\max}-\pi_{\min})^2}\right).$$

This corresponds to Inequality (29) with $\psi_{\max}^\star(x) = x^2\sigma^2/(2(\pi_{\max}-\pi_{\min})^2)$, which gives the desired result. ☐

The following corollary is a direct consequence of the previous lemma and the fact that Assumption 5 is satisfied in this case with $L_0 = \pi_{\max}/\sigma^2$.

Corollary 8. *Consider the model defined by* (1) *where* $f(\cdot;\pi)$ *is a Gaussian distribution with mean value* π *and fixed variance* σ^2. *Under ii) of Assumption 1 and Assumptions 6, 10, the conclusions of Theorem 1 and Corollaries 1 to 3 are valid.*

4.7 Gaussian scale model

In this section, we are interested in Gaussian observations with fixed mean and different variances. We thus 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, we have $f(x;\pi_{ij}) = c(\pi_{ij})^{-1/2}\exp\{-(x-m)^2/(2\pi_{ij})\}$, where c is a normalizing constant. Note that point *i)* in Assumption 1 is satisfied. We also impose bounded values for $\pi \in \Pi$, namely Assumption 9, for concentration inequalities to be uniformly satisfied.

Lemma 8. *Under Assumption 9, we obtain that Assumption 3 is satisfied with* $\psi^\star(x) = \mu_{\min}^2 \pi_{\min} x / \{8(\pi_{\max} - \pi_{\min})\} + \mu_{\min}^2 \log\{1 + 2\pi_{\min} x/(\pi_{\max} - \pi_{\min})\}/16$.

Proof. Since the distributions $f(\cdot;\pi)$ are absolutely continuous with respect to the Lebesgue measure, the log ratio of the posterior probabilities $\delta^{\boldsymbol{\pi}}$ as well as its conditional expectation $\Delta^{\boldsymbol{\pi}}$ are always finite. Here, we have

$$Y_{ij} = \frac{\pi_{z_i^\star w_j^\star} - \pi_{z_i w_j}}{2\pi_{z_i w_j}} \times \left(\frac{(X_{ij} - m)^2}{\pi_{z_i^\star w_j^\star}} - 1 \right).$$

Thus, up to a scale factor, Y_{ij} follows a centered $\chi^2(1)$ (χ-square with 1 degree of freedom) distribution. A Cramer-Chernoff bound for $\chi^2(1)$ random variables gives, for any $x > 0$ (see for instance Massart, 2007),

$$\begin{aligned} \mathbb{P}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}(|Y_{ij}| \geq x) = \mathbb{P}_\theta \left(|X - 1| \geq \frac{2\pi_{z_i w_j} x}{|\pi_{z_i^\star w_j^\star} - \pi_{z_i w_j}|} \right) \leq \mathbb{P}_\theta \left(|X - 1| \geq \frac{2\pi_{\min} x}{\pi_{\max} - \pi_{\min}} \right) \\ \leq 2 \exp\left\{ -\frac{\pi_{\min} x}{\pi_{\max} - \pi_{\min}} + \frac{1}{2} \log\left(1 + \frac{2\pi_{\min} x}{\pi_{\max} - \pi_{\min}} \right) \right\}, \end{aligned}$$

where $X \sim \chi^2(1)$. Combining this bound with the conditional independence of the Y_{ij}'s, Assumption 9 and letting $D = \mathrm{diff}(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m)$, we obtain that for any $\epsilon > 0$,

$$\mathbb{P}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star} \left(\Big| \sum_{(i,j) \in \mathcal{I}} Y_{ij} \Big| \geq \epsilon D \right) \leq 2 \exp\left\{ -\frac{\pi_{\min} \epsilon D}{\pi_{\max} - \pi_{\min}} + \frac{D}{2} \log\left(1 + \frac{2\pi_{\min} \epsilon}{\pi_{\max} - \pi_{\min}} \right) \right\}.$$

This corresponds to Inequality (29) and leads to the desired result. □

The following corollary is a direct consequence of the previous lemma and the fact that Assumption 5 is satisfied in this case with $L_0 = 1/(2\pi_{\min})$.

Corollary 9. *Consider the model defined by* (1) *where* $f(\cdot;\pi)$ *is a Gaussian distribution with fixed mean value* m *and variance* $\pi \in \Pi$. *Under ii) of Assumption 1 and Assumptions 6, 9, the conclusions of Theorem 1 and Corollaries 1 to 3 are valid.*

4.8 Mixture of Dirac and continuous distribution

In this section, we assume that X_{ij} follows a mixture of a Dirac mass at zero and a continuous distribution (on $\mathbb{R}$ for instance). This situation is particularly relevant for modeling sparse matrices (Ambroise and Matias, 2011). In this context, the former parameter π becomes now $(\pi, \gamma) \in (0, 1) \times \Gamma$ and we let

$$f(\cdot;\pi,\gamma) = \pi \tilde{f}(\cdot;\gamma) + (1 - \pi)\delta_0(\cdot), \tag{30}$$

where δ_0 is the Dirac mass at 0. 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 $\{\tilde{f}(\cdot;\gamma); \gamma \in \Gamma\}$ have exactly the same support so that for any $\gamma \in \Gamma$, the random variable $\tilde{f}(X_{ij};\gamma)$ is $\mathbb{P}_\star$-almost surely non zero.

Assumption 11. *Each distribution in* $\{\tilde{f}(\cdot;\gamma);\gamma\in\Gamma\}$ *admits a continuous c.d.f. at zero. Moreover, the distributions* $\{\tilde{f}(\cdot;\gamma);\gamma\in\Gamma\}$ *have exactly the same support.*

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 $\tilde{f}(k;\gamma)=\gamma^k/(k!)(e^\gamma-1)^{-1}$. 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 π satisfies Assumption 7 and that the family $\{\tilde{f}(\cdot;\gamma);\gamma\in\Gamma\}$ satisfies a concentration property on its likelihood ratio statistics as follows.

Assumption 12. *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* $\tilde{Y}_{ij}=\log[\tilde{f}(X_{ij};\gamma_{z_i^\star w_j^\star})/\tilde{f}(X_{ij};\gamma_{z_iw_j})]+c$*, 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* $(i,j)\in\mathcal{I}$*,*

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

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

Lemma 9. *Under Assumptions 7, 11 and 12, we obtain that Assumption 3 is satisfied, up to an extra factor 2, with* $\psi^\star(x)=\mu_{\min}^2(\tilde{\psi}^\star_{\max}(x/2)\wedge x^2/\{8[\log(1-a)-\log a]^2\})/8$*. Namely, using the same notation as in Assumption 3, we get*

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

Proof. According to Assumptions 7 and 11, the log ratio of the posterior probabilities $\delta^{\boldsymbol{\pi}}$ as well as its conditional expectation $\Delta^{\boldsymbol{\pi}}$ are always finite. Here, we have $Y_{ij}=Y_{ij}^{(1)}+Y_{ij}^{(2)}$ where

$$\begin{aligned}
Y_{ij}^{(1)} &= 1\{X_{ij}=0\}\log\left(\frac{1-\pi_{z_i^\star w_j^\star}}{1-\pi_{z_iw_j}}\right)+1\{X_{ij}\neq 0\}\log\left(\frac{\pi_{z_i^\star w_j^\star}}{\pi_{z_iw_j}}\right)+c_1,\\
Y_{ij}^{(2)} &= 1\{X_{ij}\neq 0\}\log\left(\frac{\tilde{f}(X_{ij};\gamma_{z_i^\star w_j^\star})}{\tilde{f}(X_{ij};\gamma_{z_iw_j})}\right)+c_2=1\{X_{ij}\neq 0\}\tilde{Y}_{ij},
\end{aligned}$$

where c_1,c_2 are centering constants. Conditional on $(\mathbf{Z}_n,\mathbf{W}_m)=(\mathbf{z}_n^\star,\mathbf{w}_m^\star)$, the families $\{Y_{ij}^{(1)}\}_{(i,j\in\mathcal{I})}$ and $\{Y_{ij}^{(2)}\}_{(i,j\in\mathcal{I})}$ respectively contain independent and centered random variables. Moreover, as the random variables $\{Y_{ij}^{(1)}\}_{(i,j\in\mathcal{I})}$ are bounded by $2[\log(1-a)-\log a]$, we can apply a Cramer-Chernoff bound on the deviation of each $Y_{ij}^{(1)}$. For any $x>0$, we write

$$\begin{aligned}
\mathbb{P}_\star^{\mathbf{z}_n^\star\mathbf{w}_m^\star}(|Y_{ij}|\geq x) &\leq \mathbb{P}_\star^{\mathbf{z}_n^\star\mathbf{w}_m^\star}\left(|Y_{ij}^{(1)}|\geq x/2\right)+\mathbb{P}_\star^{\mathbf{z}_n^\star\mathbf{w}_m^\star}\left(|Y_{ij}^{(2)}|\geq x/2\right)\\
&\leq 2\exp\left\{-\frac{(x/2)^2}{2[\log(1-a)-\log a]^2}\right\}+p_{z_iw_j}\mathbb{P}_\star^{\mathbf{z}_n^\star\mathbf{w}_m^\star}\left(|\tilde{Y}_{ij}|\geq x/2\Big|X_{ij}\neq 0\right)\\
&\leq 2\exp\left\{-\frac{x^2}{8[\log(1-a)-\log a]^2}\right\}+2\exp\{-\sup_{\lambda\in I}(\lambda x/2-\tilde{\psi}_{\max}(\lambda))\}.
\end{aligned}$$

Let $D = \text{diff}(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m)$. Combining these Cramer-Chernoff bounds with the respective conditional independence of the $\{Y_{ij}^{(1)}\}_{i,j}$ and $\{Y_{ij}^{(2)}\}_{i,j}$ yields

$$\begin{aligned}
&\mathbb{P}_\star^{\mathbf{z}_n^\star \mathbf{w}_m^\star}\Big(\big|\sum_{ij} Y_{ij}\big| \geq \epsilon D\Big) \\
\leq& 2\exp\left\{-\frac{D\epsilon^2}{8[\log(1-a)-\log a]^2}\right\} + 2\exp\{-D\sup_{\lambda\in I}(\lambda\epsilon/2 - \tilde{\psi}_{\max}(\lambda))\} \\
\leq& 4\exp\left\{-D\Big(\tilde{\psi}^\star_{\max}(\epsilon/2) \wedge \frac{\epsilon^2}{8[\log(1-a)-\log a]^2}\Big)\right\}.
\end{aligned}$$

This corresponds to Inequality 29 (up to an extra factor 2) and yields the result. □

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

Assumption 13. *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*

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

Note that we provided in the previous sections many examples of families for which this assumption is satisfied. Combined with Assumption 7, this ensures that Assumption 5 is satisfied with $L_0 = a^{-1} + \tilde{L}_0$. Then, the following corollary is a direct consequence from the previous results and Assumption 11 ensuring that $\kappa_{\max}$ is always finite.

Corollary 10. *Consider the model defined by* (1) *where* $f(\cdot;\pi,\gamma)$ *is a mixture given by* (30)*. Under Assumptions 1, 6, 7, 11, 12 and 13, the conclusions of Theorem 1 and Corollaries 1 to 3 are valid.*

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 4.2 and the second one being built on the weighted case from Section 4.8. 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 $(\xi_n)_{n\geq1}$ 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 Bickel and Chen (2009). We shall now assume that the unscaled connection/sparsity probabilities are well-behaved, as in Assumption 7 and shall 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 14. *The parameter sets Π_n and $\Pi_{\mathcal{QL},n}$ depend on the number of observations and we have*

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

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

$$\frac{\log n}{n\xi_n^2} \to 0 \quad \textit{and} \quad \frac{\log n}{m_n \xi_n^2} \to 0.$$

5.1 Binary block models with a vanishing density

In this setup, 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 (10) 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.

Lemma 10. *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 14, we have for all n and all $(q,l),(q',l') \in \mathcal{Q}\times\mathcal{L}$*

$$\kappa_{\min,n} := \kappa_{\min}(\boldsymbol{\pi}^\star_n) \geq \xi_n c_{\min}(\boldsymbol{\pi}^\star), \tag{31}$$

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

$$\psi^\star_n(x) := \psi^\star(x) = x^2 \mu^2_{\min}/[16\{\log(1-a) - \log a\}^2], \tag{33}$$

where

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

Proof. For any $\pi, \pi' \in \Pi$ and any $\xi \in (0,1)$, the Kullback-Leibler divergence $D(\xi\pi||\xi\pi')$ writes

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

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

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

Coming back to the definition (10) of $\kappa_{\min}(\boldsymbol{\pi}_n^\star)$ yields

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

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

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

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

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

Finally, we need the correct scaling for $\psi_n^\star(x)$ that appears in Assumption 3. Following the scheme of proof developed in Section 4.1, it turns out that the random variables $Y_{ij,n}$ defined by

$$Y_{ij,n} = X_{ij,n}\log\left(\frac{\pi_{z_i^\star w_j^\star,n}}{\pi_{z_i w_j,n}}\right) + (1-X_{ij,n})\log\left(\frac{1-\pi_{z_i^\star w_j^\star,n}}{1-\pi_{z_i w_j,n}}\right) + c_n,$$

(where c_n is a centering constant) still satisfy that, conditional on $(\mathbf{Z}_n, \mathbf{W}_m) = (\mathbf{z}_n^\star, \mathbf{w}_m^\star)$, these are independent and bounded by $2[\log(1-a) - \log a]$. The same Hoeffding's Inequality then yields (29) with $\psi_n^\star(x) = \psi^\star(x) = x^2\mu_{\min}^2/[16\{\log(1-a) - \log a\}^2]$. □

Corollary 11. *Under Assumption 1 on the unscaled parameter set $\Pi_{\mathcal{QL}}$ and Assumption 14 Theorem 1 and Corollaries 1 to 3 remain valid with the following modifications*

1. $c = \mu_{\min}^2 c_{\min}/16$;
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 4. *Assumption 6 is not in force in this theorem. Indeed, the scaling imposed with ξ_n in Assumption 14 implies it: it forces $\log n/(n\xi_n^2) \to 0$ and $\log n/(m_n\xi_n^2) \to 0$ and thus makes the speed at which m_n/n can go to 0 depend on ξ_n. Note that $m_n\xi_n^2$ plays in Assumption 14 the same role as m_n in Assumption 6.*

Proof. The proof is essentially the same as the proof of Theorem 1. We will only highlight the differences and show how the scaling $\log n/(m_n\xi_n^2) \to 0$ is derived. First Equation (12) from Proposition 1 now depends on n and should be

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

where the original $c = \mu_{\min}^2\kappa_{\min}/16$ has been changed to $c' = \mu_{\min}^2 c_{\min}/16$. Next the set $A(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m)$ 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

$$\begin{aligned}\mathbb{P}_\star(A(\mathbf{z}_n^\star, \mathbf{w}_m^\star, \mathbf{z}_n, \mathbf{w}_m) \cap \{(\mathbf{Z}_n, \mathbf{W}_m) = (\mathbf{z}_n^\star, \mathbf{w}_m^\star)\})& \\ \leq 2\exp[-\psi^\star(\xi_n(c' - 2L_0\eta))(mr_1 + nr_2)]\boldsymbol{\mu}(\mathbf{z}_n^\star, \mathbf{w}_m^\star).& \end{aligned} \tag{35}$$

The set Ω_1 is still defined as in Equation (23) and on this set, Inequality (21) and (16) are still statisfied. However the upper bound on $\mathbb{P}_\star(\overline{\Omega}_1)$ is modified as follows

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

Combining the latter with the control of the probability of $\overline{\Omega}_0$ given in Proposition 1 and the quadratic nature of $\psi^\star$, we obtain

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

where $d_{n,m} = [n\exp\{-m\xi_n^2\psi^\star(c' - 2L_0\eta)\} + m\exp\{-n\xi_n^2\psi^\star(c' - 2L_0\eta)\}]$. The condition required to make the $\varepsilon_{n,m}$ summable and conclude the proof is $\log n/(m\xi_n^2) \to 0$. This condition holds under Assumption 14. □

5.2 Weighted models with a vanishing density

We now consider the setup introduced in Section 4.8, 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 14). Note that the parameters $(\gamma_{ql})_{(q,l)\in\mathcal{Q}\times\mathcal{L}}$ remain fixed. Moreover, Assumptions 11 and 12 are assumed to hold.

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 13 is sufficient to guarantee the adequate scaling of the Lipschitz condition.

Lemma 11. *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 11 to 14, we have for all* n *and all* $(q,l),(q',l') \in \mathcal{Q}\times\mathcal{L}$

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

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

$$\psi_n^\star(x) := \psi^\star(x) = \frac{\mu_{\min}^2}{8}\left(\tilde{\psi}^\star(\frac{x}{2}) \wedge \frac{x^2}{8\{\log(1-a)-\log a\}^2} \right), \tag{38}$$

where

$$\kappa_{\min} := \kappa_{\min}(\boldsymbol{\gamma}^\star) = \min\left\{ D(\gamma^\star_{ql}\|\gamma^\star_{q'l'}); (q,l),(q',l') \in \mathcal{Q}\times\mathcal{L}, \gamma^\star_{ql} \neq \gamma^\star_{q'l'} \right\} > 0$$

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

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

$$\int_{\mathcal{X}} \log \frac{f(x;\xi\pi,\gamma)}{f(x;\xi\pi',\gamma')} f(x;\xi\pi'',\gamma'')dx = \xi\pi'' \log\frac{\pi}{\pi'} + (1-\xi\pi'')\log\frac{1-\pi}{1-\pi'} + \xi\pi'' \int_{\mathcal{X}} \log\frac{\tilde{f}(x;\gamma)}{\tilde{f}(x;\gamma')}\tilde{f}(x;\gamma'')dx. \tag{39}$$

When $(\pi'',\gamma'') = (\pi,\gamma)$, Equation (39) turns to

$$D\left((\xi\pi,\gamma)\|(\xi\pi',\gamma')\right) = D(\xi\pi\|\xi\pi') + \xi\pi D(\gamma\|\gamma') \geq \xi\frac{(\pi-\pi')^2}{2\pi}\left(\frac{a}{1-a}\right)^2 + \xi a D(\gamma\|\gamma'),$$

from which we can deduce Inequality (36).

For general (π'',γ''), Equation (39) combined with Inequality (32) and Assumption 13 gives

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

from which we can deduce Inequality (37).

Finally, in this setup Lemma 9 is still valid and gives Equation (38) □

Now, we introduce an assumption about the quadratic nature of the function $\tilde{\psi}^\star_{\max}$ introduced in Assumption 12.

Assumption 15. *With the notation of Assumption 12, for all* $x > 0$*, there exists some positive* m_x *such that for all* $\xi \in (0,m_x)$

$$\tilde{\psi}^\star_{\max}(\xi x) \geq \xi^2 \tilde{\psi}^\star_{\max}(x).$$

Remark 5. *Assumption 15 ensures that* $\psi^\star(\xi x)$ *defined in Equation (38) does not decrease faster than* ξ^2 *with* ξ *and that the condition* $\log(n)/(m_n\xi_n^2) \to 0$ *is the correct asymptotics in Corollary 12. Note also that Assumption 15 holds for all distributions considered in Section 4.*

Corollary 12. *Under Assumption 1 on the unscaled parameter set* $\Pi_{\mathcal{QL}}$ *and Assumptions 11 to 15, 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 11, with some changes in key quantities as listed in the corollary. □

A Technical proofs

Proof of Lemma 2. Let us recall that this proof is a generalization of the proof of (Proposition B.5 in Celisse et al., 2011).

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

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

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 $k \in \{1, \ldots, \lceil n\mu_{\min}/2 \rceil\}$ (resp. each $j \in \{1, \ldots, \lceil m\mu_{\min}/2 \rceil\}$), 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

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

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

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

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

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

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

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

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

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 $\mathbf{z}^\star_n$. 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 $z_i = s(z^\star_i)$ 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 $s(z^\star_i)$, 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 $w_i = t(w^\star_i)$ 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

$$\begin{aligned}
\mathbf{z}_n &= (s_1(1), \ldots, s_1(Q), \ldots, s_{d_1}(1), \ldots, s_{d_1}(Q), s(z^\star_{d_1Q+1}), \ldots, s(z^\star_n)),\\
\mathbf{w}_m &= (t_1(1), \ldots, t_1(L), \ldots, t_{d_2}(1), \ldots, t_{d_2}(L), t(w^\star_{d_2L+1}), \ldots, t(w^\star_m)),
\end{aligned}$$

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

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

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

$$\begin{aligned}
\mathrm{diff}(\mathbf{z}_n, \mathbf{w}_m, \mathbf{z}^\star_n, \mathbf{w}^\star_m) &\geq \sum_{k=1}^{\lceil n\mu_{\min}/2\rceil} \sum_{j=1}^{\lceil m\mu_{\min}/2\rceil} B_{kj}\\
&\geq d_1\lceil m\mu_{\min}/2\rceil + d_2\lceil n\mu_{\min}/2\rceil - d_1 d_2\\
&\geq \frac{d_1\lceil m\mu_{\min}/2\rceil + d_2\lceil n\mu_{\min}/2\rceil}{2}\\
&\geq \frac{r_1\lceil m\mu_{\min}/2\rceil + r_2\lceil n\mu_{\min}/2\rceil}{2Q}\\
&\geq \frac{\mu^2_{\min}}{4}(mr_1 + nr_2),
\end{aligned}$$

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

□

References

E. Airoldi, D. Blei, S. Fienberg, and E. Xing. Mixed-membership stochastic blockmodels. *Journal of Machine Learning Research*, 9:1981–2014, 2008.

E. Allman, C. Matias, and J. Rhodes. Identifiability of parameters in latent structure models with many observed variables. *Ann. Statist.*, 37(6A):3099–3132, 2009.

E. Allman, C. Matias, and J. Rhodes. Parameters identifiability in random graph mixture models. *Journal of Statistical Planning and Inference*, 141(5):1719–1736, 2011.

C. Ambroise and C. Matias. New consistent and asymptotically normal parameter estimates for random graph mixture models. *Journal of the Royal Statistical Society: Series B*, 2011.

P. Bickel and A. Chen. A nonparametric view of network models and Newman-Girvan and other modularities. *PNAS*, 106(50):21068–21073, 2009. doi: 10.1073/pnas.0907096106.

A. Celisse, J.-J. Daudin, and L. Pierre. Consistency of maximum-likelihood and variational estimators in the Stochastic Block Model. Technical report, ArXiv:1105.3288, 2011.

D. S. Choi, P. J. Wolfe, and E. M. Airoldi. Stochastic blockmodels with a growing number of classes. *Biometrika*, 2012. doi: 10.1093/biomet/asr053. URL `http://biomet.oxfordjournals.org/content/early/2012/04/14/biomet.asr053.abstract`. in press.

J.-J. Daudin, F. Picard, and S. Robin. A mixture model for random graphs. *Stat. Comput.*, 18(2):173–183, 2008.

W. DeSarbo, D. Fong, J. Liechty, and M. Kim Saxton. A hierarchical bayesian procedure for two-mode cluster analysis. *Psychometrika*, 69:547–572, 2004.

O. Frank and F. Harary. Cluster inference by using transitivity indices in empirical graphs. *J. Amer. Statist. Assoc.*, 77(380):835–840, 1982.

S. Gazal, J.-J. Daudin, and S. Robin. Accuracy of variational estimates for random graph mixture models. *Journal of Statistical Computation and Simulation*, 0(0):1–14, 2011. doi: 10.1080/00949655.2011.560117.

G. Govaert and M. Nadif. Clustering with block mixture models. *Pattern Recognition*, 36 (2):463–473, 2003.

G. Govaert and M. Nadif. Block clustering with Bernoulli mixture models: Comparison of different approaches. *Computational Statistics and Data Analysis*, 52(6):3233–3245, 2008.

G. Govaert and M. Nadif. Latent block model for contingency table. *Communications in Statistics - Theory and Methods*, 39(3):416–425, 2010.

J. A. Hartigan. Direct Clustering of a Data Matrix. *Journal of the American Statistical Association*, 67(337):123–129, 1972.

P. Holland, K. Laskey, and S. Leinhardt. Stochastic blockmodels: some first steps. *Social networks*, 5:109–137, 1983.

P. Latouche, E. Birmelé, and C. Ambroise. Overlapping stochastic block models with application to the french political blogosphere. *Ann. Appl. Stat.*, 5(1):309–336, 2011.

P. Latouche, E. Birmelé, and C. Ambroise. Variational bayesian inference and complexity control for stochastic block models. *Statistical Modelling*, 12(1):93–115, 2012.

M. Mariadassou, S. Robin, and C. Vacher. Uncovering latent structure in valued graphs: a variational approach. *Ann. Appl. Stat.*, 4(2):715–42, 2010.

P. Massart. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007.

K. Nowicki and T. Snijders. Estimation and prediction for stochastic blockstructures. *J. Amer. Statist. Assoc.*, 96(455):1077–1087, 2001.

F. Picard, V. Miele, J.-J. Daudin, L. Cottret, and S. Robin. Deciphering the connectivity structure of biological networks using MixNet. *BMC Bioinformatics*, 10:1–11, 2009.

K. Rohe, S. Chatterjee, and B. Yu. Spectral clustering and the high-dimensional stochastic block model. *Ann. Statist*, 39(4):1878–1915, 2011.

T. A. B. Snijders and K. Nowicki. Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *J. Classification*, 14(1):75–100, 1997.

J. Wyse and N. Friel. Block clustering with collapsed latent block models. *Statistics and Computing*, 22(2):415–428, 2012.

H. Zanghi, C. Ambroise, and V. Miele. Fast online graph clustering via Erdős Rényi mixture. *Pattern Recognition*, 41(12):3592–3599, 2008.

H. Zanghi, F. Picard, V. Miele, and C. Ambroise. Strategies for online inference of model-based clustering in large and growing networks. *Ann. Appl. Stat.*, 4(2):687–714, 2010.