



**HAL**  
open science

# Consistency and asymptotic normality of stochastic block models estimators from sampled data

Mahendra Mariadassou, Timothée Tabouy

► **To cite this version:**

Mahendra Mariadassou, Timothée Tabouy. Consistency and asymptotic normality of stochastic block models estimators from sampled data. *Electronic Journal of Statistics*, 2020, 14 (2), pp.3672-3704. 10.1214/20-ejs1750 . hal-04380551

**HAL Id: hal-04380551**

**<https://hal.inrae.fr/hal-04380551>**

Submitted on 8 Jan 2024

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

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



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

# Consistency and asymptotic normality of stochastic block models estimators from sampled data

Mahendra Mariadassou<sup>1</sup> and Timothée Tabouy<sup>2</sup>

<sup>1</sup>*MaIAGE, INRAE, Université Paris-Saclay, 78352 Jouy-en-Josas, France*  
*e-mail: [mahendra.mariadassou@inrae.fr](mailto:mahendra.mariadassou@inrae.fr)*

<sup>2</sup>*UMR MIA-Paris, AgroParisTech, INRAE, Université Paris-Saclay, 75005 Paris, France*  
*e-mail: [timothee.tabouy@agroparistech.fr](mailto:timothee.tabouy@agroparistech.fr)*

**Abstract:** Statistical analysis of network is an active research area and the literature counts a lot of papers concerned with network models and statistical analysis of networks. However, very few papers deal with missing data in network analysis and we reckon that, in practice, networks are often observed with missing values. In this paper we focus on the Stochastic Block Model with valued edges and consider a MCAR setting by assuming that every dyad (pair of nodes) is sampled identically and independently of the others with probability  $\rho > 0$ . We prove that maximum likelihood estimators and its variational approximations are consistent and asymptotically normal in the presence of missing data as soon as the sampling probability  $\rho$  satisfies  $\rho \gg \log(n)/n$ .

**Keywords and phrases:** Stochastic Block Model, Missing data, Asymptotic normality, Maximum likelihood, Concentration inequality.

Received April 2019.

## 1. Introduction

For the last decade, statistical network analyses has been a very active research topic and the statistical modeling of networks has found many applications in social sciences and biology for example Aicher et al. (2014), Barbillon et al. (2015), Mariadassou et al. (2010), Wasserman and Faust (1994) and Zachary (1977).

Many random graphs models have been widely studied, either from a theoretical or an empirical point of view. The first model studied was Erdős-Rényi model (Erdős and Renyi, 1959) which assumes that each pair of nodes (dyad) is connected independently to the others with the same probability. This model assumes homogeneity of all nodes across the network. In order to alleviate this constraint, many families of models have been introduced. Most are endowed with a latent structure (reviewed in Matias and Robin, 2014) to capture heterogeneity across nodes. Among those, the Stochastic Block Model (in short SBM, see Frank and Harary, 1982; Holland et al., 1983) is one of the oldest and most studied as it is highly flexible and can capture a large variety of structures

(affiliation, hub, bipartite and many other). In order to estimate this model, Bayesian approaches were first proposed (Snijders and Nowicki, 1997; Nowicki and Snijders, 2001) but have been superseded by variational methods (Daudin et al., 2008; Latouche et al., 2012). The former class of approaches are exact but lack the computational efficiency and scalability that the latter offers.

Theoretical guarantees concerning maximum likelihood estimators (in short MLE) and variational estimators (in short VE), based on variational approximations of the likelihood, for the binary SBM estimation are quite difficult to obtain. In Celisse et al. (2012), consistency of MLE and VE is proven but asymptotic normality requires that the estimators converges at rate at least  $n^{-1}$ , which is not proven in the paper, although some results were available for some particular cases (affiliation for example). Ambroise and Matias (2012) tackles the specific case of affiliation model with equal group proportion and proves the consistency and asymptotic normality of parameter estimates. Bickel et al. (2013) extends those results to arbitrary binary SBM graphs and improves Celisse et al. (2012) by removing the condition on the convergence rate, as it is automatically satisfied by the MLE. Following along the path of Bickel et al. (2013), Brault et al. (2020) proved consistency and asymptotic normality of estimators (MLE and VE) to weighted Latent Block Models where the weights distribution belongs to a one-dimensional exponential families. In particular, considering unbounded edge values invalidates several parts of the proofs for binary graphs and requires substantial adaptations and additional results, notably concentration inequalities for sums of unbounded, non-gaussian random variables.

Some results are also available for the related semi-parametric problem of assignment reconstruction. Mariadassou and Matias (2015) show that the conditional distribution of the (latent) assignments converge to a degenerate distribution and Rohe et al. (2010) prove that, when the data are generated according to a SBM model, spectral methods are consistent. Choi et al. (2012) extend those results to settings where the density of the graph goes to 0 as  $\Omega(\log^\alpha(n)/n)$  (for  $\alpha$  large enough) and/or the number of groups goes to  $+\infty$  as  $\sqrt{n}$ . Chatterjee (2015) proves also strong results for reconstruction of large matrices with noisy entries estimation and partial observation of the dyads, by means of a universal singular value thresholding (USVT). In the special case of binary SBM with  $k$  groups, he achieves a reconstruction error rate of order  $\sqrt{k/n}$  as soon as the fraction of observed dyads is at least  $\Omega(\log^\alpha(n)/n)$  for (for  $\alpha$  large enough). Since USVT replaces missing dyads with 0s, it naturally achieves the same limiting rate as the sparse setting. Finally, Wang and Bickel (2017) and Hu et al. (2017) also show that model selection for the number  $k$  of groups is consistent for dense graphs, they suggest using a penalized likelihood criteria with penalty of the form  $\frac{k(k+1)}{2} \log(n) + \lambda n \log(k)$  where  $\lambda$  is a tuning parameter.

In this paper we consider a simple setting with fixed number of groups and fixed density but weighted edges and missing values. In most network studies, there is a strong asymmetry between the presence of an edge and its absence: the lack of proof that an edge exists is taken as proof that the edge does not exist

and edges with uncertain status are considered as non existent in the graph. This is the strategy adopted in most sparse asymptotic settings where the density of edges goes to 0 asymptotically (Bickel et al., 2013). We adopt a different point of view where edges with uncertain status are considered as missing, rather than absent and explicitly accounted for their missing nature. We use the framework of Rubin (1976) and its application to network data, see Kolaczyk (2009) and Handcock and Gile (2010), for parameter inference in presence of missing values and more specifically its applications to SBM (Tabouy et al., 2019). We prove that, in the MCAR setting where each dyad is missing independently and with the same probability, the MLE and variational estimates are still consistent and asymptotically normal.

The article is organized as follows. We first present the model and missing data theory applied to our context with some examples of sampling designs. We then posit some definitions and discuss the assumptions required for our results in Section 2. In Section 3 we establish asymptotic normality for the complete-observed model (*i.e.* observed SBM where latent variables are known). Section 4 is the main result of this paper and states that the observed-likelihood behaves like the complete-observed likelihood (*i.e.* joint likelihood of the observed data and latent variables) close to its maximum. Consequences for the MLE and variational estimator are in discussed in Section 5. The proof is sketched in Section 6. Comparison to existing results are made and discussed in Section 7. Technical lemmas and details of the proofs are available in the appendices.

## 2. Statistical framework

Notation	Definition
$\mathcal{N} = \{1, \dots, n\}$	Node set
$i, j$	Node index
$\mathcal{Q} = \{1, \dots, Q\}$	Block set
$q, l$	Block index
$\mathcal{D} = \mathcal{N} \times \mathcal{N}$	Dyad set
$y_{ij}$ and $r_{ij}$	Dyad value and observation status
$z_i$ and $z_i^*$	Test and true block membership
$\alpha$ and $\alpha^*$	Test and true block proportions
$\pi$ and $\pi^*$	Test and true parameters of dyad value distribution
$\Theta$	Parameter space

### 2.1. Stochastic Block Model

In SBM, nodes from a set  $\mathcal{N} \triangleq \{1, \dots, n\}$  are distributed among a set  $\mathcal{Q} \triangleq \{1, \dots, Q\}$  of hidden blocks that model the latent structure of the graph. The block-memberships are encoded by  $(z_i, i \in \mathcal{N})$  where the  $z_i$  are independent random variables with prior probabilities  $\alpha = (\alpha_1, \dots, \alpha_Q)$ , such that  $\mathbb{P}(z_i = q) = \alpha_q$ , for all  $q \in \mathcal{Q}$ . The value  $y_{ij}$  of any dyad  $(i, j)$  in  $\mathcal{D} = \mathcal{N} \times \mathcal{N}$ , with  $i \neq j$ , only depends on the blocks  $i$  and  $j$  belong to. The variables  $(y_{ij})$ s are

thus independent conditionally on the  $(z_i)$ s:

$$y_{ij} \mid z_i = q, z_j = \ell \sim^{\text{ind}} \varphi(\cdot, \pi_{q\ell}), \quad \forall (i, j) \in \mathcal{D}, \quad i \neq j, \quad \forall (q, \ell) \in \mathcal{Q} \times \mathcal{Q}.$$

In the following,  $\mathbf{y} = (y_{ij})_{i,j \in \mathcal{D}}$  is the  $n \times n$  adjacency matrix of the random graph,  $\mathbf{z} = (z_1, \dots, z_n)$  the  $n$ -vector of the latent blocks. With a slight abuse of notation, we associate to  $z_i$  a binary vector  $(z_{i1}, \dots, z_{iQ})$  such that  $z_i = q \Leftrightarrow z_{iq} = 1, z_{i\ell} = 0$ , for all  $\ell \neq q$ . In this case  $\mathbf{z}$  is a  $n \times Q$  matrix.

We note the complete parameter set as  $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\pi}) \in \Theta$  where  $\Theta$  stands for the parameter space. When performing inference from data, we note  $\boldsymbol{\theta}^* = (\boldsymbol{\alpha}^*, \boldsymbol{\pi}^*)$  the true parameter set, *i.e.* the parameter values used to generate the data, and  $\mathbf{z}^*$  the true (and usually unobserved) memberships of nodes. For any  $\mathbf{z}$ , we also note:

- $z_{+q} = \sum_i z_{iq}$  the size of the  $q^{\text{th}}$  community (or block) for membership  $\mathbf{z}$
- $z_{+q}^*$  its counterpart for  $\mathbf{z}^*$ .

## 2.2. Missing data for SBM

Regarding SBM inference, a missing value corresponds to a missing entry in the adjacency matrix  $\mathbf{y}$ , typically denoted by NA's. We rely on the  $n \times n$  sampling matrix  $\mathbf{r}$  to record the missing state of each entry:

$$(r_{ij}) = \begin{cases} 1 & \text{if } y_{ij} \text{ is observed,} \\ 0 & \text{otherwise.} \end{cases} \quad (2.1)$$

As a shortcut, we use  $\mathbf{y}^o = \{y_{ij} : r_{ij} = 1\}$  and  $\mathbf{y}^m = \{y_{ij} : r_{ij} = 0\}$  to respectively denote the *observed* and *missing* dyads. The *sampling design* is the description of the stochastic process that generates  $\mathbf{r}$ . It is assumed that the network exists before the sampling design acts upon it, which is fully characterized by the conditional distribution  $p_\psi(\mathbf{r}|\mathbf{y})$ , the parameters of which are such that  $\psi$  and  $\theta$  live in a product space  $\Theta \times \Psi$ . In this paper we are going to focus on a specific type of missingness, called missing completely at random (MCAR) for which  $p_\psi(\mathbf{r}|\mathbf{y}) = p_\psi(\mathbf{r})$  and leave aside more complex forms of dependencies such as Missing at random (MAR) and Not missing at random (NMAR).

We then follow the framework of Rubin (1976) and Tabouy et al. (2019) for missing data and define the joint probability density function as

$$p_{\theta, \psi}(\mathbf{y}^o, \mathbf{z}, \mathbf{r}) = \int p_\theta(\mathbf{y}^o, \mathbf{y}^m, \mathbf{z}) p_\psi(\mathbf{r}|\mathbf{y}^o, \mathbf{y}^m, \mathbf{z}) d\mathbf{y}^m. \quad (2.2)$$

**Property 2.1.** *According to Equation (2.2), if the sampling design is MCAR, then maximising  $p_{\theta, \psi}(\mathbf{y}^o, \mathbf{z}, \mathbf{r})$  or  $p_{\theta, \psi}(\mathbf{y}^o, \mathbf{r})$  in  $\theta$  is equivalent to maximising  $p_\theta(\mathbf{y}^o)$  in  $\theta$ , this corresponds to the ignorability notion defined in Rubin (1976).*

### 2.3. Sampling design examples

We present here some examples of sampling designs to illustrate differences between notions of MCAR, MAR and NMAR.

**Definition 2.2** (Random dyad sampling). *Each dyad  $(i, j) \in \mathcal{D}$  has the same probability  $\mathbb{P}(r_{ij} = 1) = \rho$  of being observed, independently of the others. This design is MCAR.*

**Definition 2.3** (Random node sampling). *The random node sampling consists in selecting independently with probability  $\rho$  a set of nodes and then observing the corresponding rows and columns of matrix  $\mathbf{y}$ .*

The major point in both examples is that the probability ( $\rho$  in *random dyad sampling* and  $1 - (1 - \rho)^2$  in the *random node sampling*) of observing a dyad does not depend on its value.

In contrast, the following dyad-centered sampling design adapted to binary networks is NMAR since the probability to observe a dyad depends on its value:

**Definition 2.4** (Double standard sampling). *Each dyad  $(i, j) \in \mathcal{D}$  is observed, independently of other dyads, with a probability depending on its value:  $\mathbb{P}(r_{ij} = 1|y_{ij} = 0) = \rho_0$  and  $\mathbb{P}(r_{ij} = 1|y_{ij} = 1) = \rho_1$ .*

For non-binary networks, specifying the sampling design is more involved and requires defining the sampling density for every possible value of  $y_{ij}$ , e.g.  $(\mathbb{P}(r_{ij} = 1|y_{ij} = k))_{k \in \mathbb{N}}$  for Poisson-valued edges.

*Remark 2.5.* In this paper, we focused on data sampled according to *random dyad sampling*, which is the simplest case but already yields valuable insights into the differences between the partially and fully sampled settings.

As observed above, there are however many other ways to sample a network. In the case of node-centered sampling design, like *random node sampling*, the main difficulty to prove consistency and asymptotic normality is the dependency between the  $r_{ij}$  variables. Indeed, in random node sampling, the variable  $r_{i_0j_0}$  depends on all  $r_{ij_0}$  and  $r_{i_0j}$  (for all  $i, j \in \mathcal{N}$ ). As a consequence, a different inference strategy is required and many results proved in this paper are not valid under *random node sampling*. NMAR sampling designs raises problem of their own: each design requires its own estimation procedure (Tabouy et al., 2019) and therefore its own analysis. For example, parameter estimation under the seemingly simple *double standard sampling* for binary networks is still an open problem: numerical experiments suggest that  $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\pi})$  and  $\boldsymbol{\psi} = (\rho_0, \rho_1)$  are jointly identifiable but there is no formal proof.

### 2.4. Observed-likelihoods

When the labels are known, the *complete-observed log-likelihood* is given by:

$$\mathcal{L}_{co}(\mathbf{z}; \boldsymbol{\theta}) = \log p(\mathbf{y}^o, \mathbf{z}; \boldsymbol{\theta}) = \sum_{i,q} z_{iq} \log \alpha_q + \sum_{\substack{i,j,q,\ell \\ i \neq j}} z_{iq} z_{j\ell} r_{ij} \log \varphi(y_{ij}; \pi_{q\ell}) \quad (2.3)$$

But the labels are usually unobserved, and the *observed log-likelihood* is obtained by integration over all memberships:

$$\mathcal{L}_o(\boldsymbol{\theta}) = \log p(\mathbf{y}^o; \boldsymbol{\theta}) = \log \left( \sum_{\mathbf{z} \in \mathcal{Z}} p(\mathbf{y}^o, \mathbf{z}; \boldsymbol{\theta}) \right). \quad (2.4)$$

## 2.5. Models and assumptions

We focus here on parametric models where  $\varphi$  belongs to a regular one-dimension exponential family in canonical form:

$$\varphi(y, \pi) = b(y) \exp(\pi y - \psi(\pi)), \quad (2.5)$$

where  $\pi$  belongs to the space  $\mathcal{A}$ , so that  $\varphi(\cdot, \pi)$  is well defined for all  $\pi \in \mathcal{A}$ . Classical properties of exponential families ensure that  $\psi$  is convex, infinitely differentiable on  $\mathring{\mathcal{A}}$ , that  $(\psi')^{-1}$  is well defined on  $\psi'(\mathring{\mathcal{A}})$ . Furthermore, when  $y_\pi \sim \varphi(\cdot, \pi)$ ,  $\mathbb{E}[y_\pi] = \psi'(\pi)$  and  $\mathbb{V}[y_\pi] = \psi''(\pi)$ .

In the following, we recall assuming that missing data are produced according to a *random dyad sampling* with parameter  $\rho > 0$ .

Moreover, we make the following assumptions on the parameter space (and the asymptotics of  $\rho$ ):

$A_0$ :  $\rho$  goes to 0 but satisfies  $\rho \gg \log(n)/n$

$A_1$ : There exists a positive constant  $c$  and a compact interval  $C_\pi$  such that

$$\Theta \subset [c, 1 - c]^{\mathcal{Q}} \times C_\pi^{\mathcal{Q} \times \mathcal{Q}} \quad \text{with} \quad C_\pi \subset \mathring{\mathcal{A}}.$$

$A_2$ : The true parameter  $\boldsymbol{\theta}^* = (\boldsymbol{\alpha}^*, \boldsymbol{\pi}^*)$  lies in the interior of  $\Theta$ .

$A_3$ : The map  $\pi \mapsto \varphi(\cdot, \pi)$  is injective.

$A_4$ : The coordinates of  $\boldsymbol{\pi}^* \psi'(\boldsymbol{\alpha}^*)$ , where  $\psi'$  is applied component-wise, are pairwise distinct.

The previous assumptions are standard. Assumption  $A_0$  ensures that the fraction of observed dyad is not too small. Assumption  $A_1$  ensures that the group proportions are bounded away from 0 and 1 so that no group disappears when  $n$  goes to infinity. It also ensures that  $\pi$  is bounded away from the boundaries of the  $\mathcal{A}$ . This is essential for the subexponential properties of Propositions 2.9 and 2.10.  $A_2$  is in line with standard assumptions in parametric statistics.  $A_3$  is necessary for identifiability purposes: the model is trivially not identifiable if the map  $\pi \mapsto \varphi(\cdot, \pi)$  is not injective.  $A_4$  ensures identifiability of SBM parameters under *random dyad sampling*. Note that, combined with  $A_3$ , it implies that all columns and all rows of  $\boldsymbol{\alpha}^*$  are distincts and therefore that no two groups have the connectivity profile. In the following, we consider the number of blocks  $Q$  to be known.

## 2.6. Identifiability

Since  $\mathbf{r}$  is independent on  $\mathbf{y}$ , the identifiability of SBM with emission law in the one-dimension exponential family under *random dyad sampling* can be stated in two steps. First the sampling parameter  $\rho$  and secondly the SBM parameters  $\boldsymbol{\theta}^* = (\boldsymbol{\alpha}^*, \boldsymbol{\pi}^*)$  given  $\rho$ .

**Proposition 2.6.** *The sampling parameter  $\rho > 0$  of random dyad sampling is identifiable w.r.t. the sampling distribution.*

*Proof.* See Tabouy et al. (2019). The proof does not depend on  $\mathbf{y}$  being binary but also holds for  $\mathbf{y}$  distributed as in Eq. (2.5).  $\square$

**Proposition 2.7.** *Let  $n \geq 2Q$  and assume that for any  $1 \leq q \leq Q$ ,  $\rho > 0$ ,  $\pi_q^* > 0$  and that the coordinates of  $\boldsymbol{\alpha}^* \psi'(\boldsymbol{\pi}^*)$ , where  $\psi'$  is applied component-wise, are pairwise distinct. Then, under random dyad sampling, SBM parameters are identifiable w.r.t. the distribution of the observed part of the SBM up to label switching.*

*Proof.* The proof is nearly identical to the one written in Tabouy et al. (2019) and inspired by Celisse et al. (2012) for the binary SBM under *random dyad sampling*. However, substituting  $\mathbb{E}[y_{ij}|z_i = q]$  to  $s_q$  in the proof ensures that  $\boldsymbol{\alpha}^*$  is identifiable. Finally, the fact that  $(\psi')^{-1}$  is a one-to-one map ensures that  $\boldsymbol{\pi}^*$  is identifiable.  $\square$

Note that asymptotically, the assumption  $n \geq 2Q$  is always satisfied since  $Q$  is fixed and  $n$  grows to infinity.

## 2.7. Subexponential variables

*Remark 2.8.* Since we restricted  $\pi$  in a bounded subset of  $\mathring{\mathcal{A}}$ , the variance of  $y_\pi$  is bounded away from 0 and  $+\infty$ . We note

$$\bar{\sigma}^2 = \sup_{\pi \in C_\pi} \mathbb{V}(y_\pi) < +\infty \quad \text{and} \quad \underline{\sigma}^2 = \inf_{\pi \in C_\pi} \mathbb{V}(y_\pi) > 0. \quad (2.6)$$

Similarly, since  $\pi$  belongs to a bounded subset of an open interval, there exists a constant  $\kappa > 0$ , such that  $[\pi - \kappa, \pi + \kappa] \subset \mathring{\mathcal{A}}$  uniformly over all  $\pi \in C_\pi$

**Proposition 2.9.** *With the previous notations, if  $\pi \in C_\pi$  and  $y_\pi \sim \varphi(\cdot, \pi)$ , then  $y_\pi$  is subexponential with parameters  $(\bar{\sigma}^2, \kappa^{-1})$ .*

**Proposition 2.10.** *Considering  $x = y_\pi r_{ij} + \lambda(1 - r_{ij})$  (we recall that  $r_{ij} \sim \mathcal{B}(\rho)$ ), with  $r_{ij}$  independent of  $y_\pi$  and  $\lambda \in \mathbb{R}$  bounded. There are non-negative numbers  $\nu$  and  $b$  such that  $x$  is subexponential with parameters  $(\nu^2, b^{-1})$ .*

*Proof.* These results derive directly from theorem C.1 (statement 2.).  $\square$

## 2.8. Symmetry

We now introduce the concepts of assignments and parameter symmetries, that must be accounted for when studying the asymptotic properties of the MLE. Complications stemming from symmetries are related to but not equivalent to the problem of label-switching in mixture models.

**Definition 2.11** (Permutation). *Let  $s$  be a permutation on  $\{1, \dots, Q\}$ . If  $\mathbf{A}$  is a matrix with  $Q$  columns and  $n$  rows, we define  $\mathbf{A}^s$  as the matrix obtained by permuting the columns of  $\mathbf{A}$  according to  $s$ , i.e. for any row  $i$  and column  $q$  of  $\mathbf{A}$ ,  $A_{iq}^s = A_{is(q)}$ . If  $\mathbf{C}$  is a matrix with  $Q$  rows and  $Q$  columns,  $\mathbf{C}^s$  is defined similarly:*

$$\mathbf{A}^s = (A_{is(q)})_{i,q} \quad \mathbf{C}^s = (C_{s(q)s(\ell)})_{q,\ell}$$

**Definition 2.12** (Equivalence). *We define the following equivalence relationships:*

- Two assignments  $\mathbf{z}$  and  $\mathbf{z}'$  are equivalent, noted  $\sim$ , if they are equal up to label permutation, i.e. there exists a permutation  $s$  such that  $\mathbf{z}' = \mathbf{z}^s$ .
- Two parameters  $\boldsymbol{\theta}$  and  $\boldsymbol{\theta}'$  are equivalent, noted  $\sim$ , if they are equal up to label permutation, i.e. there exists a permutation  $s$  such that  $(\boldsymbol{\alpha}^s, \boldsymbol{\pi}^s) = (\boldsymbol{\alpha}', \boldsymbol{\pi}')$ .
- $(\boldsymbol{\theta}, \mathbf{z})$  and  $(\boldsymbol{\theta}', \mathbf{z}')$  are equivalent, noted  $\sim$ , if they are equal up to label permutation on  $\boldsymbol{\pi}$  and  $\mathbf{z}$ , i.e. there exists a permutation  $s$  such that  $(\boldsymbol{\pi}^s, \mathbf{z}^s) = (\boldsymbol{\pi}', \mathbf{z}')$ . This is label-switching.

**Definition 2.13** (Symmetry). *We say that the parameter  $\boldsymbol{\theta}$  exhibits symmetry for the permutation  $s$  if*

$$(\boldsymbol{\alpha}^s, \boldsymbol{\pi}^s) = (\boldsymbol{\alpha}, \boldsymbol{\pi}).$$

$\boldsymbol{\theta}$  exhibits symmetry if it exhibits symmetry for any non trivial permutations  $s$ . Finally the set of permutations for which  $\boldsymbol{\theta}$  exhibits symmetry is noted  $\text{Sym}(\boldsymbol{\theta})$ .

*Remark 2.14.* The set of parameters that exhibit symmetry is a manifold of null Lebesgue measure in  $\Theta$ . The notion of symmetry allows us to deal with a notion of non-identifiability of the class labels that is subtler than and different from label switching. More precisely

$$\begin{aligned} \text{Label switching is when : } p(\mathbf{y}^o, \mathbf{z}, \boldsymbol{\theta}) &= p(\mathbf{y}^o, \mathbf{z}^s, \boldsymbol{\theta}^s), \quad \boldsymbol{\theta} \neq \boldsymbol{\theta}^s \quad \forall s \\ \text{Symmetry is when : } p(\mathbf{y}^o, \mathbf{z}, \boldsymbol{\theta}) &= p(\mathbf{y}^o, \mathbf{z}^s, \boldsymbol{\theta}), \quad \forall s \in \text{Sym}(\boldsymbol{\theta}) \end{aligned}$$

In particular, in label-switching,  $\mathbf{z}$  and  $\mathbf{z}^s$  have the same likelihood but under equivalent yet different parameters  $\boldsymbol{\theta}$ s. In contrast, in the presence of symmetry,  $\mathbf{z}$  and  $\mathbf{z}^s$  have exactly the same likelihood under  $\boldsymbol{\theta}$ . This implies in particular that the posterior  $p(\mathbf{z}|\mathbf{y}^o, \boldsymbol{\theta})$  can not concentrate on a single assignment. This is instrumental for Proposition 6.11.

*Remark 2.15.* In this example we illustrate what  $\text{Sym}(\theta)$  and its cardinal can be in a simple case. Consider a network with  $n$  nodes,

$$\alpha = (1/6, 1/6, 2/3), \text{ and } \pi = \begin{pmatrix} 0 & 0.7 & 0.2 \\ 0.7 & 0 & 0.2 \\ 0.2 & 0.2 & 0.2 \end{pmatrix}.$$

As a consequences the two following assignments

$$\begin{aligned} z_1 &= (\underbrace{1, \dots, 1}_{n_1}, \underbrace{2, \dots, 2}_{n_2}, \underbrace{3, \dots, 3}_{n_3}) \\ z_2 &= (\underbrace{2, \dots, 2}_{n_2}, \underbrace{1, \dots, 1}_{n_1}, \underbrace{3, \dots, 3}_{n_3}) \end{aligned}$$

belong to  $\text{Sym}(\theta) = \{Id, [1, 2]\}$ . Indeed they are the only assignments belonging to  $\text{Sym}(\theta)$ , so, in this particular case  $\#\text{Sym}(\theta) = 2$ .

The issue of symmetry forces us to use a notion of distance between assignment that is invariant to label permutation.

**Definition 2.16** (Distance). *We define the following distance, up to equivalence, between configurations  $\mathbf{z}$  and  $\mathbf{z}^*$ :*

$$\|\mathbf{z} - \mathbf{z}^*\|_{0, \sim} = \inf_{\mathbf{z}' \sim \mathbf{z}} \|\mathbf{z}' - \mathbf{z}^*\|_0$$

where, for all matrix  $\mathbf{z}$ , we use the Hamming norm  $\|\cdot\|_0$  defined by

$$\|\mathbf{z}\|_0 = \frac{1}{2} \sum_{i,q} \mathbb{1}\{z_{iq} \neq 0\}.$$

**Definition 2.17** (Set of local assignments). *We note  $S(\mathbf{z}^*, r)$  the set of configurations that have a representative (for  $\sim$ ) within relative radius  $r$  of  $\mathbf{z}^*$ :*

$$S(\mathbf{z}^*, r) = \{\mathbf{z} : \|\mathbf{z} - \mathbf{z}^*\|_{0, \sim} \leq rn\}$$

## 2.9. Other definitions

We finally introduce a few useful notions that will be instrumental in the proofs. The first is “regular” assignments, for which each group has “enough” nodes:

**Definition 2.18** ( $c$ -regular assignments). *Let  $\mathbf{z} \in \mathcal{Z}$ . For any  $c > 0$ , we say that  $\mathbf{z}$  is  $c$ -regular if*

$$\min_q z_{+q} \geq cn. \tag{2.7}$$

Class distinctness  $\delta(\boldsymbol{\pi})$  captures the differences between groups: lower values of  $\delta(\boldsymbol{\pi})$  means that at least two classes have very similar connectivity profiles.  $\delta(\boldsymbol{\pi})$  is intrinsically linked to the convergence rate of several estimates.

**Definition 2.19** (Class distinctness). For  $\theta = (\alpha, \pi) \in \Theta$ . We define:

$$\delta(\pi) = \min_{q, q'} \max_{\ell} \text{KL}(\pi_{q\ell}, \pi_{q'\ell})$$

with  $\text{KL}(\pi, \pi') = \mathbb{E}_{\pi}[\log(\varphi(Y, \pi)/\varphi(Y, \pi'))] = \psi'(\pi)(\pi - \pi') + \psi(\pi') - \psi(\pi)$  the Kullback divergence between  $\varphi(\cdot, \pi)$  and  $\varphi(\cdot, \pi')$ , when  $\varphi$  comes from an exponential family.

*Remark 2.20.* Since all  $\pi$  have distinct rows and columns,  $\delta(\pi) > 0$ .

Finally, the confusion matrix allows to compare groups between assignments:

**Definition 2.21** (Confusion matrix). For given assignments  $\mathbf{z}$  and  $\mathbf{z}^*$ , we define the confusion matrix between  $\mathbf{z}$  and  $\mathbf{z}^*$ , noted  $\mathbb{R}(\mathbf{z})$ , as follows:

$$\mathbb{R}(\mathbf{z})_{qq'} = \frac{1}{n} \sum_i z_{iq}^* z_{iq'} \tag{2.8}$$

**Definition 2.22.** For more conciseness, we define

$$\mathbf{S}^* = (S_{q\ell}^*)_{q\ell} = (\psi'(\pi_{q\ell}^*))_{q\ell} \tag{2.9}$$

### 3. Complete-observed model

Hereafter and in the rest of the text, we use the term "complete" to say that true assignments  $\mathbf{z}^*$  are known, and "observed" to say that only some dyads are observed.

**Proposition 3.1.** Under random dyad sampling, defining  $N_i = \sum_j r_{ij}$  and  $\Omega_{0,n} = \cap_{i=0}^n \{N_i \geq 1\}$  the set of nodes with at least one dyad observed. Then

$$\mathbb{P} \left( \lim_{n \rightarrow +\infty} \Omega_{0,n} \right) = 1.$$

*Proof.* This proposition is a direct consequence of Borel-Cantelli's theorem. Details are available in appendix A. □

*Remark 3.2.* This result shows that, with high probability, the network has no unobserved node. In the remainder, we work conditionnally on  $\Omega_{0,n}$ .

Let  $\hat{\theta}_c = (\hat{\alpha}, \hat{\pi})$  be the MLE of  $\theta$  in the complete-observed data model. Simple manipulations of Equation (2.3) yield:

$$\begin{aligned} \hat{\alpha}_q &= \hat{\alpha}_q(\mathbf{z}) = \frac{z+q}{n} \\ \hat{y}_{q\ell}(\mathbf{z}) &= \frac{\sum_{i \neq j} y_{ij} r_{ij} z_{iq} z_{j\ell}}{\sum_{i \neq j} r_{ij} z_{iq} z_{j\ell}} \quad \hat{\pi}_{q\ell} = \hat{\pi}_{q\ell}(\mathbf{z}) = (\psi')^{-1}(\hat{y}_{q\ell}(\mathbf{z})) \end{aligned} \tag{3.1}$$

**Proposition 3.3.** Let  $\Sigma_{\alpha^*} = \text{Diag}(\alpha^*) - \alpha^* (\alpha^*)^T$ . Then  $\Sigma_{\alpha^*}$  is semi-definite positive, of rank  $\mathcal{Q} - 1$ , and  $\hat{\alpha}$  is asymptotically normal:

$$\sqrt{n}(\hat{\alpha}(\mathbf{z}^*) - \alpha^*) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \Sigma_{\alpha^*}) \quad (3.2)$$

Similarly, let  $V(\pi^*)$  be the matrix defined by  $[V(\pi^*)]_{q\ell} = 1/\psi''(\pi_{q\ell}^*)$  and  $\Sigma_{\pi^*} = \rho^{-1} \text{Diag}^{-1}(\alpha^*) V(\pi^*) \text{Diag}^{-1}(\alpha^*)$ . Then the estimates  $\hat{\pi}_{q\ell}(\mathbf{z}^*)$  are independent and asymptotically Gaussian with limit distribution:

$$\sqrt{n(n-1)}(\hat{\pi}_{q\ell}(\mathbf{z}^*) - \pi_{q\ell}^*) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \Sigma_{\pi^*, q\ell}) \quad \text{for all } q, \ell \quad (3.3)$$

*Proof.* The proof is postponed to appendix A. The first part is a direct application of central limit theorem for i.i.d. variables and the second part relies on a variant of the central limit theorem for random sums of random variables.  $\square$

*Remark 3.4.* The main differences with Bickel et al. (2013) are (i) the scaling of  $\Sigma_{\alpha^*}$  as  $\rho^{-1}$  and (ii) the need for a central limit theorem for random sums of random variables, as the sums involved in (3.1) are over a random number of terms.

**Proposition 3.5** (Local asymptotic normality). Let  $\mathcal{L}_{co}^*$  be the complete likelihood function defined on  $\Theta$  by  $\mathcal{L}_{co}^*(\alpha, \pi) = \log p(\mathbf{y}^o, \mathbf{z}^*; \theta)$ . For any  $s$  and  $u$  in a compact set, we have:

$$\begin{aligned} \mathcal{L}_{co}^* \left( \alpha^* + \frac{s}{\sqrt{n}}, \pi^* + \frac{u}{\sqrt{n(n-1)}} \right) &= \mathcal{L}_{co}^*(\theta^*) + s^T \mathbf{Y}_{\alpha^*} + \text{Tr}(u^T \mathbf{Y}_{\pi^*}) \\ &- \left( \frac{1}{2} s^T \Sigma_{\alpha^*} s + \frac{1}{2} \text{Tr}((u \odot u)^T \Sigma_{\pi^*}) \right) \\ &+ o_P(1) \end{aligned}$$

where  $\odot$  denote the Hadamard product of two matrices (element-wise product) and  $\Sigma_{\alpha^*}$  and  $\Sigma_{\pi^*}$  are defined in Proposition 3.3.  $\mathbf{Y}_{\alpha^*}$  is asymptotically Gaussian with zero mean and variance matrix  $\Sigma_{\alpha^*}$ .  $\mathbf{Y}_{\pi^*}$  is a random matrix with independent entries that are asymptotically gaussian zero mean and variance  $\Sigma_{\pi^*}$ .

*Proof.* This result is based on a Taylor expansion of  $\mathcal{L}_{co}^*$  in a neighborhood of  $(\alpha^*, \pi^*)$ . Details are available in appendix A.  $\square$

#### 4. Main result

Our main result compares the observed likelihood ratio  $p(\mathbf{y}^o; \theta)/p(\mathbf{y}^o; \theta^*)$  with the complete likelihood  $p(\mathbf{y}^o, \mathbf{z}^*; \theta')/p(\mathbf{y}^o, \mathbf{z}^*; \theta^*)$  to show that they have the same argmax. To ease the comparison, we work only on the high probability set  $\Omega_1$  of  $c/2$ -regular configurations, i.e. that have  $\Omega(n)$  nodes in each group as defined in Section 2,

**Proposition 4.1.** Define  $\mathcal{Z}_1$  as the subset of  $\mathcal{Z}$  made of  $c/2$ -regular assignments, with  $c$  defined in assumption  $H_1$ . Note  $\Omega_1$  the event  $\{\mathbf{z}^* \in \mathcal{Z}_1\}$ , then:

$$\mathbb{P}_{\theta^*}(\bar{\Omega}_1) \leq \mathcal{Q} \exp\left(-\frac{nc^2}{2}\right).$$

*Proof.* This proposition is a consequence of Hoeffding’s inequality. See appendix A for more details.  $\square$

We can now state our main result:

**Theorem 4.2** (Complete-observed). Assume that  $A_1$  to  $A_4$  with random-dyad sampling hold for the Stochastic Block Model of known order with  $n \times n$  observations coming from an univariate exponential family and define  $\#\text{Sym}(\theta)$  as the set of permutation  $s$  for which  $\theta = (\alpha, \pi)$  exhibits symmetry. Then, for  $n$  tending to infinity and  $\rho \gg \log(n)/n$ , the observed likelihood ratio behaves like the complete likelihood ratio, up to a bounded multiplicative factor:

$$\frac{p(\mathbf{y}^o; \theta)}{p(\mathbf{y}^o; \theta^*)} = \frac{\#\text{Sym}(\theta)}{\#\text{Sym}(\theta^*)} \max_{\theta' \sim \theta} \frac{p(\mathbf{y}^o, \mathbf{z}^*; \theta')}{p(\mathbf{y}^o, \mathbf{z}^*; \theta^*)} (1 + o_P(1)) + o_P(1)$$

where the  $o_P$  is uniform over all  $\theta \in \Theta$ .

The maximum over all  $\theta'$  that are equivalent to  $\theta$  stems from the fact that because of label-switching,  $\theta$  is only identifiable up to its  $\sim$ -equivalence class from the observed likelihood, whereas it is completely identifiable from the complete likelihood. The multiplicative factor arises from the fact that equivalent assignments have exactly the same complete likelihood and contribute equally to the observed likelihood.

*Remark 4.3.* This result is very similar to the one of Brault et al. (2020) and corrects an error in the main result of Bickel et al. (2013): the missing terms  $\#\text{Sym}(\theta)$  and  $\#\text{Sym}(\theta^*)$ .

**Corollary 4.4.** If  $\Theta$  contains only parameters with no symmetry:

$$\frac{p(\mathbf{y}^o; \theta)}{p(\mathbf{y}^o; \theta^*)} = \max_{\theta' \sim \theta} \frac{p(\mathbf{y}^o, \mathbf{z}^*; \theta')}{p(\mathbf{y}^o, \mathbf{z}^*; \theta^*)} (1 + o_P(1)) + o_P(1)$$

where the  $o_P$  is uniform over all  $\Theta$ .

## 5. Variational and Maximum Likelihood Estimates

This section is devoted to the asymptotic of the MLE and the VE in the incomplete data model as a consequence of the main result 4.2. Note that, with high probability, both estimators have no symmetry since the set  $\{\theta : \#\text{Sym}(\theta) > 1\}$  is a manifold of null Lebesgue’s measure in  $\Theta$  and thus  $p_{\theta^*}(\#\text{Sym}(\theta) > 1) \rightarrow 0$ .

### 5.1. ML estimator

The asymptotic behavior of the maximum likelihood estimator in the incomplete data model is a direct consequence of Theorem 4.2 and Proposition 3.5.

**Corollary 5.1** (Asymptotic behavior of  $\theta_{MLE}$ ). *Denote  $\theta_{MLE}$  the maximum likelihood estimator and use the notations of Proposition 3.3. There exist permutations  $s$  of  $\{1, \dots, Q\}$  such that*

$$\begin{aligned}\hat{\alpha}(\mathbf{z}^*) - \alpha_{MLE}^s &= o_P(n^{-1/2}), \\ \hat{\pi}(\mathbf{z}^*) - \pi_{MLE}^s &= o_P(n^{-1}).\end{aligned}$$

Hence, the maximum likelihood estimator for the SBM under random-dyad sampling condition is consistent and asymptotically normal, with the same behavior as the maximum likelihood estimator in the complete data model. The proof is postponed to appendix B.10.

### 5.2. Variational estimator

Due to the complex dependency structure of the observations, the maximum likelihood estimator of the SBM is not numerically tractable, even with the *Expectation Maximisation* algorithm. In practice, a variational approximation is often used (see Daudin et al., 2008): for any joint distribution  $\mathbb{Q} \in \mathcal{Q}$  on  $\mathcal{Z}$  a lower bound of  $\mathcal{L}(\theta)$  is given by

$$\begin{aligned}J(\mathbb{Q}, \theta) &= \mathcal{L}(\theta) - KL(\mathbb{Q}, p(\cdot; \theta, \mathbf{y}^o)) \\ &= \mathbb{E}_{\mathbb{Q}}[\mathcal{L}_{co}(\mathbf{z}; \theta)] + \mathcal{H}(\mathbb{Q}).\end{aligned}$$

where  $\mathcal{H}(\mathbb{Q}) = -\mathbb{E}_{\mathbb{Q}}[\log(\mathbb{Q})]$ . Choosing  $\mathcal{Q}$  to be the set of product distributions, such that for all  $\mathbf{z}$

$$\mathbb{Q}(\mathbf{z}) = \prod_{i,q} \mathbb{Q}(z_{iq} = 1)^{z_{iq}}$$

allows us to obtain tractable expressions of  $J(\mathbb{Q}, \theta)$ . The variational estimate  $\hat{\theta}_{var}$  of  $\theta$  is defined as

$$\hat{\theta}_{var} \in \operatorname{argmax}_{\theta \in \Theta} \max_{\mathbb{Q} \in \mathcal{Q}} J(\mathbb{Q}, \theta).$$

The following corollary states that  $\hat{\theta}_{var}$  has the same asymptotic properties as  $\hat{\theta}_{MLE}$  and  $\hat{\theta}_{MC}$ , in particular is consistent and asymptotically normal.

**Corollary 5.2** (Variational estimate). *Under the assumptions of Theorem 4.2, there exist permutations  $s$  of  $\{1, \dots, Q\}$  such that*

$$\begin{aligned}\hat{\alpha}(\mathbf{z}^*) - \hat{\alpha}_{var}^s &= o_P(n^{-1/2}), \\ \hat{\pi}(\mathbf{z}^*) - \hat{\pi}_{var}^s &= o_P(n^{-1}).\end{aligned}$$

The proof is very similar to the proof of Corollary 5.1 and postponed to appendix B.10.

## 6. Proof sketch

The proof of the theorem relies on controlling deviations of the log-likelihood ratios from their expectations. We first define those quantities.

### 6.1. Log-likelihood ratios

**Definition 6.1.** We define the conditional log-likelihood ratio  $LR$  and its expectation  $ELR$  as:

$$LR(\boldsymbol{\theta}, \mathbf{z}) = \log \frac{p(\mathbf{y}^\circ | \mathbf{z}; \boldsymbol{\theta})}{p(\mathbf{y}^\circ | \mathbf{z}^*; \boldsymbol{\theta}^*)} \quad \text{and} \quad ELR(\boldsymbol{\theta}, \mathbf{z}) = \mathbb{E}_{\boldsymbol{\theta}^*} [LR(\boldsymbol{\theta}, \mathbf{z}) | \mathbf{z}^*] \quad (6.1)$$

We also define the profile ratio  $\Lambda$  and its counterpart  $\tilde{\Lambda}$  as:

$$\Lambda(\mathbf{z}) = \max_{\boldsymbol{\theta}} LR(\boldsymbol{\theta}, \mathbf{z}) \quad \text{and} \quad \tilde{\Lambda}(\mathbf{z}) = \max_{\boldsymbol{\theta}} ELR(\boldsymbol{\theta}, \mathbf{z}). \quad (6.2)$$

The following decomposition of  $p(\mathbf{y}^\circ; \boldsymbol{\theta})$  highlights the importance of  $LR(\boldsymbol{\theta}, \mathbf{z})$ :

$$p(\mathbf{y}^\circ; \boldsymbol{\theta}) = \sum_{(\mathbf{z})} p(\mathbf{y}^\circ, \mathbf{z}; \boldsymbol{\theta}) = p(\mathbf{y}^\circ | \mathbf{z}^*; \boldsymbol{\theta}^*) \sum_{(\mathbf{z})} p(\mathbf{z}; \boldsymbol{\theta}) \exp(LR(\boldsymbol{\theta}, \mathbf{z})).$$

Since  $LR(\boldsymbol{\theta}, \mathbf{z}) \leq \Lambda(\mathbf{z})$ , the profile ratio is useful to remove the dependency on  $\boldsymbol{\theta}$  and reduce the study to a series of problems depending only on  $\mathbf{z}$ . The following propositions shows that those quantities are constraints which are maximum (in expectation) at the true parameter value (up to group relabeling) and have negative curvature at those points. This allows us to prove that, asymptotically, only one (or a few)  $z$  contribute to the above sum.

**Proposition 6.2.** Conditionally on  $\mathbf{z}^*$ , we have

$$\bar{y}_{q\ell}(\mathbf{z}) := \mathbb{E}_{\boldsymbol{\theta}^*} [\hat{y}_{q\ell}(\mathbf{z}) | \mathbf{z}^*] = \frac{[\mathbb{R}(\mathbf{z})^T \mathbf{S}^* \mathbb{R}(\mathbf{z})]_{q\ell}}{\hat{\alpha}_q(\mathbf{z}) \hat{\alpha}_\ell(\mathbf{z})} \quad (6.3)$$

with  $\bar{y}_{q\ell}(\mathbf{z}) = 0$  for  $\mathbf{z}$  such that  $\hat{\alpha}_q(\mathbf{z}) = 0$  or  $\hat{\alpha}_\ell(\mathbf{z}) = 0$ , i.e. no dyad observed in class  $(q, \ell)$ .

*Remark 6.3.* Note the absence of the random variable  $\mathbf{r}$  in  $\bar{y}_{q\ell}(\mathbf{z})$ , which is integrated out in the expectation  $\mathbb{E}_{\boldsymbol{\theta}^*}$ .

**Proposition 6.4** (Maximum of  $ELR$  and  $\tilde{\Lambda}$  in  $\boldsymbol{\theta}$ ). The functions  $LR(\boldsymbol{\theta}, \mathbf{z})$  and  $ELR(\boldsymbol{\theta}, \mathbf{z})$  are maximum respectively in  $\boldsymbol{\pi}$  for  $\hat{\boldsymbol{\pi}}(\mathbf{z})$  and  $\bar{\boldsymbol{\pi}}(\mathbf{z})$  defined by:

$$\hat{\boldsymbol{\pi}}(\mathbf{z})_{q\ell} = (\psi')^{-1}(\hat{y}_{q\ell}(\mathbf{z})) \quad \text{and} \quad \bar{\boldsymbol{\pi}}(\mathbf{z})_{q\ell} = (\psi')^{-1}(\bar{y}_{q\ell}(\mathbf{z}))$$

so that

$$\Lambda(\mathbf{z}) = LR(\hat{\boldsymbol{\pi}}(\mathbf{z}), \mathbf{z}) \quad \text{and} \quad \tilde{\Lambda}(\mathbf{z}) = ELR(\bar{\boldsymbol{\pi}}(\mathbf{z}), \mathbf{z}).$$

**Proposition 6.5** (Local upperbound for  $\tilde{\Lambda}$ ). *Conditionally upon  $\Omega_1$ , there exists a positive constant  $C$  such that for all  $\mathbf{z} \in S(\mathbf{z}^*, C)$ :*

$$\tilde{\Lambda}(\mathbf{z}) \leq -c\rho n \frac{3\delta(\boldsymbol{\pi}^*)}{4} \|\mathbf{z} - \mathbf{z}^*\|_{0,\sim} \quad (6.4)$$

**Proposition 6.6** (Maximum of  $ELR$  and  $\tilde{\Lambda}$  in  $(\boldsymbol{\theta}, \mathbf{z})$ ).  *$ELR$  can be written:*

$$ELR(\boldsymbol{\theta}, \mathbf{z}) = -\rho n^2 \sum_{q,q'} \sum_{\ell,\ell'} \mathbb{R}(\mathbf{z})_{q,q'} \mathbb{R}(\mathbf{z})_{\ell,\ell'} \text{KL}(\pi_{q\ell}^*, \pi_{q'\ell'}) \leq 0. \quad (6.5)$$

Conditionally on the set  $\Omega_1$  of regular assignments and for  $n > 2/c$ ,

- (i)  $ELR$  is maximized at  $(\boldsymbol{\pi}^*, \mathbf{z}^*)$  and its equivalence class and  $ELR(\boldsymbol{\pi}^*, \mathbf{z}^*) = 0$ .
- (ii)  $\tilde{\Lambda}$  is maximized at  $\mathbf{z}^*$  and its equivalence class and  $\tilde{\Lambda}(\mathbf{z}^*) = 0$ .
- (iii) The maximum of  $\tilde{\Lambda}$  (and hence the maximum of  $ELR$ ) is well separated.

Proofs of Propositions 6.2, 6.4, 6.5 and 6.6 are postponed to Appendix B.

## 6.2. High level view of the proof

The proof proceeds by splitting  $p(\mathbf{y}^o; \boldsymbol{\theta})$  as a sum over three types of configurations that partition  $\mathcal{Z}$  and studying the asymptotic behavior of  $LR$  and on each type:

1. *global control*: for  $\mathbf{z}$  such that  $\tilde{\Lambda}(\mathbf{z}) = \Omega(-n^2)$ , Proposition 6.7 proves a large deviation behavior and shows that  $LR = -\Omega_P(n^2)$ . In turn, those assignments contribute a  $o_P$  of  $p(\mathbf{y}^o, \mathbf{z}^*; \boldsymbol{\theta}^*)$  to the sum (Proposition 6.8).
2. *local control*: a small deviation result (Proposition 6.9) is needed to show that the combined contribution of assignments close to but not equivalent to  $\mathbf{z}^*$  is also a  $o_P$  of  $p(\mathbf{y}^o, \mathbf{z}^*; \boldsymbol{\theta}^*)$  (Proposition 6.10).
3. *equivalent assignments*: Proposition 6.11 examines which of the remaining assignments, all equivalent to  $\mathbf{z}^*$ , contribute to the sum.

These results are presented in next section 6.3 and their proofs postponed to appendix B. They are then put together in section 6.4 to prove our main result. The remainder of the section is devoted to the asymptotics of the ML and variational estimators as a consequence of the main result.

## 6.3. Different asymptotic behaviors

### 6.3.1. Global control

**Proposition 6.7** (Large deviations of  $LR$ ). *Let  $\text{Diam}(\Theta) = \sup_{\boldsymbol{\theta}, \boldsymbol{\theta}'} \|\boldsymbol{\theta} - \boldsymbol{\theta}'\|_\infty$ . For all  $\varepsilon_n < \nu b$  and  $n$  large enough that  $2\sqrt{2n^2\varepsilon_n} \geq Q^2$*

$$\sup_{\boldsymbol{\theta}, \mathbf{z}} \{LR(\boldsymbol{\theta}, \mathbf{z}) - \tilde{\Lambda}(\mathbf{z})\} = \mathcal{O}_p(n^2\varepsilon_n) \quad (6.6)$$

**Proposition 6.8** (Contribution of global assignments). *Choose  $t_n$  decreasing to 0 slowly enough that  $\frac{\rho n t_n}{\sqrt{\log(n)}} \rightarrow +\infty$ . Then conditionally on  $\Omega_1$  and for  $n$  large enough that  $2\sqrt{2n^2}\epsilon_n \geq \mathcal{Q}^2$ , we have:*

$$\sup_{\theta \in \Theta} \sum_{\mathbf{z} \notin S(\mathbf{z}^*, t_n)} p(\mathbf{z}, \mathbf{y}^o; \theta) = o_P(p(\mathbf{z}^*, \mathbf{y}^o; \theta^*))$$

### 6.3.2. Local control

**Proposition 6.9** (Small deviations LR). *Conditionally on  $\Omega_1$ ,*

$$\sup_{\mathbf{z} \sim \mathbf{z}^*} \frac{\Lambda(\mathbf{z}) - \tilde{\Lambda}(\mathbf{z}^*)}{n \|\mathbf{z} - \mathbf{z}^*\|_{0, \sim}} = o_P(1) \tag{6.7}$$

The next proposition uses Propositions 6.9 and 6.6 to show that the combined contribution to the observed likelihood of assignments close to  $\mathbf{z}^*$  is also a  $o_P$  of  $p(\mathbf{z}^*, \mathbf{y}^o; \theta^*)$ :

**Proposition 6.10** (Contribution of local assignments). *With the previous notations and  $C$  the positive constant defined in Proposition 6.5:*

$$\sup_{\theta \in \Theta} \sum_{\substack{\mathbf{z} \in S(\mathbf{z}^*, C) \\ \mathbf{z} \sim \mathbf{z}^*}} p(\mathbf{z}, \mathbf{y}^o; \theta) = o_P(p(\mathbf{z}^*, \mathbf{y}^o; \theta^*))$$

### 6.3.3. Equivalent assignments

It remains to study the contribution of equivalent assignments.

**Proposition 6.11** (Contribution of equivalent assignments). *For all  $\theta \in \Theta$ , we have*

$$\sum_{\mathbf{z} \sim \mathbf{z}^*} \frac{p(\mathbf{y}^o, \mathbf{z}; \theta)}{p(\mathbf{y}^o, \mathbf{z}^*; \theta^*)} = \# \text{Sym}(\theta) \max_{\theta' \sim \theta} \frac{p(\mathbf{y}^o, \mathbf{z}^*; \theta')}{p(\mathbf{y}^o, \mathbf{z}^*; \theta^*)} (1 + o_P(1))$$

where the  $o_P$  is uniform in  $\theta$ .

## 6.4. Proof of the main result

*Proof.* We work conditionally on  $\Omega_1$ . Choose  $\mathbf{z}^* \in \mathcal{Z}_1$  and a sequence  $t_n$  decreasing to 0 but satisfying  $\frac{\rho n t_n}{\sqrt{\log(n)}} \rightarrow +\infty$ . According to Proposition 6.8,

$$\sup_{\theta \in \Theta} \sum_{\mathbf{z} \notin S(\mathbf{z}^*, t_n)} p(\mathbf{z}, \mathbf{y}^o; \theta) = o_P(p(\mathbf{z}^*, \mathbf{y}^o; \theta^*))$$

Since  $t_n$  decreases to 0, it gets smaller than  $C$  (used in Proposition 6.10) for  $n$  large enough. As this point, Proposition 6.10 ensures that:

$$\sup_{\theta \in \Theta} \sum_{\substack{\mathbf{z} \in S(\mathbf{z}^*, t_n) \\ \mathbf{z} \sim \mathbf{z}^*}} p(\mathbf{z}, \mathbf{y}^o; \theta) = o_P(p(\mathbf{z}^*, \mathbf{y}^o; \theta^*))$$

And therefore the observed likelihood ratio reduces as:

$$\begin{aligned} \frac{p(\mathbf{y}^o; \theta)}{p(\mathbf{y}^o; \theta^*)} &= \frac{\sum_{\mathbf{z} \sim \mathbf{z}^*} p(\mathbf{y}^o, \mathbf{z}; \theta) + \sum_{\mathbf{z} \sim \mathbf{z}^*} p(\mathbf{y}^o, \mathbf{z}; \theta)}{\sum_{\mathbf{z} \sim \mathbf{z}^*} p(\mathbf{y}^o, \mathbf{z}; \theta^*) + \sum_{\mathbf{z} \sim \mathbf{z}^*} p(\mathbf{y}^o, \mathbf{z}; \theta^*)} \\ &= \frac{\sum_{\mathbf{z} \sim \mathbf{z}^*} p(\mathbf{y}^o, \mathbf{z}; \theta) + p(\mathbf{y}^o; \mathbf{z}^*, \theta^*) o_P(1)}{\sum_{\mathbf{z} \sim \mathbf{z}^*} p(\mathbf{y}^o, \mathbf{z}; \theta^*) + p(\mathbf{y}^o; \mathbf{z}^*, \theta^*) o_P(1)} \end{aligned}$$

And Proposition 6.11 allows us to conclude

$$\frac{p(\mathbf{y}^o; \theta)}{p(\mathbf{y}^o; \theta^*)} = \frac{\#\text{Sym}(\theta)}{\#\text{Sym}(\theta^*)} \max_{\theta' \sim \theta} \frac{p(\mathbf{y}^o, \mathbf{z}^*; \theta')}{p(\mathbf{y}^o, \mathbf{z}^*; \theta^*)} (1 + o_P(1)) + o_P(1). \quad \square$$

## 7. Discussion

Close examination of the different proofs, especially of Prop. 6.10, reveals that the quantities driving convergence of the estimates are  $\rho n \delta(\boldsymbol{\pi}^*)$ , which must go to  $+\infty$  with  $n$  to ensure validity of Prop. 6.8, and  $\rho n t_n \delta(\boldsymbol{\pi}^*)$ , which must be larger than  $\sqrt{\log(n)}$  while  $t_n \rightarrow 0$ , to ensure validity of Prop. 6.10. Both conditions are met as soon as  $\rho \gg \log(n)/n$ , allowing for a large fraction of missing edges. Note that this limiting rate for missingness is the same as the one found for graph density in sparse settings to achieve consistency and local asymptotic normality of  $\boldsymbol{\theta}$  (Bickel et al., 2013). It's also the same as the one found by Chatterjee (2015) for the structured matrix reconstruction problem. Note also that in the fixed  $\rho$  setting, both MLE and VE are consistent and asymptotically normal but the cost of missingness is an expected blow up of the asymptotic variance matrix by a factor of  $\rho^{-1}$ .

The proof follows along the line of Bickel et al. (2013) but differs in some significant ways. First, since the number of observed dyads is random, we must rely on variants of the central limit theorem that hold for random sums of random variables. Second, the move from the binary to unbounded dyads invalidates a counting argument used in Bickel et al. (2013) and requires different concentration inequalities. We leverage the facts that random variables with distribution in natural exponential families are subexponential and that the subexponential property is preserved by summation and multiplication to derive Bernstein-type inequality. Finally, we add the missing terms  $\#\text{Sym}(\boldsymbol{\theta})$  which have little impact for the corollaries but are required for the rigorous statement of the main result.

**Appendix A: Technical results**

**A.1. Proof of Proposition 3.1**

*Proof.* Noticing that  $N_i \sim \text{Bin}(n - 1, \rho)$ , then  $\mathbb{P}(N_i \geq 1) = 1 - (1 - \rho)^{n-1}$ . As a consequence  $\mathbb{P}(\overline{\Omega}_{0,n}) \leq \sum_i \mathbb{P}(N_i = 0) = n(1 - \rho)^{n-1} \xrightarrow{n \rightarrow +\infty} 0$ , and  $\mathbb{P}(\Omega_{0,n}) \xrightarrow{n \rightarrow +\infty} 1$ . Then  $\mathbb{P}(\limsup(\overline{\Omega}_{0,n})) = 0$  by Borel-Cantelli theorem (because  $\sum_n \mathbb{P}(\overline{\Omega}_{0,n})$  converge), and as  $\overline{\limsup \overline{\Omega}_{0,n}} = \overline{\bigcap_{n \geq 0} \bigcup_{q \geq n} \overline{\Omega}_{0,n}} = \bigcup_{n \geq 0} \bigcap_{q \geq n} \Omega_{0,n} = \liminf \Omega_{0,n}$ , the result follow.  $\square$

**A.2. Technical Lemma A.1**

**Lemma A.1.**

$$U_n = \frac{1}{n(n-1)} \sum_{i \neq j} r_{ij} z_{iq} z_{jl} \xrightarrow{n \rightarrow +\infty} \rho \alpha_q \alpha_l$$

*Proof.* Noticing that  $\mathbb{E}[r_{ij} z_{iq} z_{jl}] = \rho \alpha_q \alpha_l$  and defining  $q_{i,j}^{q,\ell} = r_{ij} z_{iq} z_{jl} - \rho \alpha_q \alpha_l$ . By Hoeffding decomposition for U-statistics (see Hoeffding, 1948)

$$U'_n = \frac{1}{n(n-1)} \sum_{i \neq j} (r_{ij} z_{iq} z_{jl} - \rho \alpha_q \alpha_l) = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \frac{1}{\lfloor \frac{n}{2} \rfloor} \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} q_{\sigma(i), \sigma(i+\lfloor \frac{n}{2} \rfloor)}^{q,\ell}, \quad (\text{A.1})$$

where for each permutation  $\sigma \in \mathfrak{S}$ ,  $\sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} q_{\sigma(i), \sigma(i+\lfloor \frac{n}{2} \rfloor)}^{q,\ell}$  is a sum of independant r.v. Then, for  $\gamma > 0$  by Jensen's inequality and Hoeffding's lemma about bounded r.v.

$$\begin{aligned} \mathbb{E}[\exp(\gamma U'_n)] &\leq \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \mathbb{E} \exp \left( \frac{\gamma}{\lfloor \frac{n}{2} \rfloor} \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} q_{\sigma(i), \sigma(i+\lfloor \frac{n}{2} \rfloor)}^{q,\ell} \right) \\ &\leq \exp \left( \frac{\gamma^2}{8 \lfloor \frac{n}{2} \rfloor} \right). \end{aligned}$$

Finally, using the same proof than Hoeffding's inequality allows us to conclude.  $\square$

**A.3. Proof of Proposition 3.3**

*Proof.* Since  $\hat{\alpha}(\mathbf{z}^*) = (\hat{\alpha}_1(\mathbf{z}^*), \dots, \hat{\alpha}_g(\mathbf{z}^*))$  is the sample mean of  $n$  i.i.d. multinomial random variables with parameters 1 and  $\alpha^*$ , a simple application of the central limit theorem (CLT) gives:

$$\Sigma_{\alpha^*, qq'} = \begin{cases} \alpha_q^*(1 - \alpha_q^*) & \text{if } q = q' \\ -\alpha_q^* \alpha_{q'}^* & \text{if } q \neq q' \end{cases}$$

which proves Equation (3.2) where  $\Sigma_{\alpha^*}$  is semi-definite positive of rank  $Q - 1$ .

Similarly,  $\psi'(\widehat{\pi}_{q\ell}(\mathbf{z}^*))$  is the average of  $\sum_{i \neq j} r_{ij} z_{iq}^* z_{j\ell}^*$  i.i.d. random variables with mean  $\psi'(\pi_{q\ell}^*)$  and variance  $\psi''(\pi_{q\ell}^*)$ .  $\sum_{i \neq j} r_{ij} z_{iq}^* z_{j\ell}^*$  is itself random but thanks to Lemma A.1:  $\frac{1}{n(n-1)} \sum_{i \neq j} r_{ij} z_{iq}^* z_{j\ell}^* \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \rho \alpha_q^* \alpha_\ell^*$ . Therefore, by Slutsky's lemma and the CLT for random sums of random variables (Shanthikumar and Sumita, 1984), we have:

$$\begin{aligned} & \sqrt{n(n-1)} \rho \alpha_q^* \alpha_\ell^* (\psi'(\widehat{\pi}_{q\ell}(\mathbf{z}^*)) - \psi'(\pi_{q\ell}^*)) \\ &= \sqrt{n(n-1)} \rho \alpha_q^* \alpha_\ell^* \left( \frac{\sum_{i \neq j} y_{ij} r_{ij} z_{iq}^* z_{j\ell}^*}{\sum_{i \neq j} r_{ij} z_{iq}^* z_{j\ell}^*} - \psi'(\pi_{q\ell}^*) \right) \\ & \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N}(0, \psi''(\pi_{q\ell}^*)) \end{aligned}$$

The differentiability of  $(\psi')^{-1}$  and the delta method then gives:

$$\sqrt{n(n-1)} (\widehat{\pi}_{q\ell}(\mathbf{z}^*) - \pi_{q\ell}^*) \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{\rho \alpha_q^* \alpha_\ell^* \psi''(\pi_{q\ell}^*)}\right)$$

and the independence results from the independence of  $\widehat{\pi}_{q\ell}(\mathbf{z}^*)$  and  $\widehat{\pi}_{q'\ell'}(\mathbf{z}^*)$  as soon as  $q \neq q'$  or  $\ell \neq \ell'$ , as they involve different sets of i.i.d. variables.  $\square$

#### A.4. Proof of Proposition 3.5

*Proof.* By Taylor expansion,

$$\begin{aligned} & \mathcal{L}_{co}^* \left( \boldsymbol{\alpha}^* + \frac{s}{\sqrt{n}}, \boldsymbol{\pi}^* + \frac{u}{\sqrt{n(n-1)}} \right) \\ &= \mathcal{L}_{co}^*(\boldsymbol{\theta}^*) + \frac{1}{\sqrt{n}} s^T \nabla \mathcal{L}_{co\boldsymbol{\alpha}}^*(\boldsymbol{\theta}^*) + \frac{1}{\sqrt{n(n-1)}} \text{Tr}(u^T \nabla \mathcal{L}_{co\boldsymbol{\pi}}^*(\boldsymbol{\theta}^*)) \\ & \quad + \frac{1}{n} s^T \mathbf{H}_\alpha(\boldsymbol{\theta}^*) s + \frac{1}{n(n-1)} \text{Tr}((u \odot u)^T \mathbf{H}_\pi(\boldsymbol{\theta}^*)) + o_P(1) \end{aligned}$$

where  $\nabla \mathcal{L}_{co\boldsymbol{\alpha}}^*(\boldsymbol{\theta}^*)$  and  $\nabla \mathcal{L}_{co\boldsymbol{\pi}}^*(\boldsymbol{\theta}^*)$  denote the respective components of the gradient of  $\mathcal{L}_{co}^*$  evaluated at  $\boldsymbol{\theta}^*$  and  $\mathbf{H}_\alpha$  and  $\mathbf{H}_\pi$  denote the conditional hessian of  $\mathcal{L}_{co}^*$  evaluated at  $\boldsymbol{\theta}^*$ . By inspection,  $\mathbf{H}_\alpha/n$  and  $\mathbf{H}_\pi/(n(n-1))$  converge in probability to constant matrices  $\Sigma_\alpha, \Sigma_\pi$  and the random vectors  $\nabla \mathcal{L}_{co\boldsymbol{\alpha}}^*(\boldsymbol{\theta}^*)/\sqrt{n}$  and  $\nabla \mathcal{L}_{co\boldsymbol{\pi}}^*(\boldsymbol{\theta}^*)/\sqrt{n(n-1)}$  converge in distribution by central limit theorem.  $\square$

#### A.5. Proof of Proposition 4.1

*Proof.* In regular configurations, each group has  $\Omega(n)$  members, where  $u_n = \Omega(n)$  if there exists two constant  $a, b > 0$  such that for  $n$  enough large  $an \leq$

$u_n \leq bn$ .  $c/2$ -regular assignments, with  $c$  defined in Assumption  $H_1$ , have high  $\mathbb{P}_{\theta^*}$ -probability in the space of all assignments, uniformly over all  $\theta^* \in \Theta$ .

Each  $z_{+q}$  is a sum of  $n$  i.i.d Bernoulli r.v. with parameter  $\alpha_q \geq \alpha_{\min} \geq c$ . A simple Hoeffding bound shows that

$$\mathbb{P}_{\theta^*} \left( z_{+q} \leq n \frac{c}{2} \right) \leq \mathbb{P}_{\theta^*} \left( z_{+q} \leq n \frac{\alpha_q}{2} \right) \leq \exp \left( -2n \left( \frac{\alpha_q}{2} \right)^2 \right) \leq \exp \left( -\frac{nc^2}{2} \right)$$

taking a union bound over  $\mathcal{Q}$  values of  $q$  leads to Proposition 4.1. □

## Appendix B: Main results

### B.1. Proof of Proposition 6.2

*Proof.* First of all we will prove equation (6.3),

$$\begin{aligned} \bar{y}_{q\ell}(\mathbf{z}) &= \mathbb{E}_{\theta^*} \left[ \frac{\sum_{i \neq j} z_{iq} z_{j\ell} r_{ij} y_{ij}}{\sum_{i \neq j} z_{iq} z_{j\ell} r_{ij}} \middle| \mathbf{z}^* \right] \\ &= \mathbb{E}_{\theta^*} \left[ \mathbb{E}_{\theta^*} \left[ \frac{\sum_{i \neq j} z_{iq} z_{j\ell} r_{ij} y_{ij}}{\sum_{i \neq j} z_{iq} z_{j\ell} r_{ij}} \middle| R, \mathbf{z}^* \right] \middle| \mathbf{z}^* \right] \\ &= \mathbb{E}_{\theta^*} \left[ \frac{\sum_{i \neq j} z_{iq} z_{j\ell} r_{ij} S_{Z_i Z_j}^*}{\sum_{i \neq j} z_{iq} z_{j\ell} r_{ij}} \middle| \mathbf{z}^* \right], \end{aligned}$$

where  $Z_i = q \Leftrightarrow z_{iq} = 1$ . Noticing that the  $(i, j)$  for which  $z_{iq} z_{j\ell} = 0$  does not contribute in any of the two terms of the ratio. The calculus of this expectation is then equivalent to calculate an expectation of the general form

$$\mathbb{E}_{\theta^*} \left[ \frac{\sum_{i=1}^n a_i R_i}{\sum_{i=1}^n R_i} \right], (a_i)_{i \in \{1, \dots, n\}} \in \mathbb{R}^n \text{ and } T_i \stackrel{iid}{\sim} \mathcal{B}(\rho).$$

**Lemma B.1.**

$$\mathbb{E}_{\theta^*} \left[ \frac{\sum_{i=1}^n a_i T_i}{\sum_{i=1}^n T_i} \right] = \frac{\sum_{i=1}^n a_i}{n}.$$

*Proof.* Define  $N = \sum_{i=1}^n T_i$  and noticing that  $\mathbb{E}[T_i | N = k] = \frac{k}{n}$ . Conditionally to  $N \geq 1$

$$\begin{aligned} \mathbb{E} \left[ \frac{\sum_{i=1}^n a_i T_i}{\sum_{i=1}^n T_i} \right] &= \mathbb{E} \left[ \mathbb{E} \left[ \frac{\sum_{i=1}^n a_i T_i}{N} \middle| N \right] \right] \\ &= \frac{\sum_{i=1}^n a_i}{n}. \end{aligned} \quad \square$$

Now, applying Lemma B.1 with  $N_{q\ell}^o(z) = \sum_{i \neq j} z_{iq} z_{j\ell} r_{ij}$  leads to

$$\mathbb{E}_{\theta^*} \left[ \frac{\sum_{i \neq j} z_{iq} z_{j\ell} r_{ij} S_{Z_i Z_j}^*}{\sum_{i \neq j} z_{iq} z_{j\ell} r_{ij}} \middle| \mathbf{z}^*, N_{q\ell}^o(z) \geq 1 \right] = \frac{[\mathbf{R}(\mathbf{z})^T \mathbf{S}^* \mathbf{R}(\mathbf{z})]_{q\ell}}{\hat{\alpha}_q(\mathbf{z}) \hat{\alpha}_\ell(\mathbf{z})} \mathbb{1}_{N_{q\ell}^o(z) \geq 1}.$$

Finally,  $\mathbb{E}_{\theta^*} [\hat{y}_{q\ell}(\mathbf{z}) | \mathbf{z}^*, N_{q\ell}^o(z) = 0]$  can be arbitrarily defined at the same value than  $\mathbb{E}_{\theta^*} [\hat{y}_{q\ell}(\mathbf{z}) | \mathbf{z}^*, N_{q\ell}^o(z) \geq 1]$  which concluded the proof. □

### B.2. Proof of Proposition 6.4

*Proof.* Defining  $\nu(y, \pi) = y\pi - \psi(\pi)$ . For  $y$  fixed,  $\nu(y, \pi)$  is maximized at  $\pi = (\psi')^{-1}(y)$ . Manipulations yield

$$\begin{aligned} LR(\boldsymbol{\theta}, \mathbf{z}) &= \log p(\mathbf{y}^\circ; \mathbf{z}, \boldsymbol{\theta}) - \log p(\mathbf{y}^\circ; \mathbf{z}^*, \boldsymbol{\theta}^*) \\ &= \left[ \sum_q \sum_\ell N_{q\ell}^o(z) \nu(\widehat{y}_{q\ell}(\mathbf{z}), \pi_{q\ell}) - \sum_q \sum_\ell N_{q\ell}^o(z^*) \nu(\widehat{y}_{q\ell}(\mathbf{z}^*), \pi_{q\ell}^*) \right] \end{aligned}$$

which is maximized at  $\pi_{q\ell} = (\psi')^{-1}(\widehat{y}_{q\ell}(\mathbf{z}))$ . Similarly with  $N_{q\ell}(z) = \sum_{i \neq j} z_{iq} z_{j\ell}$ ,

$$\begin{aligned} ELR(\boldsymbol{\theta}, \mathbf{z}) &= \mathbb{E}_{\boldsymbol{\theta}^*} [\log p(\mathbf{y}^\circ; \mathbf{z}, \boldsymbol{\theta}) - \log p(\mathbf{y}^\circ; \mathbf{z}^*, \boldsymbol{\theta}^*) | \mathbf{z}^*] \\ &= \rho \left[ \sum_q \sum_\ell N_{q\ell}(z) \nu(\bar{y}_{q\ell}(\mathbf{z}), \pi_{q\ell}) - \sum_q \sum_\ell N_{q\ell}(z^*) \nu(\psi'(\pi_{q\ell}^*), \pi_{q\ell}^*) \right] \end{aligned}$$

is maximized at  $\pi_{q\ell} = (\psi')^{-1}(\bar{y}_{q\ell}(\mathbf{z}))$ .  $\square$

### B.3. Proof of Proposition 6.6 (maximum of ELR and $\tilde{\Lambda}$ )

*Proof.* We condition on  $\mathbf{z}^*$  and prove Equation (6.5):

$$\begin{aligned} ELR(\boldsymbol{\theta}, \mathbf{z}) &= \mathbb{E}_{\boldsymbol{\theta}^*} \left[ \log \frac{p(\mathbf{y}^\circ; \mathbf{z}, \boldsymbol{\theta})}{p(\mathbf{y}^\circ; \mathbf{z}^*, \boldsymbol{\theta}^*)} \middle| \mathbf{z}^* \right] \\ &= \sum_i \sum_j \sum_{q, q'} \sum_{\ell, \ell'} \mathbb{E}_{\boldsymbol{\theta}^*} [y_{ij}(\pi_{q'\ell'} - \pi_{q\ell}^*) - (\psi(\pi_{q'\ell'}) - \psi(\pi_{q\ell}^*))] \rho z_{iq}^* z_{iq'} z_{j\ell}^* z_{j\ell'} \\ &= n^2 \rho \sum_{q, q'} \sum_{\ell, \ell'} \mathbb{R}(\mathbf{z})_{q, q'} \mathbb{R}(\mathbf{z})_{\ell, \ell'} [\psi'(\pi_{q\ell}^*)(\pi_{q'\ell'} - \pi_{q\ell}^*) + \psi(\pi_{q\ell}^*) - \psi(\pi_{q'\ell'})] \\ &= -n^2 \rho \sum_{q, q'} \sum_{\ell, \ell'} \mathbb{R}(\mathbf{z})_{q, q'} \mathbb{R}(\mathbf{z})_{\ell, \ell'} \text{KL}(\pi_{q\ell}^*, \pi_{q'\ell'}) \end{aligned}$$

If  $\mathbf{z}^*$  is regular, and for  $n > 2/c$ , all the rows of  $\mathbb{R}(\mathbf{z})$  have at least one positive element and we can apply Lemma 3.2 of Bickel et al. (2013) to characterize the maximum for  $ELR$ .

The maximality of  $\tilde{\Lambda}(\mathbf{z}^*)$  results from the fact that  $\tilde{\Lambda}(\mathbf{z}) = ELR(\bar{\boldsymbol{\pi}}(\mathbf{z}), \mathbf{z})$  where  $\bar{\boldsymbol{\pi}}(\mathbf{z})$  is a particular value of  $\boldsymbol{\pi}$ ,  $\tilde{\Lambda}$  is immediately maximum at  $\mathbf{z} \sim \mathbf{z}^*$ , and for those, we have  $\bar{\boldsymbol{\pi}}(\mathbf{z}) \sim \boldsymbol{\pi}^*$ .

The separation and local behavior of  $G$  around  $\mathbf{z}^*$  is a direct consequence of Proposition 6.5.  $\square$

### B.4. Proof of Proposition 6.5 (local upper bound for $\tilde{\Lambda}$ )

*Proof.* We work conditionally on  $\mathbf{z}^*$ . The principle of the proof relies on the extension of  $\tilde{\Lambda}$  to a continuous subspace of  $\mathcal{M}_{\mathcal{Q}}([0, 1])$ , in which the confusion

matrix is naturally embedded. The regularity assumption allows us to work on a subspace that is bounded away from the borders of  $\mathcal{M}_{\mathcal{Q}}([0, 1])$ . The proof then proceeds by computing the gradient of  $\tilde{\Lambda}$  at and around its argmax and using those gradients to control the local behavior of  $\tilde{\Lambda}$  around its argmax. The local behavior allows us in turn to show that  $\tilde{\Lambda}$  is well-separated.

Note that  $\tilde{\Lambda}$  only depends on  $\mathbf{z}$  through  $\mathbb{R}(\mathbf{z})$ . We can therefore extend it to matrix  $U \in \mathcal{U}_c$  where  $\mathcal{U}$  is the subset of matrices  $\mathcal{M}_{\mathcal{Q}}([0, 1])$  with each row sum higher than  $c/2$ .

$$\tilde{\Lambda}(U) = -\rho n^2 \sum_{q,q'} \sum_{\ell,\ell'} U_{qq'} U_{\ell\ell'} \text{KL}(\pi_{q\ell}^*, \bar{\pi}_{q'\ell'})$$

where

$$\bar{\pi}_{q\ell} = \bar{\pi}_{q\ell}(U) = (\psi')^{-1} \left( \frac{[U^T \mathbf{S}^* U]_{q\ell}}{[U^T \mathbf{1} U]_{q\ell}} \right)$$

and  $\mathbf{1}$  is the  $\mathcal{Q} \times \mathcal{Q}$  matrix filled with 1. Confusion matrix  $\mathbb{R}(\mathbf{z})$  satisfy  $\mathbb{R}(\mathbf{z})\mathbb{I} = \boldsymbol{\alpha}(\mathbf{z}^*)$ , with  $\mathbb{I} = (1, \dots, 1)^T$  a vector only containing 1 values, and are obviously in  $\mathcal{U}_c$  as soon as  $\mathbf{z}^*$  is  $c/2$  regular.

The maps  $f_{q,q',\ell,\ell'} : (U) \mapsto \text{KL}(\pi_{q\ell}^*, \bar{\pi}_{q'\ell'})$  are twice differentiable with second derivatives bounded over  $\mathcal{U}_c$  and therefore so is  $\tilde{\Lambda}(U)$ . Tedious but straightforward computations show that the derivative of  $\tilde{\Lambda}$  at  $D_\alpha := \text{Diag}(\boldsymbol{\alpha}(\mathbf{z}^*))$  is:

$$A_{qq'}(\mathbf{z}^*) := \frac{-1}{n^2} \frac{\partial \tilde{\Lambda}}{\partial U_{qq'}}(D_\alpha) = 2\rho \sum_{\ell} \alpha_{\ell}(\mathbf{z}^*) \text{KL}(\pi_{q\ell}^*, \pi_{q'\ell}^*)$$

$A(\mathbf{z}^*)$  is the matrix-derivative of  $-\tilde{\Lambda}/n^2$  at  $D_\alpha$ . Since  $\mathbf{z}^*$  is  $c/2$ -regular and by definition of  $\delta(\boldsymbol{\pi}^*)$ ,  $A(\mathbf{z}^*)_{qq'} \geq c\rho\delta(\boldsymbol{\pi}^*)$  if  $q \neq q'$  and  $A(\mathbf{z}^*)_{qq} = 0$  for all  $q$ . By boundedness of the second derivative, there exists  $C > 0$  such that for all  $D_\alpha$  and all  $H \in B(D_\alpha, C)$ , we have:

$$\frac{-1}{n^2} \frac{\partial \tilde{\Lambda}}{\partial U_{qq'}}(H) \begin{cases} \geq \rho \frac{7c\delta(\boldsymbol{\pi}^*)}{8} & \text{if } q \neq q' \\ \leq \rho \frac{c\delta(\boldsymbol{\pi}^*)}{8} & \text{if } q = q' \end{cases}$$

Choose  $U$  in  $\mathcal{U}_c \cap B(D_\alpha, C)$  satisfying  $U\mathbb{I} = \boldsymbol{\alpha}(\mathbf{z}^*)$ .  $U - D_\alpha$  have nonnegative off diagonal coefficients and negative diagonal coefficients. Furthermore, the coefficients of  $U, D_\alpha$  sum up to 1 and  $\text{Tr}(D_\alpha) = 1$ . By Taylor expansion, there exists  $H$  also in  $\mathcal{U}_c \cap B(D_\alpha, C)$  such that

$$\begin{aligned} \frac{-1}{n^2} \tilde{\Lambda}(U) &= \frac{-1}{n^2} \tilde{\Lambda}(D_\alpha) + \text{Tr} \left( (U - D_\alpha) \frac{-1}{n^2} \frac{\partial \tilde{\Lambda}}{\partial U}(H) \right) \\ &\geq \rho \frac{c\delta(\boldsymbol{\pi}^*)}{8} \left[ 7 \sum_{q \neq q'} (U - D_\alpha)_{qq'} - \sum_q (U - D_\alpha)_{qq} \right] \\ &= c\rho \frac{3\delta(\boldsymbol{\pi}^*)}{4} (1 - \text{Tr}(U)) \end{aligned}$$

To conclude the proof, assume without loss of generality that  $\mathbf{z} \in S(\mathbf{z}^*, C)$  achieves the  $\|\cdot\|_{0,\sim}$  norm (i.e. it is the closest to  $\mathbf{z}^*$  in its representative class). Then  $U = \mathbb{R}(\mathbf{z})$  is in  $(\mathcal{U}_c \cap B(D_\alpha, C))$  and satisfy  $U\Pi = \boldsymbol{\alpha}(\mathbf{z}^*)$ . We just need to note  $n(1 - \text{Tr}(\mathbb{R}(\mathbf{z}))) = \|\mathbf{z} - \mathbf{z}^*\|_{0,\sim}$  to end the proof.  $\square$

**B.5. Proof of Proposition 6.7 (global convergence LR)**

*Proof.* Conditionally upon  $\mathbf{z}^*$ ,

$$\begin{aligned} LR(\boldsymbol{\theta}, \mathbf{z}) - \tilde{\Lambda}(\mathbf{z}) &\leq LR(\boldsymbol{\theta}, \mathbf{z}) - ELR(\boldsymbol{\theta}, \mathbf{z}) \\ &= \sum_i \sum_j (\pi_{z_i z_j} - \pi_{z_i^* z_j^*}) \left( y_{ij} r_{ij} - \psi'(\pi_{z_i^* z_j^*}) \rho \right) \\ &\quad + \sum_i \sum_j (\psi(\pi_{z_i z_j}) - \psi(\pi_{z_i^* z_j^*})) (r_{ij} - \rho) \\ &= \sum_{qq'} \sum_{\ell\ell'} (\pi_{q'\ell'} - \pi_{q\ell}^*) W_{qq'\ell\ell'} \\ &\leq \sup_{\substack{\Gamma \in \mathbb{R}^{\mathcal{Q}^2 \times \mathcal{Q}^2} \\ \|\Gamma\|_\infty \leq \text{Diam}(\Theta)}} \sum_{qq'} \sum_{\ell\ell'} \Gamma_{qq'\ell\ell'} W_{qq'\ell\ell'} := Z \end{aligned}$$

uniformly in  $\boldsymbol{\theta}$ , where the  $W_{qq'\ell\ell'}$  are independent and by Taylor expansion defined by:

$$\begin{aligned} W_{qq'\ell\ell'} &= \sum_i \sum_j z_{iq}^* z_{jl}^* z_{i,q'} z_{j\ell'} (y_{ij} r_{ij} - \psi'(\pi_{q\ell}^*) \rho - (r_{ij} - \rho) C_{qq'\ell\ell'}) , \\ C_{qq'\ell\ell'} &\in \psi'(\Theta) \end{aligned}$$

is the sum of  $n^2 \mathbb{R}(\mathbf{z})_{qq'} \mathbb{R}(\mathbf{z})_{\ell\ell'}$  sub-exponential variables with parameters  $(\nu^2, 1/b)$  and is therefore itself sub-exponential with parameters  $(n^2 \mathbb{R}(\mathbf{z})_{qq'} \mathbb{R}(\mathbf{z})_{\ell\ell'} \nu^2, 1/b)$ . According to Proposition B.3 of Brault et al. (2020),  $\mathbb{E}_{\boldsymbol{\theta}^*}[Z|\mathbf{z}^*] \leq Q^2 \text{Diam}(\Theta) \sqrt{n^2 \nu^2}$  and  $Z$  is sub-exponential with parameters  $(n^2 \text{Diam}(\Theta)^2 (2\sqrt{2})^2 \nu^2, 2\sqrt{2} \text{Diam}(\Theta)/b)$ . In particular, for all  $\varepsilon_n < \nu b$

$$\begin{aligned} \mathbb{P}_{\boldsymbol{\theta}^*} \left( Z \geq \nu Q^2 \text{Diam}(\Theta) n \left\{ 1 + \frac{\sqrt{8n^2 \varepsilon_n}}{Q^2} \right\} \middle| \mathbf{z}^* \right) \\ \leq \mathbb{P}_{\boldsymbol{\theta}^*} \left( Z \geq \mathbb{E}_{\boldsymbol{\theta}^*}[Z|\mathbf{z}^*] + \nu \text{Diam}(\Theta) n^2 2\sqrt{2} \varepsilon_n \middle| \mathbf{z}^* \right) \\ \leq \exp \left( -\frac{n^2 \varepsilon_n^2}{2} \right) \end{aligned}$$

We can then remove the conditioning and take a union bound.  $\square$

**B.6. Proof of Proposition 6.8 (contribution of far away assignments)**

*Proof.* Conditionally on  $\mathbf{z}^*$ , we know from Proposition 6.6 that  $\tilde{\Lambda}$  is maximal in  $\mathbf{z}^*$  and its equivalence class. Choose  $0 < t_n$  decreasing to 0 but satisfying

$\frac{n\rho t_n}{\sqrt{\log(n)}} \rightarrow +\infty$ . According to 6.6 (iii), for all  $\mathbf{z} \notin S(\mathbf{z}^*, t_n)$

$$\tilde{\Lambda}(\mathbf{z}) \leq -c\rho n \frac{3\delta(\boldsymbol{\pi}^*)}{4} \|\mathbf{z} - \mathbf{z}^*\|_{0,\sim} \leq -c\rho \frac{3\delta(\boldsymbol{\pi}^*)}{4} n^2 t_n \tag{B.1}$$

since  $\|\mathbf{z} - \mathbf{z}^*\|_{0,\sim} \geq nt_n$ .

Set  $\varepsilon_n = \inf(5c\rho\delta(\boldsymbol{\pi}^*)t_n/(\sqrt{2}\nu \text{Diam}(\boldsymbol{\Theta})), \nu b)$  and  $n$  large enough that  $\varepsilon_n \geq \frac{Q^2}{n\sqrt{8}}$ . By Proposition 6.7, and with our choice of  $\varepsilon_n$ , with probability higher than  $1 - \Delta_n^1(\varepsilon_n)$ ,

$$\begin{aligned} & \sum_{\mathbf{z} \notin S(\mathbf{z}^*, t_n)} p(\mathbf{y}^\circ, \mathbf{z}; \boldsymbol{\theta}) \\ &= p(\mathbf{y}^\circ | \mathbf{z}^*, \boldsymbol{\theta}^*) \sum_{\mathbf{z} \notin S(\mathbf{z}^*, t_n)} p(\mathbf{z}; \boldsymbol{\theta}) e^{LR(\boldsymbol{\theta}, \mathbf{z}) - \tilde{\Lambda}(\mathbf{z}) + \tilde{\Lambda}(\mathbf{z}^*)} \\ &\leq p(\mathbf{y}^\circ | \mathbf{z}^*, \boldsymbol{\theta}^*) \sum_{\mathbf{z}} p(\mathbf{z}; \boldsymbol{\theta}) e^{LR(\boldsymbol{\theta}, \mathbf{z}) - \tilde{\Lambda}(\mathbf{z}) - 3n^2 t_n c\rho\delta(\boldsymbol{\pi}^*)/4} \\ &\leq p(\mathbf{y}^\circ | \mathbf{z}^*, \boldsymbol{\theta}^*) \sum_{\mathbf{z}} p(\mathbf{z}; \boldsymbol{\theta}) e^{-n^2 t_n c\rho\delta(\boldsymbol{\pi}^*)/8} \\ &= \frac{p(\mathbf{y}^\circ, \mathbf{z}^*; \boldsymbol{\theta}^*)}{p(\mathbf{z}^*; \boldsymbol{\theta}^*)} e^{-n^2 t_n c\rho\delta(\boldsymbol{\pi}^*)/8} \\ &\leq p(\mathbf{y}^\circ, \mathbf{z}^*; \boldsymbol{\theta}^*) \exp\left(-n^2 t_n \frac{c\rho\delta(\boldsymbol{\pi}^*)}{8} + n \log \frac{1}{c}\right) \\ &= p(\mathbf{y}^\circ, \mathbf{z}^*; \boldsymbol{\theta}^*) o(1) \end{aligned}$$

where the second line comes from inequality (B.1), the third from the global control studied in Proposition 6.7 and the definition of  $\varepsilon_n$ , the fourth from the definition of  $p(\mathbf{y}^\circ, \mathbf{z}^*; \boldsymbol{\theta}^*)$ , the fifth from the bounds on  $\boldsymbol{\alpha}^*$  and the last from  $\frac{n\rho t_n}{\sqrt{\log(n)}} \rightarrow +\infty$ .

In addition, with our choice of  $t_n$ , we have  $\varepsilon_n \gg \sqrt{\log(n)}/n$  so that the series  $\sum_n \Delta_n^1(\varepsilon_n)$  converges and:

$$\sum_{\mathbf{z} \notin S(\mathbf{z}^*, t_n)} p(\mathbf{y}^\circ, \mathbf{z}; \boldsymbol{\theta}) = p(\mathbf{y}^\circ; \mathbf{z}^*, \boldsymbol{\theta}^*) o_P(1) \quad \square$$

**B.7. Proof of Proposition 6.9 (local convergence LR)**

*Proof.* We work conditionally on  $\mathbf{z}^* \in \mathcal{Z}_1$ . Choose  $\varepsilon \leq \kappa \varrho^2$  small. Assignments  $\mathbf{z}$  at  $\|\cdot\|_{0,\sim}$ -distance less than  $c/4$  of  $\mathbf{z}^*$  are  $c/4$ -regular. According to Proposition B.1 of Brault et al. (2020),  $\hat{y}_{q\ell}$  and  $\bar{y}_{q\ell}$  are at distance at most  $\varepsilon$  with probability higher than  $1 - 2 \exp\left(-\frac{n^2 c^2 \varepsilon^2}{32(\nu^2 + b^{-1}\varepsilon)}\right)$ . Defining

$$\tilde{\tilde{\Lambda}}(\mathbf{z}) = \sum_q \sum_\ell N_{q\ell}^o(\mathbf{z}) \nu(\bar{y}_{q\ell}(\mathbf{z}), \pi_{q\ell}) - \sum_q \sum_\ell N_{q\ell}^o(\mathbf{z}^*) \nu(\psi'(\pi_{q\ell}^*), \pi_{q\ell}^*),$$

where  $\tilde{\Lambda}(\mathbf{z}) = \mathbb{E} [\tilde{\tilde{\Lambda}}(\mathbf{z})|\mathbf{z}^*]$ . Manipulation of  $\Lambda$ ,  $\tilde{\Lambda}$  and  $\tilde{\tilde{\Lambda}}$  yield

$$\begin{aligned} \frac{\Lambda(\mathbf{z}) - \tilde{\Lambda}(\mathbf{z})}{n^2} &\leq \frac{\Lambda(\mathbf{z}) - \tilde{\tilde{\Lambda}}(\mathbf{z})}{n^2} + \frac{\tilde{\tilde{\Lambda}}(\mathbf{z}) - \tilde{\Lambda}(\mathbf{z})}{n^2} \\ &= \frac{1}{n^2} \sum_q \sum_\ell (N_{q\ell}^o(\mathbf{z}) [f(\hat{y}_{q\ell}) - f(\bar{y}_{q\ell})] - N_{q\ell}^o(\mathbf{z}^*) \pi_{q\ell}^* (\hat{y}_{q\ell}^* - \bar{y}_{q\ell}^*)) \\ &\quad + \frac{1}{n^2} \sum_q \sum_\ell f(\bar{y}_{q\ell}^*) \underbrace{[N_{q\ell}^o(\mathbf{z}) - N_{q\ell}^o(\mathbf{z}^*) - \rho(N_{q\ell}(\mathbf{z}) - N_{q\ell}(\mathbf{z}^*))]}_{=A_{q\ell}} \\ &\quad + \frac{1}{n^2} \sum_q \sum_\ell [N_{q\ell}^o(\mathbf{z}) - \rho N_{q\ell}(\mathbf{z})] (f(\bar{y}_{q\ell}^*) - f(\bar{y}_{q\ell})) \end{aligned}$$

where  $f(x) = x(\psi')^{-1}(x) - \psi \circ (\psi')^{-1}(x)$ ,  $\hat{y}_{q\ell}^* = \hat{y}_{q\ell}(\mathbf{z}^*)$  and  $\bar{y}_{q\ell}^* = \psi'(\pi_{q\ell}^*)$ .

**Concerning the first term** The function  $f$  is twice differentiable on  $\mathcal{A}$  with  $f'(x) = (\psi')^{-1}(x)$  and  $f''(x) = 1/\psi'' \circ (\psi')^{-1}(x)$ .  $f'$  (resp.  $f''$ ) are bounded over  $I = \psi'(C_\pi)$  by  $C_\pi$  (resp.  $1/\sigma^2$ ) so that:

$$f(\hat{y}_{q\ell}) - f(\bar{y}_{q\ell}) = f'(\bar{y}_{q\ell})(\hat{y}_{q\ell} - \bar{y}_{q\ell}) + \Omega((\hat{y}_{q\ell} - \bar{y}_{q\ell})^2)$$

By Proposition B.1 (adapted for SBM) of Brault et al. (2020),  $(\hat{y}_{q\ell} - \bar{y}_{q\ell})^2 = \mathcal{O}_P(1/n^2)$  where the  $\mathcal{O}_P$  is uniform in  $\mathbf{z}$  and does not depend on  $\mathbf{z}^*$ . Similarly,

$$f'(\bar{y}_{q\ell}) = f'(\bar{y}_{q\ell}^*) + \Omega(\bar{y}_{q\ell} - \bar{y}_{q\ell}^*) = \pi_{q\ell}^* + \Omega(\bar{y}_{q\ell} - \bar{y}_{q\ell}^*)$$

$\bar{y}_{q\ell}$  is a convex combination of the  $S_{q\ell}^* = \psi'(\pi_{q\ell}^*)$  therefore,

$$\begin{aligned} |\bar{y}_{q\ell} - \bar{y}_{q\ell}^*| &= \left| \frac{[\mathbb{R}(\mathbf{z})^T \mathbf{S}^* \mathbb{R}(\mathbf{z})]_{q\ell}}{\hat{\alpha}_q(\mathbf{z}) \hat{\alpha}_\ell(\mathbf{z})} - \bar{y}_{q\ell}^* \right| \\ &\leq \left( 1 - \frac{\mathbb{R}(\mathbf{z})_{qq} \mathbb{R}(\mathbf{z})_{\ell\ell}}{\hat{\alpha}_q(\mathbf{z}) \hat{\alpha}_\ell(\mathbf{z})} \right) (S_{\max}^* - S_{\min}^*) \end{aligned}$$

Note that:

$$\begin{aligned} \sum_{q,\ell} N_{q\ell}^o(\mathbf{z}) \left( 1 - \frac{\mathbb{R}(\mathbf{z})_{qq} \mathbb{R}(\mathbf{z})_{\ell\ell}}{\hat{\alpha}_q(\mathbf{z}) \hat{\alpha}_\ell(\mathbf{z})} \right) &= n^2 \rho(1 + o_P(1)) \sum_{q,\ell} [1 - \mathbb{R}(\mathbf{z})_{qq} \mathbb{R}(\mathbf{z})_{\ell\ell}] \\ &= n^2 \rho(1 + o_P(1)) [1 - \text{Tr}(\mathbb{R}(\mathbf{z}))^2] \\ &\leq n \rho(1 + o_P(1)) 2 \|\mathbf{z} - \mathbf{z}^*\|_{0,\sim} \end{aligned}$$

and  $\hat{y}_{q\ell} - \bar{y}_{q\ell} = o_P(1)$ . Therefore

$$\frac{1}{n^2} \sum_{q,\ell} N_{q\ell}^o(\mathbf{z}) \Omega(\bar{y}_{q\ell} - \bar{y}_{q\ell}^*) \times (\hat{y}_{q\ell} - \bar{y}_{q\ell}) = o_P \left( \frac{\|\mathbf{z} - \mathbf{z}^*\|_{0,\sim}}{n} \right)$$

The remaining term writes

$$\frac{1}{n^2} \sum_{q,\ell} \pi_{q\ell}^* [N_{q\ell}^o(\mathbf{z})(\hat{y}_{q\ell} - \bar{y}_{q\ell}) - N_{q\ell}^o(\mathbf{z}^*)(\hat{y}_{q\ell}^* - \bar{y}_{q\ell}^*)]$$

and is also  $o_P(\|\mathbf{z} - \mathbf{z}^*\|_{0,\sim}/n)$  uniformly in  $\mathbf{z}$  and  $\mathbf{z}^* \in \Omega_1$  by Proposition C.2.

**Concerning the second term** For all  $q, \ell$ , defining

$$\begin{cases} N_{q\ell}^+(\mathbf{z}, \mathbf{z}^*) = n^2 \sum_{q'} \mathbb{R}(\mathbf{z})_{qq'}(\mathbf{z}) \sum_{\ell'} \mathbb{R}(\mathbf{z})_{\ell\ell'}(\mathbf{z}) - n^2 \mathbb{R}(\mathbf{z})_{qq} \mathbb{R}(\mathbf{z})_{\ell\ell} \\ N_{q\ell}^-(\mathbf{z}, \mathbf{z}^*) = n^2 \sum_q \mathbb{R}(\mathbf{z})_{qq'}(\mathbf{z}) \sum_{\ell} \mathbb{R}(\mathbf{z})_{\ell\ell'}(\mathbf{z}) - n^2 \mathbb{R}(\mathbf{z})_{qq} \mathbb{R}(\mathbf{z})_{\ell\ell} \end{cases}$$

and noticing that  $N_{q\ell}^+(\mathbf{z}, \mathbf{z}^*) = \#\{(i, j) : z_{iq} = 1, z_{j\ell} = 1, (z_{q\ell}, z_{j\ell}) \neq (z_{q\ell}^*, z_{j\ell}^*)\}$  and  $N_{q\ell}^-(\mathbf{z}, \mathbf{z}^*) = \#\{(i, j) : z_{iq}^* = 1, z_{j\ell}^* = 1, (z_{q\ell}, z_{j\ell}) \neq (z_{q\ell}^*, z_{j\ell}^*)\}$ . Using the following notations

$$\hat{\rho}_{q\ell}^+ = \frac{1}{N_{q\ell}^+(\mathbf{z}, \mathbf{z}^*)} \sum_{(i,j) \in N_{q\ell}^+(\mathbf{z}, \mathbf{z}^*)} R_{ij}, \quad \hat{\rho}_{q\ell}^- = \frac{1}{N_{q\ell}^-(\mathbf{z}, \mathbf{z}^*)} \sum_{(i,j) \in N_{q\ell}^-(\mathbf{z}, \mathbf{z}^*)} R_{ij}$$

we are able to write

$$\begin{aligned} A_{q\ell} &= \sum_{\substack{i < j \\ z_{iq}=1, z_{j\ell}=1}} (R_{ij} - \rho) - \sum_{\substack{i < j \\ z_{iq}^*=1, z_{j\ell}^*=1}} (R_{ij} - \rho) \\ &= N_{q\ell}^+(\mathbf{z}, \mathbf{z}^*) (\hat{\rho}_{q\ell}^+ - \rho) - N_{q\ell}^-(\mathbf{z}, \mathbf{z}^*) (\hat{\rho}_{q\ell}^- - \rho). \end{aligned}$$

Where the second equality is the sum of independent random variables. Note that:

$$\begin{aligned} \sum_{q\ell} N_{q\ell}^+(\mathbf{z}, \mathbf{z}^*) &= \sum_{q\ell} N_{q\ell}^-(\mathbf{z}, \mathbf{z}^*) \\ &= n^2 \sum_{q,\ell} [1 - \mathbb{R}(\mathbf{z})_{qq} \mathbb{R}(\mathbf{z})_{\ell\ell}] \\ &= n^2 [1 - \text{Tr}(\mathbb{R}(\mathbf{z}))]^2 \\ &\leq n2 \|\mathbf{z} - \mathbf{z}^*\|_{0,\sim} \end{aligned}$$

also that  $\hat{\rho}_{q\ell}^+ - \rho = o_P(1)$  and  $\hat{\rho}_{q\ell}^- - \rho = o_P(1)$ . Therefore

$$\frac{1}{n^2} \sum_q \sum_{\ell} f(\bar{y}_{q\ell}^*) A_{q\ell} = o_P\left(\frac{\|\mathbf{z} - \mathbf{z}^*\|_{0,\sim}}{n}\right).$$

**Concerning the third term** Using arguments developed previously leads to the same conclusion than before:

$$\frac{1}{n^2} \sum_q \sum_{\ell} [N_{q\ell}^o(\mathbf{z}) - \rho N_{q\ell}(\mathbf{z})] (f(\bar{y}_{q\ell}^*) - f(\bar{y}_{q\ell})) = o_P\left(\frac{\|\mathbf{z} - \mathbf{z}^*\|_{0,\sim}}{n}\right).$$

As a conclusion, writing

$$\sup_{\mathbf{z} \sim \mathbf{z}^*} \frac{\Lambda(\mathbf{z}) - \tilde{\Lambda}(\mathbf{z}^*)}{n \|\mathbf{z} - \mathbf{z}^*\|_{0,\sim}} = \sup_{\mathbf{z} \sim \mathbf{z}^*} \left( \frac{\Lambda(\mathbf{z}) - \tilde{\Lambda}(\mathbf{z})}{n \|\mathbf{z} - \mathbf{z}^*\|_{0,\sim}} + \frac{\tilde{\Lambda}(\mathbf{z}) - \tilde{\Lambda}(\mathbf{z}^*)}{n \|\mathbf{z} - \mathbf{z}^*\|_{0,\sim}} \right)$$

and noticing that  $\frac{\tilde{\Lambda}(\mathbf{z}) - \tilde{\Lambda}(\mathbf{z}^*)}{n \|\mathbf{z} - \mathbf{z}^*\|_{0,\sim}} \leq 0$  since  $\tilde{\Lambda}$  is maximized in  $\mathbf{z}^*$  (see 6.6). We have

$$\sup_{\mathbf{z} \sim \mathbf{z}^*} \frac{\Lambda(\mathbf{z}) - \tilde{\Lambda}(\mathbf{z}^*)}{n \|\mathbf{z} - \mathbf{z}^*\|_{0,\sim}} = o_P(1). \quad \square$$

### B.8. Proof of Proposition 6.10 (contribution of local assignments)

*Proof.* By Proposition 4.1, it is enough to prove that the sum is small compared to  $p(\mathbf{z}^*, \mathbf{y}^o; \boldsymbol{\theta}^*)$  on  $\Omega_1$ . We work conditionally on  $\mathbf{z}^* \in \mathcal{Z}_1$ . Choose  $\mathbf{z}$  in  $S(\mathbf{z}^*, C)$  with  $C$  defined in Proposition 6.8.

$$\log \left( \frac{p(\mathbf{z}, \mathbf{y}^o; \boldsymbol{\theta})}{p(\mathbf{z}^*, \mathbf{y}^o; \boldsymbol{\theta}^*)} \right) = \log \left( \frac{p(\mathbf{z}; \boldsymbol{\theta})}{p(\mathbf{z}^*; \boldsymbol{\theta}^*)} \right) + LR(\boldsymbol{\theta}, \mathbf{z})$$

For  $C$  small enough, we can assume without loss of generality that  $\mathbf{z}$  is the representative closest to  $\mathbf{z}^*$  and note  $r = \|\mathbf{z} - \mathbf{z}^*\|_0$ . Then:

$$\begin{aligned} LR(\boldsymbol{\theta}, \mathbf{z}) &\leq \Lambda(\mathbf{z}) - \tilde{\Lambda}(\mathbf{z}) + \tilde{\Lambda}(\mathbf{z}) \\ &\leq \Lambda(\mathbf{z}) - \tilde{\Lambda}(\mathbf{z}) - c\rho \frac{3\delta(\boldsymbol{\pi}^*)}{4} nr \\ &\leq c\rho \frac{3\delta(\boldsymbol{\pi}^*)}{4} nr(1 + o_P(1)) \end{aligned}$$

where the first line comes from the definition of  $\Lambda$ , the second line from Proposition 6.6 and the third from Proposition 6.9. Thanks to Proposition D.1, we also know that:

$$\log \left( \frac{p(\mathbf{z}; \boldsymbol{\theta})}{p(\mathbf{z}^*; \boldsymbol{\theta}^*)} \right) \leq \mathcal{O}_P(1) \exp \{M_{c/4} r\}$$

There are at most  $\binom{n}{r} Q^r$  assignments  $\mathbf{z}$  at distance  $r$  of  $\mathbf{z}^*$  and each of them has at most  $Q^Q$  equivalent configurations. Therefore,

$$\begin{aligned} &\frac{\sum_{\substack{\mathbf{z} \in S(\mathbf{z}^*, \bar{c}) \\ \mathbf{z} \sim \mathbf{z}^*}} p(\mathbf{z}, \mathbf{y}^o; \boldsymbol{\theta})}{p(\mathbf{z}^*, \mathbf{y}^o; \boldsymbol{\theta}^*)} \\ &\leq \mathcal{O}_P(1) \sum_{r \geq 1} \binom{n}{r} Q^{Q+r} \exp \left( r M_{c/4} - c\rho \frac{3\delta(\boldsymbol{\pi}^*)}{4} nr(1 + o_P(1)) \right) \\ &= \mathcal{O}_P(1) \left( 1 + e^{(Q+1) \log Q + M_{c/4} - c\rho n \frac{3\delta(\boldsymbol{\pi}^*)(1+o_P(1))}{4}} \right)^n - 1 \\ &\leq \mathcal{O}_P(1) a_n \exp(a_n) \end{aligned}$$

where  $a_n = n e^{(Q+1) \log Q + M_{c/4} - c\rho n \frac{3\delta(\boldsymbol{\pi}^*)(1+o_P(1))}{4}} = o_P(1)$ .  $\square$

### B.9. Proof of Proposition 6.11 (contribution of equivalent assignments)

*Proof.* Choose  $s$  permutations of  $\{1, \dots, Q\}$  and assume that  $\mathbf{z} = \mathbf{z}^{*,s}$ . Then  $p(\mathbf{y}^o, \mathbf{z}; \boldsymbol{\theta}) = p(\mathbf{y}^o, \mathbf{z}^{*,s}; \boldsymbol{\theta}) = p(\mathbf{y}^o, \mathbf{z}^*; \boldsymbol{\theta}^s)$ . If furthermore  $s \in \text{Sym}(\boldsymbol{\theta})$ ,  $\boldsymbol{\theta}^s = \boldsymbol{\theta}$  and immediately  $p(\mathbf{y}^o, \mathbf{z}; \boldsymbol{\theta}) = p(\mathbf{y}^o, \mathbf{z}^*; \boldsymbol{\theta})$ . We can therefore partition the sum as

$$\sum_{\mathbf{z} \sim \mathbf{z}^*} p(\mathbf{y}^o, \mathbf{z}; \boldsymbol{\theta}) = \sum_s p(\mathbf{y}^o, \mathbf{z}^{*,s}; \boldsymbol{\theta})$$

$$\begin{aligned}
 &= \sum_s p(\mathbf{y}^\circ, \mathbf{z}^*; \boldsymbol{\theta}^s) \\
 &= \sum_{\boldsymbol{\theta}' \sim \boldsymbol{\theta}} \# \text{Sym}(\boldsymbol{\theta}') p(\mathbf{y}^\circ, \mathbf{z}^*; \boldsymbol{\theta}') \\
 &= \# \text{Sym}(\boldsymbol{\theta}) \sum_{\boldsymbol{\theta}' \sim \boldsymbol{\theta}} p(\mathbf{y}^\circ, \mathbf{z}^*; \boldsymbol{\theta}')
 \end{aligned}$$

$p(\mathbf{y}^\circ, \mathbf{z}^*; \boldsymbol{\theta})$  unimodal in  $\boldsymbol{\theta}$ , with a mode in  $\widehat{\boldsymbol{\theta}}_{MC}$ . By consistency of  $\widehat{\boldsymbol{\theta}}_{MC}$ , either  $p(\mathbf{y}^\circ, \mathbf{z}^*; \boldsymbol{\theta}) = o_P(p(\mathbf{y}^\circ, \mathbf{z}^*; \boldsymbol{\theta}^*))$  or  $p(\mathbf{y}^\circ, \mathbf{z}^*; \boldsymbol{\theta}) = \mathcal{O}_P(p(\mathbf{y}^\circ, \mathbf{z}^*; \boldsymbol{\theta}^*))$  and  $\boldsymbol{\theta} \rightarrow \boldsymbol{\theta}^*$ . In the latter case, any  $\boldsymbol{\theta}' \sim \boldsymbol{\theta}$  other than  $\boldsymbol{\theta}$  is bounded away from  $\boldsymbol{\theta}^*$  and thus  $p(\mathbf{y}^\circ, \mathbf{z}^*; \boldsymbol{\theta}') = o_P(p(\mathbf{y}^\circ, \mathbf{z}^*; \boldsymbol{\theta}^*))$ . In summary,

$$\sum_{\boldsymbol{\theta}' \sim \boldsymbol{\theta}} \frac{p(\mathbf{y}^\circ, \mathbf{z}^*; \boldsymbol{\theta}')}{p(\mathbf{y}^\circ, \mathbf{z}^*; \boldsymbol{\theta}^*)} = \max_{\boldsymbol{\theta}' \sim \boldsymbol{\theta}} \frac{p(\mathbf{y}^\circ, \mathbf{z}^*; \boldsymbol{\theta}')}{p(\mathbf{y}^\circ, \mathbf{z}^*; \boldsymbol{\theta}^*)} (1 + o_P(1)) \quad \square$$

**B.10. Proof of Corollary 5.1: Behavior of  $\widehat{\boldsymbol{\theta}}_{MLE}$**

We may prove the corollary by contradiction. Note first that unless  $\Theta$  is constrained and with high probability,  $\widehat{\boldsymbol{\theta}}_{MLE}$  and  $\widehat{\boldsymbol{\theta}}(\mathbf{z}^*)$  exhibit no symmetries. Indeed, equalities like  $\widehat{y}_{q\ell} = \widehat{y}_{q'\ell'}$  have vanishingly small probabilities of being simultaneously true when  $y_{ij}$  is discrete, and even null when  $y_{ij}$  is continuous. Assume then  $\min_s (\widehat{\boldsymbol{\alpha}}_{MLE}^s - \widehat{\boldsymbol{\alpha}}(\mathbf{z}^*)) \neq o_P(1/\sqrt{n})$  or  $\min_s (\widehat{\boldsymbol{\pi}}_{MLE}^s - \widehat{\boldsymbol{\pi}}(\mathbf{z}^*)) \neq o_P(1/n)$  where  $s$  is a permutation of  $\{1, \dots, Q\}$ . Then, by Proposition 3.5 and the consistency of  $\widehat{\boldsymbol{\theta}}(\mathbf{z}^*)$

$$\min_s \mathcal{L}_{co}^* \left( \widehat{\boldsymbol{\theta}}(\mathbf{z}^*) \right) - \mathcal{L}_{co}^* \left( \widehat{\boldsymbol{\theta}}_{MLE}^s \right) = \Omega_P(1). \tag{B.2}$$

But, since  $\widehat{\boldsymbol{\theta}}(\mathbf{z}^*)$  and  $\widehat{\boldsymbol{\theta}}_{MLE}$  maximise respectively  $\frac{p(\mathbf{y}^\circ, \mathbf{z}^*; \boldsymbol{\theta}')}{p(\mathbf{y}^\circ, \mathbf{z}^*; \boldsymbol{\theta}^*)}$  and  $\frac{p(\mathbf{y}^\circ; \boldsymbol{\theta})}{p(\mathbf{y}^\circ; \boldsymbol{\theta}^*)}$  and have no symmetries, it follows by Theorem 4.2 that

$$\left| \frac{p(\mathbf{y}^\circ, \mathbf{z}^*; \widehat{\boldsymbol{\theta}}(\mathbf{z}^*))}{p(\mathbf{y}^\circ, \mathbf{z}^*; \boldsymbol{\theta}^*)} - \max_s \frac{p(\mathbf{y}^\circ, \mathbf{z}^*; \widehat{\boldsymbol{\theta}}_{MLE}^s)}{p(\mathbf{y}^\circ, \mathbf{z}^*; \boldsymbol{\theta}^*)} \right| = o_P(1)$$

which contradicts Eq (B.2) and concludes the proof.

**B.11. Proof of Corollary 5.2: Behavior of  $J(\mathbb{Q}, \boldsymbol{\theta})$**

Remark first that for every  $\boldsymbol{\theta}$  and for every  $\mathbf{z}$ ,

$$p(\mathbf{y}^\circ, \mathbf{z}; \boldsymbol{\theta}) \leq \exp [J(\delta_{\mathbf{z}}, \boldsymbol{\theta})] \leq \max_{\mathbb{Q} \in \mathbb{Q}} \exp [J(\mathbb{Q}, \boldsymbol{\theta})] \leq p(\mathbf{y}^\circ; \boldsymbol{\theta})$$

where  $\delta_{\mathbf{z}}$  denotes the dirac mass on  $\mathbf{z}$ . By dividing by  $p(\mathbf{y}^\circ; \boldsymbol{\theta}^*)$ , we obtain

$$\frac{p(\mathbf{y}^\circ, \mathbf{z}; \boldsymbol{\theta})}{p(\mathbf{y}^\circ; \boldsymbol{\theta}^*)} \leq \frac{\max_{\mathbb{Q} \in \mathbb{Q}} \exp [J(\mathbb{Q}, \boldsymbol{\theta})]}{p(\mathbf{y}^\circ; \boldsymbol{\theta}^*)} \leq \frac{p(\mathbf{y}^\circ; \boldsymbol{\theta})}{p(\mathbf{y}^\circ; \boldsymbol{\theta}^*)}.$$

As this inequality is true for every couple  $\mathbf{z}$ , we have in particular:

$$\max_{\mathbf{z} \sim \mathbf{z}^*} \frac{p(\mathbf{y}^o; \mathbf{z}; \boldsymbol{\theta})}{p(\mathbf{y}^o; \boldsymbol{\theta}^*)} = \max_{\boldsymbol{\theta}' \sim \boldsymbol{\theta}} \frac{p(\mathbf{y}^o; \mathbf{z}^*; \boldsymbol{\theta}')}{p(\mathbf{y}^o; \boldsymbol{\theta}^*)} \leq \frac{\max_{\mathbb{Q} \in \mathcal{Q}} \exp[J(\mathbb{Q}, \boldsymbol{\theta})]}{p(\mathbf{y}^o; \boldsymbol{\theta}^*)}.$$

Noticing that  $p(\mathbf{y}^o; \boldsymbol{\theta}^*) = \#\text{Sym}(\boldsymbol{\theta}^*)p(\mathbf{y}^o, \mathbf{z}^*; \boldsymbol{\theta}^*)(1 + o_p(1))$ , Theorem 4.2 therefore leads to the following bounds:

$$\begin{aligned} \max_{\boldsymbol{\theta}' \sim \boldsymbol{\theta}} \frac{p(\mathbf{y}^o, \mathbf{z}^*; \boldsymbol{\theta}')}{p(\mathbf{y}^o, \mathbf{z}^*; \boldsymbol{\theta}^*)} (1 + o_P(1)) &\leq \frac{\max_{\mathbb{Q} \in \mathcal{Q}} \exp[J(\mathbb{Q}, \boldsymbol{\theta})]}{p(\mathbf{y}^o, \mathbf{z}^*; \boldsymbol{\theta}^*)} \\ &\leq \#\text{Sym}(\boldsymbol{\theta}) \max_{\boldsymbol{\theta}' \sim \boldsymbol{\theta}} \frac{p(\mathbf{y}^o, \mathbf{z}^*; \boldsymbol{\theta}')}{p(\mathbf{y}^o, \mathbf{z}^*; \boldsymbol{\theta}^*)} (1 + o_P(1)) + o_P(1). \end{aligned}$$

Again unless  $\Theta$  is constrained,  $\widehat{\boldsymbol{\theta}}_{VAR}$  exhibits no symmetries with high probability and the same proof by contradiction as in appendix B.10 gives the result.

### Appendix C: Sub-exponential random variables

We now prove two propositions regarding subexponential variables. Recall first that a random variable  $X$  is sub-exponential with parameters  $(\tau^2, b)$  if for all  $\lambda$  such that  $|\lambda| \leq 1/b$ ,

$$\mathbb{E}[e^{\lambda(X - \mathbb{E}(X))}] \leq \exp\left(\frac{\lambda^2 \tau^2}{2}\right).$$

In particular, all distributions coming from a natural exponential family are sub-exponential. Sub-exponential variables satisfy a large deviation Bernstein-type inequality:

$$\mathbb{P}(X - \mathbb{E}[X] \geq t) \leq \begin{cases} \exp\left(-\frac{t^2}{2\tau^2}\right) & \text{if } 0 \leq t \leq \frac{\tau^2}{b} \\ \exp\left(-\frac{t}{2b}\right) & \text{if } t \geq \frac{\tau^2}{b} \end{cases} \quad (\text{C.1})$$

So that

$$\mathbb{P}(X - \mathbb{E}[X] \geq t) \leq \exp\left(-\frac{t^2}{2(\tau^2 + bt)}\right)$$

The subexponential property is preserved by summation and multiplication.

- If  $X$  is sub-exponential with parameters  $(\tau^2, b)$  and  $\alpha \in \mathbb{R}$ , then so is  $\alpha X$  with parameters  $(\alpha^2 \tau^2, \alpha b)$
- If the  $X_i$ ,  $i = 1, \dots, n$  are sub-exponential with parameters  $(\tau_i^2, b_i)$  and independent, then so is  $X = X_1 + \dots + X_n$  with parameters  $(\sum_i \tau_i^2, \max_i b_i)$

**Theorem C.1** (Equivalent characterizations of sub-exponential variables). *For a zero-mean random variable  $X$ , the following statements are equivalent:*

1. *There are non-negative numbers  $(\nu, b^{-1})$  such that*

$$\mathbb{E}[e^{\lambda X}] \leq \exp\left(\frac{\lambda^2 \nu^2}{2}\right) \quad \text{for all } |\lambda| < b.$$

2. There is a positive number  $c_0 > 0$  such that  $\mathbb{E}[e^{\lambda X}] < \infty$  for all  $|\lambda| < c_0$ .  
 3. There are constants  $c_1, c_2 > 0$  such that

$$\mathbb{P}(|X| \geq t) \leq c_1 e^{-c_2 t} \quad \text{for all } t > 0.$$

4. The quantity  $\gamma := \sup_{k \geq 2} \left[ \frac{\mathbb{E}[X^k]}{k!} \right]^{1/k}$  is finite.

*Proof.* A proof of this theorem can be found in Wainwright (2015).  $\square$

**Proposition C.2** (Maximum in  $\mathbf{z}$ ). *Let  $(\bar{\mathbf{z}}$  be any configuration and  $\mathbf{z}$  the  $\sim$ -equivalent configuration that achieves  $\|\mathbf{z} - \mathbf{z}^*\|_0 = \|\bar{\mathbf{z}} - \mathbf{z}^*\|_{0, \sim}$  let  $\hat{y}_{q\ell} = \hat{y}_{q,\ell}(\mathbf{z})$  (resp.  $\bar{y}_{q\ell}(\mathbf{z})$ ) and  $\hat{y}_{q\ell}^* = \hat{y}_{q,\ell}(\mathbf{z}^*)$  (resp.  $\bar{y}_{q\ell}^* = \bar{y}_{q\ell}(\mathbf{z}^*) = \psi'(\pi_{q\ell}^*)$ ) be as defined in Equations (3.1) and (6.3). Under the assumptions of the section 2.5, for all  $\varepsilon \leq \kappa \bar{\sigma}^2$ ,*

$$\mathbb{P} \left( \max_{\bar{\mathbf{z}} \sim \mathbf{z}^*} \max_{k,\ell} \frac{N_{q\ell}^o(\mathbf{z})(\hat{y}_{q,\ell} - \bar{y}_{q\ell}) - N_{q\ell}^o(\mathbf{z}^*)(\hat{y}_{q\ell}^* - \bar{y}_{q\ell}^*)}{n \|\mathbf{z} - \mathbf{z}^*\|_0} > \varepsilon \right) = o(1)$$

*Proof.* Note  $r = \|\mathbf{z} - \mathbf{z}^*\|_0$ . The numerator within the max in the fraction can be expanded to

$$Z_{q\ell}(\mathbf{z}) = \sum_{i,j} (z_{iq}z_{j\ell} - z_{iq}^*z_{j\ell}^*) (y_{ij}r_{ij} - \pi_{z_{iq}^*z_{j\ell}^*} \rho)$$

and is thus a sum of at most  $N = nr$  non-null centered subexponential random variables with parameters  $(a^2, 1/w)$ . It is therefore a centered subexponential with parameters  $(Na^2, 1/w)$ . By Bernstein inequality, for all  $\varepsilon \leq \kappa a^2$  we have

$$\mathbb{P}(Z \geq \varepsilon nr) \leq \exp \left( -\frac{nr\varepsilon^2}{2a^2} \right).$$

There are at most  $n^r Q^r Q^Q$   $\mathbf{z}$  at  $\|\cdot\|_{0, \sim}$  distance  $r$  of  $\mathbf{z}^*$ . A union bound shows that:

$$\begin{aligned} \mathbb{P} \left( \max_{\bar{\mathbf{z}} \sim \mathbf{z}^*} \max_{q,\ell} \frac{Z_{q\ell}(\mathbf{z})}{n \|\mathbf{z} - \mathbf{z}^*\|_0} \geq \varepsilon \right) &\leq \sum_{r \geq 1} \sum_{r = \|\bar{\mathbf{z}} - \mathbf{z}^*\|_{0, \sim}} Q^2 \mathbb{P}(Z_{q\ell}(\mathbf{z}) \geq \varepsilon nr) \\ &\leq \sum_{r \geq 1} Q^Q \exp(-nr\varepsilon^2/2a^2 + r \log(nQ) + 2 \log(Q)) = o(1) \end{aligned}$$

where the last equality is true as soon as  $n\varepsilon_n \gg \log n$ .  $\square$

## Appendix D: Likelihood ratio of assignments

**Proposition D.1.** *Let  $\mathbf{z}^*$  be  $c/2$ -regular and  $\mathbf{z}$  at  $\|\cdot\|_0$ -distance  $c/4$  of  $\mathbf{z}^*$ . Then, for all  $\boldsymbol{\theta} \in \Theta$*

$$\log \frac{p(\mathbf{z}; \boldsymbol{\theta})}{p(\mathbf{z}^*; \boldsymbol{\theta}^*)} \leq \mathcal{O}_P(1) \exp \{ M_{c/4} \|\mathbf{z} - \mathbf{z}^*\|_0 \}$$

*Proof.* Note then that:

$$\begin{aligned}
 \frac{p(\mathbf{z}; \boldsymbol{\theta})}{p(\mathbf{z}^*; \boldsymbol{\theta}^*)} &= \frac{p(\mathbf{z}; \boldsymbol{\alpha})}{p(\mathbf{z}^*; \boldsymbol{\alpha}^*)} = \frac{p(\mathbf{z}; \boldsymbol{\alpha})}{p(\mathbf{z}^*; \hat{\boldsymbol{\alpha}}(\mathbf{z}^*))} \frac{p(\mathbf{z}^*; \hat{\boldsymbol{\alpha}}(\mathbf{z}^*))}{p(\mathbf{z}^*; \boldsymbol{\alpha}^*)} \\
 &\leq \frac{p(\mathbf{z}; \hat{\boldsymbol{\alpha}}(\mathbf{z}))}{p(\mathbf{z}^*; \hat{\boldsymbol{\alpha}}(\mathbf{z}^*))} \frac{p(\mathbf{z}^*; \hat{\boldsymbol{\alpha}}(\mathbf{z}^*))}{p(\mathbf{z}^*; \boldsymbol{\alpha}^*)} \\
 &\leq \exp \{M_{c/4} \|\mathbf{z} - \mathbf{z}^*\|_0\} \times \frac{p(\mathbf{z}^*; \hat{\boldsymbol{\alpha}}(\mathbf{z}^*))}{p(\mathbf{z}^*; \boldsymbol{\alpha}^*)} \\
 &\leq \mathcal{O}_P(1) \exp \{M_{c/4} \|\mathbf{z} - \mathbf{z}^*\|_0\}
 \end{aligned}$$

where the first inequality comes from the definition of  $\hat{\boldsymbol{\alpha}}(\mathbf{z})$  and the second from Lemma B.6 of Brault et al. (2020) and the fact that  $\mathbf{z}^*$  and  $\mathbf{z}$  are  $c/4$ -regular. Finally, local asymptotic normality of the MLE for multinomial proportions ensures that  $\frac{p(\mathbf{z}^*; \hat{\boldsymbol{\alpha}}(\mathbf{z}^*))}{p(\mathbf{z}^*; \boldsymbol{\alpha}^*)} = \mathcal{O}_P(1)$ .  $\square$

## Acknowledgment

The authors thank Pierre Barbillon (INRAE-MIA, AgroParisTech), Julien Chiquet (INRAE-MIA, AgroParisTech), Stéphane Robin (INRAE-MIA, AgroParisTech) and James Ridgway (CFM) for their helpful remarks and suggestions.

This work is supported by two public grants overseen by the French National research Agency (ANR): first as part of the «Investissement d’Avenir» program, through the «IDI 2017» project funded by the IDEX Paris-Saclay, ANR-11-IDEX-0003-02, and second by the «EcoNet» project.

## References

- C. Aicher, A. Z. Jacobs, and A. Clauset. Learning latent block structure in weighted networks. *J. Compl. Net.*, 3(2):221–248, 2014. [MR3365464](#)
- 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 (Statistical Methodology)*, 74(1):3–35, 2012. [MR2885838](#)
- P. Barbillon, S. Donnet, E. Lazega, and A. Bar-Hen. Stochastic block models for multiplex networks: an application to networks of researchers. *J. R. Stat. Soc. C-Appl.*, 2015. [MR3600512](#)
- P. Bickel, D. Choi, X. Chang, H. Zhang, et al. Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *The Annals of Statistics*, 41(4):1922–1943, 2013. [MR3127853](#)
- V. Brault, C. Keribin, and M. Mariadassou. Consistency and asymptotic normality of latent blocks model estimators. *Electronic Journal of Statistics*, 14(1):1234–1268, 2020. [MR4079457](#)
- S. Chatterjee. Matrix estimation by universal singular value thresholding. *The Annals of Statistics*, 43(1):177–214, 2015. [MR3285604](#)

- A. Celisse, J.-J. Daudin, L. Pierre, et al. Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electronic Journal of Statistics*, 6:1847–1899, 2012. [MR2988467](#)
- D. S. Choi, P. J. Wolfe, and E. M. Airoidi. Stochastic blockmodels with growing number of classes. *Biometrika*, 99 2:273–284, 2012. [MR2931253](#)
- J.-J. Daudin, F. Picard, and S. Robin. A mixture model for random graphs. *Stat. Comp.*, 18(2):173–183, 2008. [MR2390817](#)
- P. Erdős and A. Rényi. On random graphs. *Publicationes Mathematicae*, 6:290–297, 1959. [MR0120167](#)
- O. Frank and F. Harary. Cluster inference by using transitivity indices in empirical graphs. *J. Am. Stat. Soc.*, 77(380):835–840, 1982. [MR0686407](#)
- M. S. Handcock and K. J. Gile. Modeling social networks from sampled data. *The Annals of Applied Statistics*, 4(1):5–25, 2010. [MR2758082](#)
- W. Hoeffding. A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19(3):293–325, 1948. [MR0026294](#)
- P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983. [MR0718088](#)
- J. Hu, H. Qin, T. Yan, and Y. Zhao. On consistency of model selection for stochastic block models. [arXiv:1611.01238](#), 2017. [MR3217945](#)
- E. D. Kolaczyk. *Statistical Analysis of Network Data, Methods and Models*. Springer, 2009. [MR2724362](#)
- P. Latouche, É. Birmelé, and C. Ambroise. Variational bayesian inference and complexity control for stochastic block models. *Stat. Modelling*, 12(1):93–115, 2012. [MR2953099](#)
- M. Mariadassou and C. Matias. Convergence of the groups posterior distribution in latent or stochastic block models. *Bernoulli*, 21(1):537–573, 2015. [MR3322330](#)
- M. Mariadassou, S. Robin, and C. Vacher. Uncovering latent structure in valued graphs: A variational approach. *Ann. Appl. Stat.*, 4(2):715–742, 06 2010. [MR2758646](#)
- C. Matias and S. Robin. Modeling heterogeneity in random graphs through latent space models: a selective review. *ESAIM Proc. Sur.*, 47:55–74, 2014. [MR3419385](#)
- K. Nowicki and T. A. B. Snijders. Estimation and prediction for stochastic blockstructures. *J. Am. Stat. Soc.*, 96(455):1077–1087, September 2001. [MR1947255](#)
- K. Rohe, S. Chatterjee, and B. Yu. Spectral clustering and the high-dimensional stochastic block model. *Ann. Stat.*, 2010. [MR2893856](#)
- D. B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976. [MR0455196](#)
- J. Shanthikumar and U. Sumita. A central limit theorem for random sums of random variables. *Operations Research Letters*, 3(3):153–155, 1984. [https://doi.org/10.1016/0167-6377\(84\)90008-7](https://doi.org/10.1016/0167-6377(84)90008-7). [MR0761508](#)
- T. A. Snijders and K. Nowicki. Estimation and prediction for stochastic block-models for graphs with latent block structure. *J. Class.*, 14(1):75–100, 1997. [MR1449742](#)

- T. Tabouy, P. Barbillon, and J. Chiquet. Variational inference for stochastic block models from sampled data. *Journal of the American Statistical Association*, 115(529):455–466, 2020. <https://doi.org/10.1080/01621459.2018.1562934>. MR4078475
- M. J. Wainwright. Basic tail and concentration bounds. <https://www.stat.berkeley.edu/~mhwain/stat210b/Chap2/>, 2015.
- Y. X. R. Wang and P. J. Bickel. Likelihood-based model selection for stochastic block models. *Ann. Statist.*, 45(2):500–528, 04 2017. <https://doi.org/10.1214/16-AOS1457>. MR3650391
- S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Structural Analysis in the Social Sciences. Cambridge University Press, 1994. <https://doi.org/10.1017/CB09780511815478>.
- W. W. Zachary. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33(4):452–473, 1977. <https://doi.org/10.1086/jar.33.4.3629752>.