A Proofs

In this section, we prove Theorem 2. This proof follows to some extend that of Theorem 3, so we underline the main differences. Because of missing links, we introduce new techniques to compare the restricted and unrestricted maximum likelihood estimators. We also need to establish the strong consistency of the maximum likelihood estimator for the conditional SBM (in the full observation setting, this result is a direct consequence of [8]). Similarly, the proof of Theorem 3 relies heavily on the fact that the likelihood function at the parameters and the profile likelihood function at the parameters are asymptotically equivalent, which is a direct consequence of Lemma 3 [7]. This result does not hold under missing observations, and we develop new arguments to prove the strong consistency of the variational estimate of the labels.

A.1 Proof of Theorem 2

To prove Theorem 2, we first show that $\mathbb{P}\left(\cdot|\boldsymbol{X}\odot\boldsymbol{A},\widehat{\alpha}^{Var},\widehat{\boldsymbol{Q}}^{Var}\right)$, i.e. the posterior distribution of z at the variational estimator $(\widehat{\alpha}^{Var},\widehat{\boldsymbol{Q}}^{Var})$, concentrates around $\delta_{z'}$, the dirac distribution at some label function z' such that $z'\sim z^*$:

$$\mathbb{P}\left(z'|\boldsymbol{X}\odot\boldsymbol{A},\widehat{\alpha}^{Var},\widehat{\boldsymbol{Q}}^{Var}\right)=1-o_p(1). \tag{11}$$

Then, we show that it implies the concentration of the estimator $\widehat{z}^{Var}$:

$$\mathbb{P}\left(\widehat{z}^{Var}=z'|\boldsymbol{X}\odot\boldsymbol{A}\right)=1-o_p(1). \tag{12}$$

Since $\mathbb{P}\left(\widehat{z}^{Var}=z'|\boldsymbol{X}\odot\boldsymbol{A}\right)$ is bounded, this also implies that it converges to 1 in expectation :

$$\mathbb{P}\left(\widehat{z}^{Var}=z'\right)\to 1. \tag{13}$$

Finally, we show that with probability going to one,

$$\mathbb{P}\left(\widehat{z}\sim z^*\right)\to 1. \tag{14}$$

Combing Equations (12) and (14), we prove the first part of Theorem 2 :

$$\mathbb{P}\left(\widehat{z}\sim\widehat{z}^{VAR}\right)\to 1. \tag{15}$$

To establish the second part of Theorem 2, we show that the maximum likelihood estimator defined in (9) is equal to the restricted maximum estimator (4). Theorem 3 then follows from Theorem 1.

Define $c_{min}=\min_{a,b}\boldsymbol{Q}^*_{a,b}$ and $c_{max}=\max_{a,b}\boldsymbol{Q}^*_{a,b}$. Theorem 1 implies that for some absolute constant $C>0$,

$$\mathbb{P}\left(\left\|\boldsymbol{\Theta}^*-\widehat{\boldsymbol{\Theta}}^r\right\|_2^2\le C(c_{max}/c_{min})^2\left(k^2+n\log(k)\right)\right)\to 1,$$

where the restricted maximum likelihood estimator $\widehat{\boldsymbol{\Theta}}^r$ is defined as

$$\begin{aligned}&\widehat{\boldsymbol{\Theta}}^r_{i<j}=\widehat{\boldsymbol{Q}}^r_{\widehat{z}^r(i)\widehat{z}^r(j)},\ \widehat{\boldsymbol{\Theta}}^r_{ii}=0\\ &(\widehat{\boldsymbol{Q}}^r,\widehat{z}^r)\in\underset{\boldsymbol{Q}\in[c_{min}/2,2c_{max}]^{k\times k}_{\text{sym}},z\in\mathcal{Z}_{n,k}}{\arg\max}\sum_{i\neq j}\mathcal{L}_{\boldsymbol{X}}(\boldsymbol{A}_{ij},\boldsymbol{Q}_{z(i)z(j)}).\end{aligned}$$

Now, Equation (15) implies that with probability going to one, the variational estimator of the probabilities of connections $\widehat{\boldsymbol{\Theta}}^{VAR}$ is equal to the maximum likelihood estimator $\widehat{\boldsymbol{\Theta}}$ given by

$$\begin{aligned}&\widehat{\boldsymbol{\Theta}}_{i<j}=\widehat{\boldsymbol{Q}}_{\widehat{z}(i)\widehat{z}(j)},\ \widehat{\boldsymbol{\Theta}}_{ii}=0\\ &\text{for }(\widehat{\boldsymbol{Q}},\widehat{z})\in\underset{\boldsymbol{Q}\in\mathcal{Q},z\in\mathcal{Z}_{n,k}}{\arg\min}\sum_{i\neq j}\mathcal{K}(\boldsymbol{A}_{ij},\boldsymbol{Q}_{z(i)z(j)}).\end{aligned}$$

Thus, it is enough to show that $\widehat{\boldsymbol{\Theta}}=\widehat{\boldsymbol{\Theta}}^r$ with large probability to prove the second part of Theorem 3. To do so, we show that

$$\mathbb{P}\left(Q(\widehat{z})\in[c_{min}/2,2c_{max}]^{k\times k}\right)\to 1. \tag{16}$$

Equation (16) implies that with probability going to 1, the maximum likelihood estimator of the probabilities of connections between nodes coincides $\widehat{\boldsymbol{\Theta}}$ with the restricted maximum likelihood estimator $\widehat{\boldsymbol{\Theta}}^r$. This concludes the proof of Theorem 3.

Proof of Equation (11)

For any $z \in \mathcal{Z}_{n,k}$ and $(\alpha, \boldsymbol{Q}) \in \mathcal{Q}$, let $l'_{\boldsymbol{X}}(\boldsymbol{A}, z; \alpha, \boldsymbol{Q}) = \left(\prod_{i \le n} \alpha_{z(i)}\right) \exp\left(\mathcal{L}_{\boldsymbol{X}}(\boldsymbol{A}; z, \boldsymbol{Q})\right)$ be the profile likelihood of the parameters $(z, \boldsymbol{Q})$. Then,

$$l'_{\boldsymbol{X}}(\boldsymbol{A}, z; \alpha, \boldsymbol{Q}) \le \sup_{\tau \in \mathcal{T}} \exp\left(\mathcal{J}_{\boldsymbol{X}}(\boldsymbol{A}; \tau, \alpha, \boldsymbol{Q})\right) \le l_{\boldsymbol{X}}(\boldsymbol{A}; \alpha, \boldsymbol{Q}). \tag{17}$$

Let $z' = \arg\max_{z: z \sim z^*} l'_{\boldsymbol{X}}\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)$. By definition of $l_{\boldsymbol{X}}$,

$$l_{\boldsymbol{X}}\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) = \sum_{z \sim z'} l'_{\boldsymbol{X}}\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) + \sum_{z \not\sim z'} l'_{\boldsymbol{X}}\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right). \tag{18}$$

On the one hand, we bound the sum $\sum_{z \sim z'} l'_{\boldsymbol{X}}\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)$ using the following result, proven in [37] :

Proposition 1 (Proposition 6.11 in [37])**.** *For any* $(\alpha, \boldsymbol{Q}) \in \mathcal{Q}$*,*

$$\frac{\sum_{z \sim z^*} l'_{\boldsymbol{X}}(\boldsymbol{A}, z; \alpha, \boldsymbol{Q})}{l'_{\boldsymbol{X}}(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*)} = \#Sym(\alpha, \boldsymbol{Q}) \max_{z' \sim z^*} \frac{l'_{\boldsymbol{X}}(\boldsymbol{A}, z'; \alpha, \boldsymbol{Q})}{l'_{\boldsymbol{X}}(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*)} (1 + o_p(1))$$

where the $o_p(1)$ *is uniform in* $(\alpha, \boldsymbol{Q})$ *and*

$$Sym(\alpha, \boldsymbol{Q}) = \left\{\sigma \in \mathcal{S}_k : \left(\alpha_{\sigma(a)}\right)_{a \le k} = (\alpha_a)_{a \le k} \text{ and } \left(\boldsymbol{Q}_{\sigma(a), \sigma(b)}\right)_{a,b \le k} = \left(\boldsymbol{Q}_{a,b}\right)_{a,b \le k}\right\}$$

for $\mathcal{S}_k$ *the set of permutations of* $[k]$.

Now, with probability going to one, $(\widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR})$ exhibits no symmetry, i.e. $\#Sym(\widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}) = 1$ (see Section B.11 in [37] for a proof of this result). Then, Proposition 1 implies that

$$\sum_{z \sim z'} l'_{\boldsymbol{X}}\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) = l'_{\boldsymbol{X}}\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)(1 + o_p(1))$$

which in turn implies

$$\sum_{z \sim z'} l'_{\boldsymbol{X}}\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) = l'_{\boldsymbol{X}}\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) + l_{\boldsymbol{X}}\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) o_p(1). \tag{19}$$

On the other hand, we bound the term $\sum_{z \not\sim z'} l'_{\boldsymbol{X}}\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)$ by combining the two following propositions from [37] :

Proposition 2 (Proposition 6.8 in [37])**.** *Let* $(t_n)_{n \in \mathbb{N}}$ *be a positive sequence such that* $t_n \to 0$ *and* $pnt_n / \sqrt{\log(n)} \to +\infty$*. Then, on an event of probability going to 1 and for* n *large enough,*

$$\sup_{(\alpha, \boldsymbol{Q}) \in \mathcal{Q}} \sum_{z \notin S(z^*, t_n)} l'_{\boldsymbol{X}}(\boldsymbol{A}, z; \alpha, \boldsymbol{Q}) = o_p\left(l'_{\boldsymbol{X}}(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*)\right)$$

where $S(z^*, t_n) = \{z \in \mathcal{Z}_{n,k} : \exists z' \sim z, \sum |z_i^* - z_i'| \le nt_n\}$.

Proposition 3 (Proposition 6.10 in [37]). *There exists a positive constant C such that*

$$\sup_{(\alpha,\boldsymbol{Q})\in\mathcal{Q}} \sum_{z\in S(z^*,C),z\nsim z^*} l'_{\boldsymbol{X}}\left(\boldsymbol{A},z;\alpha,\boldsymbol{Q}\right) = o_p\left(l'_{\boldsymbol{X}}\left(\boldsymbol{A},z^*;\alpha^*,\boldsymbol{Q}^*\right)\right).$$

Combining Propositions 2 and 3, we find that on a event of probability going to 1,

$$\sum_{z\nsim z^*} l'_{\boldsymbol{X}}\left(\boldsymbol{A},z;\widehat{\alpha}^{VAR},\widehat{\boldsymbol{Q}}^{VAR}\right) = l'_{\boldsymbol{X}}\left(\boldsymbol{A},z^*;\alpha^*,\boldsymbol{Q}^*\right)o_p(1).$$

Now, we use the definition of the variational estimator and Equation (17), and find that

$$l'_{\boldsymbol{X}}\left(\boldsymbol{A},z^*;\alpha^*,\boldsymbol{Q}^*\right) \le \sup_{\tau\in\mathcal{T}} \exp\left(\mathcal{J}_{\boldsymbol{X}}\left(\boldsymbol{A};\tau,\alpha^*,\boldsymbol{Q}^*\right)\right) \le \exp\left(\mathcal{J}_{\boldsymbol{X}}\left(\boldsymbol{A};\widehat{\tau}^{VAR},\widehat{\alpha}^{VAR},\widehat{\boldsymbol{Q}}^{VAR}\right)\right) \le l_{\boldsymbol{X}}\left(\boldsymbol{A};\widehat{\alpha}^{VAR},\widehat{\boldsymbol{Q}}^{VAR}\right).$$

Thus,

$$\sum_{z\nsim z^*} l'_{\boldsymbol{X}}\left(\boldsymbol{A},z;\widehat{\alpha}^{VAR},\widehat{\boldsymbol{Q}}^{VAR}\right) = l_{\boldsymbol{X}}\left(\boldsymbol{A};\widehat{\alpha}^{VAR},\widehat{\boldsymbol{Q}}^{VAR}\right)o_p(1). \tag{20}$$

Combining Equations (18), (19) and (20), we find that

$$l_{\boldsymbol{X}}\left(\boldsymbol{A};\widehat{\alpha}^{VAR},\widehat{\boldsymbol{Q}}^{VAR}\right) = l'_{\boldsymbol{X}}\left(\boldsymbol{A},z';\widehat{\alpha}^{VAR},\widehat{\boldsymbol{Q}}^{VAR}\right) + l_{\boldsymbol{X}}\left(\boldsymbol{A};\widehat{\alpha}^{VAR},\widehat{\boldsymbol{Q}}^{VAR}\right)o_p(1).$$

Dividing both sides by $l_{\boldsymbol{X}}\left(\boldsymbol{A};\widehat{\alpha}^{VAR},\widehat{\boldsymbol{Q}}^{VAR}\right)$, we find that

$$\mathbb{P}\left(z'|\boldsymbol{X}\odot\boldsymbol{A},\widehat{\alpha}^{Var},\widehat{\boldsymbol{Q}}^{Var}\right) = \frac{l'_{\boldsymbol{X}}\left(\boldsymbol{A},z';\widehat{\alpha}^{VAR},\widehat{\boldsymbol{Q}}^{VAR}\right)}{l_{\boldsymbol{X}}\left(\boldsymbol{A};\widehat{\alpha}^{VAR},\widehat{\boldsymbol{Q}}^{VAR}\right)} = 1 + o_p(1)$$

which proves Equation (11).

Proof of Equation (12)

By definition of $\mathcal{J}_{\boldsymbol{X}}$,

$$KL\left(\mathbb{P}_{\widehat{\tau}^{VAR}}(\cdot)||\mathbb{P}\left(\cdot|\boldsymbol{X}\odot\boldsymbol{A},\widehat{\alpha}^{VAR},\widehat{\boldsymbol{Q}}^{VAR}\right)\right) = \log\left(l_{\boldsymbol{X}}(\boldsymbol{A};\widehat{\alpha}^{VAR},\widehat{\boldsymbol{Q}}^{VAR})\right) - \mathcal{J}_{\boldsymbol{X}}(\boldsymbol{A};\widehat{\tau}^{VAR},\widehat{\alpha}^{VAR},\widehat{\boldsymbol{Q}}^{VAR}).$$

Equation (17) implies that

$$\mathcal{J}_{\boldsymbol{X}}(\boldsymbol{A};\widehat{\tau}^{VAR},\widehat{\alpha}^{VAR},\widehat{\boldsymbol{Q}}^{VAR}) \ge \log\left(l'_{\boldsymbol{X}}\left(\boldsymbol{A},z';\widehat{\alpha}^{VAR},\widehat{\boldsymbol{Q}}^{VAR}\right)\right)$$

so

$$KL\left(\mathbb{P}_{\widehat{\tau}^{VAR}}(\cdot)||\mathbb{P}\left(\cdot|\boldsymbol{X}\odot\boldsymbol{A},\widehat{\alpha}^{VAR},\widehat{\boldsymbol{Q}}^{VAR}\right)\right) \le \log\left(l_{\boldsymbol{X}}(\boldsymbol{A};\widehat{\alpha}^{VAR},\widehat{\boldsymbol{Q}}^{VAR})\right) - \log\left(l'_{\boldsymbol{X}}\left(\boldsymbol{A},z';\widehat{\alpha}^{VAR},\widehat{\boldsymbol{Q}}^{VAR}\right)\right).$$

Note that Equation (11) implies

$$\log\left(l_{\boldsymbol{X}}(\boldsymbol{A};\widehat{\alpha}^{VAR},\widehat{\boldsymbol{Q}}^{VAR})\right) - \log\left(l'_{\boldsymbol{X}}\left(\boldsymbol{A},z';\widehat{\alpha}^{VAR},\widehat{\boldsymbol{Q}}^{VAR}\right)\right) = o_p(1).$$

Now, using Pinsker's inequality, we see that

$$\left|\mathbb{P}_{\widehat{\tau}^{VAR}}(z') - \mathbb{P}\left(z'|\boldsymbol{X}\odot\boldsymbol{A},\widehat{\alpha}^{VAR},\widehat{\boldsymbol{Q}}^{VAR}\right)\right| = o_p(1).$$

We use Equation (11) and the definition of $\widehat{z}^{(VAR)}$ to conclude the proof of Equation (12).

Proof of Equation (14)

For $z\in\mathcal{Z}_{n,k}$, define

$$\begin{aligned}\Lambda(z) &= max_{\boldsymbol{Q}\in\mathcal{Q}}\mathcal{L}_{\boldsymbol{X}}(\boldsymbol{A};z,\boldsymbol{Q}) - \mathcal{L}_{\boldsymbol{X}}(\boldsymbol{A};z^*,\boldsymbol{Q}^*) \quad \text{and}\\ \widetilde{\Lambda}(z) &= max_{\boldsymbol{Q}\in\mathcal{Q}}\mathbb{E}\left[\mathcal{L}_{\boldsymbol{X}}(\boldsymbol{A};z,\boldsymbol{Q}) - \mathcal{L}_{\boldsymbol{X}}(\boldsymbol{A};z^*,\boldsymbol{Q}^*)\middle|z^*\right].\end{aligned}$$

Moreover, for $z\in\mathcal{Z}_{n,k}$ and $(\alpha,\boldsymbol{Q})$, define

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

where $\|z'-z^*\|_0$ is the Hamming distance between the label functions z' and z^*.

To prove Equation (14), we will use the following results.

Proposition 4 (Equation (B.1) in [37]). *There exists a constant $c > 0$ such that on an event of probability going to one, for all positive sequence $(t_n)_{n\in\mathbb{N}}$ such that $t_n \to 0$ and $pnt_n/\sqrt{\log(n)} \to +\infty$, $\forall z \notin S(z^*, t_n)$,*

$$\tilde{\Lambda}(z) \leq -\frac{3cpn^2 t_n \delta(\boldsymbol{Q}^*)}{4}$$

where and $\delta(\boldsymbol{Q}) = \min_{a,a'} \max_c KL\left(\boldsymbol{Q}_{ac}, \boldsymbol{Q}_{a'c}\right)$ *and* $S(z^*, t_n) = \left\{z \in \mathcal{Z}_{n,k} : \|z - z^*\|_{\sim,0} \leq nt_n\right\}$.

Proposition 5 (Proposition 6.7 in [37]). *There exists a constant $C_{\mathcal{Q}} > 0$ depending on $\mathcal{Q}$ such that for any sequence $(\epsilon_n)_{n\in\mathbb{N}}$ with $\epsilon_n < C_{\mathcal{Q}}$ and $\epsilon_n \geq k^2/(\sqrt{8}n)$,*

$$\sup_{z\in\mathcal{Z}_{n,k}} \left(\Lambda(z) - \tilde{\Lambda}(z)\right) = O_p(\epsilon_n n^2).$$

We choose $\epsilon_n = 3\delta(\boldsymbol{Q}^*)\log(n)/(8n)$. Then, Proposition 5 implies that there exists a constant $C > 0$ such that with probability going to 1, $\sup_{z\in\mathcal{Z}_{n,k}} \left(\Lambda(z) - \tilde{\Lambda}(z)\right) \leq C\epsilon_n n^2$. Moreover, we choose $t_n = 2C\log(n)/(cnp)$ and note that under the assumption $p \gg \log(n)/n$, $t_n \to 0$. Then, Propositions 4 and 5 imply that with probability going to one

$$\begin{aligned}
\sup_{z\notin S(z^*,t_n)} \Lambda(z) &\leq \sup_{z\notin S(z^*,t_n)} \widetilde{\Lambda}(z) + \sup_{z\notin S(z^*,t_n)} \left(\Lambda(z) - \widetilde{\Lambda}(z)\right) \\
&\leq -\frac{3Cpn^2 t_n \delta(\boldsymbol{Q}^*)}{4} + \frac{3Cpn^2 t_n \delta(\boldsymbol{Q}^*)}{8} \\
&\leq -\frac{3Cn\log(n)\delta(\boldsymbol{Q}^*)}{8}.
\end{aligned}$$

This implies in particular that

$$\mathbb{P}\left(\sup_{z\notin S(z^*,t_n)} \Lambda(z) < 0\right) \to 1. \tag{21}$$

We show a similar result for label functions z that are close to z^*. To do so, we use the following result.

Proposition 6 (Proposition 6.5 in [37]). *There exists a positive constant C such that on an event of probability going to 1, for all $z \in S(z^*, C)$,*

$$\tilde{\Lambda}(z) \leq -\frac{3cpn^2\delta(\boldsymbol{Q}^*)\, \|z - z^*\|_{\sim,0}}{4}.$$

We use Proposition 4, where we choose $\epsilon_n = k^2/n$. Then, there exists a constant $C' > 0$ such that with probability going to 1, $\sup_{z\in\mathcal{Z}_{n,k}} \left(\Lambda(z) - \widetilde{\Lambda}(z)\right) \leq C'nk^2$. Now, Proposition 6 implies that with probability going to 1,

$$\begin{aligned}
\sup_{z\in S(z^*,C), z\not\sim z^*} \Lambda(z) &\leq \sup_{z\in S(z^*,C), z\not\sim z^*} \widetilde{\Lambda}(z) + \sup_{z\in S(z^*,C), z\not\sim z^*} \left(\Lambda(z) - \widetilde{\Lambda}(z)\right) \\
&\leq -\frac{3cpn^2\delta(\boldsymbol{Q}^*)}{4} + C'nk^2 \\
&\leq nk^2\left(C' - \frac{3cpn\delta(\boldsymbol{Q}^*)}{8k^2}\right).
\end{aligned}$$

Since $pn \to +\infty$, this implies that

$$\mathbb{P}\left(\sup_{z\in S(z^*,C), z\not\sim z^*} \Lambda(z) < 0\right) \to 1. \tag{22}$$

Finally, since $t_n \to 0$, for n large enough $\mathcal{Z}_{n,k} = S(z^*, C) \cup \overline{S(z^*, t_n)}$. Thus, Equations (21) and (22) imply that

$$\mathbb{P}\left(\sup_{z \not\sim z^*} \Lambda(z) < 0\right) \to 1. \tag{23}$$

Now, $\Lambda(z^*) = 0$. Thus, with probability going to 1, $\arg\max \Lambda(z) \sim z^*$, so $\widehat{z} \sim z^*$.

Proof of Equation (16)

To prove Equation (16), we use Bernstein's inequality, which we recall here for sake of completeness :

Theorem 4 (Bernstein's inequality)**.** *Let $X_1, ..., X_n$ be independent centered random variables. Assume that for any $i \in [n]$, $|X_i| \le M$ almost surely, then*

$$\mathbb{P}\left(\left|\sum_{1\le i\le n} X_i\right| \ge \sqrt{2t \sum_{1\le i\le n} \mathbb{E}[X_i^2]} + \frac{2M}{3}t\right) \le 2e^{-t}.$$

For $z \in \mathcal{Z}_{n,k}$ and $(a,b) \in [k]^2$, define

$$n_{ab}(z) = \begin{cases} |(z)^{-1}(a)| \times |(z)^{-1}(b)| & \text{if } a \neq b \\ |(z)^{-1}(a)| \times \big(|(z)^{-1}(a)| - 1\big) & \text{otherwise} \end{cases}$$

and

$$n^{\boldsymbol{X}}_{ab}(z) = \sum_{\substack{i \in z^{-1}(a), j \in z^{-1}(b) \\ i \neq j}} \boldsymbol{X}_{ij}$$

the number of entries and of observed entries of the adjacency matrix between nodes of the communities a and b, and $\boldsymbol{Q}(z) = (\boldsymbol{Q}(z)_{ab})$ such that $\boldsymbol{Q}(z)_{ab} = \left(\sum_{i \in z^{-1}(a), j \in z^{-1}(b)} \boldsymbol{X}_{ij}\boldsymbol{A}_{ij}\right)/n^{\boldsymbol{X}}_{ab}(z)$. With these notations, we note that $\widehat{\boldsymbol{Q}} = \boldsymbol{Q}(\widehat{z})$.

Note that $|(z^*)^{-1}(a)|$ is a sum of n independent Bernoulli random variables with mean α_a. Using Bernstein's inequality 4, we find that for any a,

$$\mathbb{P}\left(n\alpha_a - |(z^*)^{-1}(a)| \ge 0.5n\alpha_a\right) \le 2e^{-n\alpha_a/16}.$$

Thus,

$$\mathbb{P}\left(\min_a |(z^*)^{-1}(a)| \le 0.5n \min_a \alpha_a\right) \le 2ke^{-n \min_a \alpha_a/16}.$$

Therefore, the event $\Omega = \left\{\min_{a,b} n_{a,b}(z^*) \ge n^2 \min_a (\alpha_a)^2/5\right\}$ holds with probability going to 1.

Similarly, note that conditionally on z^*, $n^{\boldsymbol{X}}_{ab}(z^*)$ is a sum of $n_{ab}(z^*)$ independent Bernoulli variables with parameter p. Then, for any two $(a,b) \in [k]^2$, Bernstein's inequality 4 implies that

$$\mathbb{P}\left(|pn_{ab}(z^*) - n^{\boldsymbol{X}}_{ab}(z^*)| \ge 0.5pn_{ab}(z^*)\big|z^*\right) \le 2e^{-pn_{ab}(z^*)/16}.$$

Thus,

$$\mathbb{P}\left(\min_{a,b} n^{\boldsymbol{X}}_{ab}(z^*) \le 0.5p \min_{a,b} n_{ab}(z^*)\big|z^*\right) \le 2ke^{-p \min_{a,b} n_{ab}(z^*)/16}.$$

This implies that

$$\mathbb{P}\left(\min_{a,b} n^{\boldsymbol{X}}_{ab}(z^*) \le 0.1n^2 p \min_a \alpha_a^2 \big|\Omega\right) \le 2ke^{-pn^2 \min_a \alpha_a/80}.$$

Since $p \gg \log(n)/n$, the event $\Omega' = \{\forall (a,b) \in [k]^2, n^{\boldsymbol{X}}_{ab}(z^*) \ge 0.1n^2 p \min_a \alpha_a^2\}$ holds with probability going to 1.

Now, we show that on the event Ω', with large probability, $\boldsymbol{Q}(z^*) \in [c_{min}/2, 2c_{max}]^{k\times k}$. Recall that for any a, b, conditionally on z^* and $\boldsymbol{X}$, $n_{ab}^{\boldsymbol{X}}(z^*)\boldsymbol{Q}(z^*)_{ab}$ is a sum of $n_{ab}^{\boldsymbol{X}}(z^*)$ independent Bernoulli random variables with mean $\boldsymbol{Q}_{ab}^*$. Then, Bernstein's inequality implies that for any $t > 0$

$$\mathbb{P}\left(\left|n_{ab}^{\boldsymbol{X}}(z^*)\boldsymbol{Q}(z^*)_{ab} - n_{ab}^{\boldsymbol{X}}(z^*)\boldsymbol{Q}_{ab}^*\right| \geq \sqrt{2tn_{ab}^{\boldsymbol{X}}(z^*)\boldsymbol{Q}_{ab}^*} + \frac{2t}{3}\middle| z^*, \boldsymbol{X}\right) \leq 2e^{-t}.$$

Choosing $t = n_{ab}^{\boldsymbol{X}}(z^*)\boldsymbol{Q}_{ab}^*/16$ yields

$$\mathbb{P}\left(\left|n_{ab}^{\boldsymbol{X}}(z^*)\boldsymbol{Q}(z^*)_{ab} - n_{ab}^{\boldsymbol{X}}(z^*)\boldsymbol{Q}_{ab}^*\right| \geq 0.5n_{ab}^{\boldsymbol{X}}(z^*)\boldsymbol{Q}_{ab}^*\middle| z^*, \boldsymbol{X}\right) \leq 2e^{-n_{ab}^{\boldsymbol{X}}(z^*)\boldsymbol{Q}_{ab}^*/16}.$$

On the event Ω', this implies that

$$\mathbb{P}\left(|\boldsymbol{Q}(z^*)_{ab} - \boldsymbol{Q}_{ab}^*| \geq 0.5\boldsymbol{Q}_{ab}^*\middle|\Omega'\right) \leq 2e^{-n^2\boldsymbol{Q}_{ab}^*(\min_a \alpha_a)^2/160}.$$

A union bound yields

$$\mathbb{P}\left(\boldsymbol{Q}(z^*) \notin [c_{min}/2, 2c_{max}]^{k\times k}\middle|\Omega'\right) \leq 2k^2 e^{-n^2\min_{a,b}\boldsymbol{Q}_{ab}^*(\min_a \alpha_a)^2/160}.$$

Since $\mathbb{P}(\Omega') \to 1$, this shows that

$$\mathbb{P}\left(\boldsymbol{Q}(z^*) \in [c_{min}/2, 2c_{max}]^{k\times k}\right) \to 1.$$

Now, Equation (14) shows that with probability going to 1, $\widehat{z} \sim z^*$. Thus,

$$\mathbb{P}\left(\boldsymbol{Q}(\widehat{z}) \in [c_{min}/2, 2c_{max}]^{k\times k}\right) \to 1.$$

A.2 Proof of Theorem 3

In the case of fully observed network, we alleviate notations and write

$$\begin{aligned}
\mathcal{L}(\boldsymbol{A}; z, \boldsymbol{Q}) &= \sum_{i\neq j}\boldsymbol{A}_{ij}\log\left(\boldsymbol{Q}_{z(i),z(j)}\right) + (1-\boldsymbol{A}_{ij})\log\left(1-\boldsymbol{Q}_{z(i),z(j)}\right),\\
l(\boldsymbol{A}; \alpha, \boldsymbol{Q}) &= \sum_{z\in\mathcal{Z}_{n,k}}\left(\prod_i \alpha_{z(i)}\right)\exp\left(\mathcal{L}(\boldsymbol{A}; z, \boldsymbol{Q})\right),\\
\text{and } \mathcal{J}(\boldsymbol{A}; \tau, \alpha, \boldsymbol{Q}) &= \log\left(l(\boldsymbol{A}; \alpha, \boldsymbol{Q})\right) - KL\left(\mathbb{P}_\tau(\cdot)\,||\mathbb{P}(\cdot|\boldsymbol{A}, \alpha, \boldsymbol{Q})\right).
\end{aligned}$$

For any $z \in \mathcal{Z}_{n,k}$ and $(\alpha, \boldsymbol{Q}) \in \mathcal{Q}$, we denote

$$l'(\boldsymbol{A}, z; \alpha, \boldsymbol{Q}) = \left(\prod_{i\leq n}\alpha_{z(i)}\right)\exp\left(\mathcal{L}(\boldsymbol{A}; z, \boldsymbol{Q})\right)$$

the likelihood of the parameters $(\alpha, \boldsymbol{Q})$ and the label function z. Then, the likelihood of the stochastic block model with parameters $(\alpha, \boldsymbol{Q})$ is given by $l(\boldsymbol{A}; \alpha, \boldsymbol{Q}) = \sum_{z\in\mathcal{Z}_{n,k}} l'(\boldsymbol{A}, z; \alpha, \boldsymbol{Q})$. Note that the likelihood functions $l(\boldsymbol{A}; \alpha, \boldsymbol{Q})$ and $l'(\boldsymbol{A}, z; \alpha, \boldsymbol{Q})$ provide lower and upper bounds on the variational objective function $\mathcal{J}(\boldsymbol{A}; \tau, \alpha, \boldsymbol{Q})$: for any parameter $(\alpha, \boldsymbol{Q})$ and any label function $z \in \mathcal{Z}_{n,k}$,

$$l'(\boldsymbol{A}, z; \alpha, \boldsymbol{Q}) \leq \sup_{\tau\in\mathcal{T}}\exp\left(\mathcal{J}(\boldsymbol{A}; \tau, \alpha, \boldsymbol{Q})\right) \leq l(\boldsymbol{A}; \alpha, \boldsymbol{Q}). \tag{24}$$

To prove Proposition 3, we first show that $\mathbb{P}\left(\cdot|\boldsymbol{A}, \widehat{\alpha}^{Var}, \widehat{\boldsymbol{Q}}^{Var}\right)$, i.e. the posterior distribution of z at the variational estimator $(\widehat{\alpha}^{Var}, \widehat{\boldsymbol{Q}}^{Var})$, concentrates around $\delta_{z'}$, the dirac distribution at the label function $z' = \arg\max_{z:z\sim z^*} l'\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)$:

$$\mathbb{P}\left(z'|\boldsymbol{A}, \widehat{\alpha}^{Var}, \widehat{\boldsymbol{Q}}^{Var}\right) = 1 - o_p(1). \tag{25}$$

Then, we show that it implies the concentration of the estimator $\widehat{z}^{Var}$:

$$\mathbb{P}\left(\widehat{z}^{Var} = z'|\boldsymbol{A}\right) = 1 - o_p(1). \tag{26}$$

Together (25) and (26) imply $\mathbb{P}\left(\widehat{z}^{Var} \sim z^*|\boldsymbol{A}\right) = 1 - o_p(1)$. Since the random variable $\mathbb{P}\left(\widehat{z}^{Var} \sim z^*|\boldsymbol{A}\right)$ is bounded, Equation (26) also implies that it converges to 1 in expectation. Finally, we show that with probability going to one, the maximum likelihood estimator of the label function is equal to the true label function (up to permutation):

$$\mathbb{P}\left(\widehat{z} \sim z^*\right) = 1 - o_p(1) \tag{27}$$

which concludes the proof of the first part of Theorem 3.

To prove the second part of Theorem 3, we show that the maximum likelihood estimator studied in Proposition 3 is equal to the restricted maximum estimator studied in Theorem 1. More precisely, define $c_{min} = \min_{a,b} \boldsymbol{Q}^0_{a,b}$ and $c_{max} = \max_{a,b} \boldsymbol{Q}^0_{a,b}$. Theorem 1 implies that for some absolute constant $C > 0$,

$$\mathbb{P}\left(\left\|\boldsymbol{\Theta}^* - \widehat{\boldsymbol{\Theta}}^r\right\|_2^2 \le C(c_{max}/c_{min})^2 \rho_n \left(k^2 + n\log(k)\right)\right) \to 1,$$

where the restricted maximum likelihood estimator $\widehat{\boldsymbol{\Theta}}^r$ is defined as

$$\begin{aligned}
&\widehat{\boldsymbol{\Theta}}^r_{i<j} = \widehat{\boldsymbol{Q}}^r_{\widehat{z}^r(i)\widehat{z}^r(j)},\ \widehat{\boldsymbol{\Theta}}^r_{ii} = 0\\
&(\widehat{\boldsymbol{Q}}^r, \widehat{z}^r) \in \underset{\boldsymbol{Q} \in [c_{min}\rho_n/2, 2c_{max}\rho_n]^{k\times k}_{\mathrm{sym}}, z \in \mathcal{Z}_{n,k}}{\arg\min} \sum_{i\neq j} \mathcal{K}(\boldsymbol{A}_{ij}, \boldsymbol{Q}_{z(i)z(j)}).
\end{aligned}$$

One the other hand, Proposition 3 implies that with probability going to one, the variational estimator of the probabilities of connections $\widehat{\boldsymbol{\Theta}}^{VAR}$ is equal to the maximum likelihood estimator $\widehat{\boldsymbol{\Theta}}$ given by

$$\begin{aligned}
&\widehat{\boldsymbol{\Theta}}_{i<j} = \widehat{\boldsymbol{Q}}_{\widehat{z}(i)\widehat{z}(j)},\ \widehat{\boldsymbol{\Theta}}_{ii} = 0\\
&\text{for } (\widehat{\boldsymbol{Q}}, \widehat{z}) \in \underset{\boldsymbol{Q} \in \mathcal{Q}, z \in \mathcal{Z}_{n,k}}{\arg\min} \sum_{i\neq j} \mathcal{K}(\boldsymbol{A}_{ij}, \boldsymbol{Q}_{z(i)z(j)}).
\end{aligned}$$

We show that

$$\mathbb{P}\left(\widehat{\boldsymbol{\Theta}} = \widehat{\boldsymbol{\Theta}}^r\right) \to 1, \tag{28}$$

which concludes the proof of Theorem 3.

Proof of Equation (25)

The proof of Equation (25) relies on results proven in [7], which we recall for the sake of completeness. For any two parameters $(\alpha, \boldsymbol{Q})$ and $(\alpha', \boldsymbol{Q}')$ in $\mathcal{Q}$, we say that $(\alpha', \boldsymbol{Q}') \in \mathcal{S}_{\alpha,\boldsymbol{Q}}$ if there exists a permutation σ of $\{1, ..., k\}$ such that for any $(a, b) \in \{1, ..., k\}^2$, $\boldsymbol{Q}'_{\sigma(a),\sigma(b)} = \boldsymbol{Q}_{a,b}$ and $\alpha'_{\sigma(a)} = \alpha_a$.

Theorem 5 (Theorem 1 in [7]). *Let (z^*, A) be generated from a stochastic block model with parameters $(\alpha^*, \boldsymbol{Q}^*) \in \mathcal{Q}$ such that $\boldsymbol{Q}^0$ has no identical columns and $\rho_n \gg \log(n)/n$. Then, for any $(\alpha, \boldsymbol{Q}) \in \mathcal{Q}$,*

$$\frac{l\left(\boldsymbol{A}; \alpha, \boldsymbol{Q}\right)}{l\left(\boldsymbol{A}; \alpha^*, \boldsymbol{Q}^*\right)} = \max_{(\alpha', \boldsymbol{Q}') \in \mathcal{S}_{\alpha,\boldsymbol{Q}}} \frac{l'\left(\boldsymbol{A}, z^*; \alpha', \boldsymbol{Q}'\right)}{l'\left(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*\right)} \left(1 + \epsilon_n\left(\left(\alpha', \boldsymbol{Q}'\right), k\right)\right) + \epsilon_n\left(\left(\alpha', \boldsymbol{Q}'\right), k\right)$$

where $\sup_{(\alpha,\boldsymbol{Q})\in\mathcal{Q}} \epsilon_n\left(\left(\alpha, \boldsymbol{Q}\right), k\right) = o_p(1)$.

Proposition 7 (Lemma 3 in [7]). *Let (z^*, A) be generated from a stochastic block model with parameters $(\alpha^*, \boldsymbol{Q}^*) \in \mathcal{Q}$ such that $\boldsymbol{Q}^0$ has no identical columns and $\rho_n \gg \log(n)/n$. Then,*

$$\frac{l'\left(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*\right)}{l\left(\boldsymbol{A}; \alpha^*, \boldsymbol{Q}^*\right)} = 1 + o_p(1).$$

Recall that $z' = \arg\max_{z: z\sim z^*} l'\left(\boldsymbol{A}, z^*; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)$. By definition of l and l',

$$\sum_{z\neq z'} l'\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) = l\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) - l'\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right).$$

Thus

$$\frac{\sum_{z\neq z'} l'\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)}{l'\left(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*\right)} = \frac{l\left(\boldsymbol{A}; \alpha^*, \boldsymbol{Q}^*\right)}{l'\left(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*\right)} \times \frac{l\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)}{l\left(\boldsymbol{A}; \alpha^*, \boldsymbol{Q}^*\right)} - \frac{l'\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)}{l'\left(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*\right)} \tag{29}$$

Using Proposition 7, we have that

$$\frac{l\left(\boldsymbol{A}; \alpha^*, \boldsymbol{Q}^*\right)}{l'\left(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*\right)} = 1 + o_p(1). \tag{30}$$

Moreover, we note that

$$\begin{aligned} \max_{(\alpha', \boldsymbol{Q}')\in \mathcal{S}_{\widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}}} l'\left(\boldsymbol{A}, z^*; \alpha', \boldsymbol{Q}'\right) &= \max_{z\sim z^*} l'\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) \\ &= l'\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) \end{aligned}$$

by the definition of z'. Then, applying Theorem 5, we get that

$$\frac{l\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)}{l\left(\boldsymbol{A}; \alpha^*, \boldsymbol{Q}^*\right)} = \frac{l'\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)}{l'\left(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*\right)} (1 + o_p(1)) + o_p(1). \tag{31}$$

Combining Equations (29), (30) and (31), we obtain that

$$\frac{\sum_{z\neq z'} l'\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)}{l'\left(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*\right)} = \frac{l'\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)}{l'\left(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*\right)} o_p(1) + o_p(1).$$

Thus,

$$\sum_{z\neq z'} l'\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) = \max\left\{ l'\left(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*\right), l'\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right\} o_p(1). \tag{32}$$

On the one hand, using Equation (24) and the definition of $(\widehat{\tau}^{VAR}, \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR})$, we find that

$$\begin{aligned} l'\left(\boldsymbol{A}, z^*; \alpha^*, \boldsymbol{Q}^*\right) &\leq \sup_{\tau\in\mathcal{T}} \exp\left(\mathcal{J}\left(\boldsymbol{A}; \tau, \alpha^*, \boldsymbol{Q}^*\right)\right) \\ &\leq \exp\left(\mathcal{J}\left(\boldsymbol{A}; \widehat{\tau}^{VAR}, \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right) \\ &\leq l\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right). \end{aligned}$$

Also, by the definition of l and l', we have that $l'\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) \leq \left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)$. Thus, Equation (32) implies

$$\sum_{z\neq z'} l'\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) = l\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) o_p(1). \tag{33}$$

Now, we can conclude the proof of Equation (25) by noticing that

$$\begin{aligned} \mathbb{P}\left(z' | \boldsymbol{A}, \widehat{\alpha}^{Var}, \widehat{\boldsymbol{Q}}^{Var}\right) &= \frac{l'\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)}{l\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)} \\ &= 1 - \frac{\sum_{z\neq z'} l'\left(\boldsymbol{A}, z; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)}{l\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)} \end{aligned}$$

and using Equation (33).

Proof of Equation (26) By the definition of $\mathcal{J}(\boldsymbol{A}; \tau, \alpha, \boldsymbol{Q})$, we have that

$$KL\left(\mathbb{P}_{\widehat{\tau}^{VAR}}(\cdot)||\mathbb{P}\left(\cdot|\boldsymbol{A}, \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right) = \log\left(l\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right) - \mathcal{J}\left(\boldsymbol{A}; \widehat{\tau}^{VAR}, \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right).$$

Equation (24) implies that $\mathcal{J}\left(\boldsymbol{A}; \widehat{\tau}^{VAR}, \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right) \geq \log\left(l\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right)$, so

$$KL\left(\mathbb{P}_{\widehat{\tau}^{VAR}}(\cdot)||\mathbb{P}\left(\cdot|\boldsymbol{A}, \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right) \leq \log\left(l\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right) - \log\left(l\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right).$$

Note that Equation (25) implies

$$\log\left(l\left(\boldsymbol{A}; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right) - \log\left(l\left(\boldsymbol{A}, z'; \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right) = o_p(1).$$

Now, using Pinsker's inequality, we see that

$$\left|\mathbb{P}_{\widehat{\tau}^{VAR}}(z') - \mathbb{P}\left(z'|\boldsymbol{A}, \widehat{\alpha}^{VAR}, \widehat{\boldsymbol{Q}}^{VAR}\right)\right| = o_p(1).$$

We use Equation (25) and the definition of $\widehat{z}^{(VAR)}$ to conclude the proof of Equation (26).

Proof of Equation (27)

Equation (27) is proven in [8]. In this work, the authors define the profile likelihood modularity $\mathcal{Q}_{LM}(A, z)$ of a label function $z \in \mathcal{Z}_{n,k}$ as

$$\mathcal{Q}_{LM}(A, z) = \frac{1}{2}\sum_{a,b} n_{ab}\left(\frac{\boldsymbol{O}_{ab}}{n_{ab}}\log\left(\frac{\boldsymbol{O}_{ab}}{n_{ab}}\right) + \left(1 - \frac{\boldsymbol{O}_{ab}}{n_{ab}}\right)\log\left(1 - \frac{\boldsymbol{O}_{ab}}{n_{ab}}\right)\right).$$

for $\boldsymbol{O}_{ab} = \sum\limits_{i \in z^{-1}(a), j \in z^{-1}(b)} \boldsymbol{A}_{ij}$ and

$$n_{ab} = \begin{cases} |z^{-1}(a)| \times |z^{-1}(b)| & \text{if } a \neq b \\ |z^{-1}(a)| \times \left(|z^{-1}(a)| - 1\right) & \text{otherwise} \end{cases}$$

For $\hat{z}^{LM} = \arg\max_{z \in \mathcal{Z}_{n,k}} \mathcal{Q}_{LM}(A, z)$, the authors of [8] prove that under the assumptions of Proposition 3, with probability going to 1, $\hat{z}^{LM} \sim z^*$. Since maximizing $\mathcal{Q}_{LM}(A, z)$ is equivalent to maximizing $\max_{\boldsymbol{Q}} \mathcal{L}(\boldsymbol{A}; \boldsymbol{Q}, z)$, this implies that $\widehat{z} \sim z^*$ with probability going to 1.

Proof of Equation (28) To do so, we show that with large probability, $\boldsymbol{Q}(\widehat{z}) \in [c_{min}\rho_n/2, 2c_{max}\rho_n]^{k\times k}$. We define

$$n_{ab}(z) = \begin{cases} |z^{-1}(a)| \times |z^{-1}(b)| & \text{if } a \neq b \\ |z^{-1}(a)| \times \left(|z^{-1}(a)| - 1\right) & \text{otherwise} \end{cases}$$

for $z \in \mathcal{Z}_{n,k}$, and $\boldsymbol{Q}(z) = (\boldsymbol{Q}(z)_{ab})$ such that $\boldsymbol{Q}(z)_{ab} = \left(\sum\limits_{i \in z^{-1}(a), j \in z^{-1}(b)} \boldsymbol{A}_{ij}\right)/n_{ab}(z)$. With these notations, we note that $\widehat{\boldsymbol{Q}} = \boldsymbol{Q}(\widehat{z})$.

Recall that $|(z^*)^{-1}(a)|$ is a sum of n independent Bernoulli random variables with mean α_a^0. Using Bernstein's inequality 4, we find that for any a,

$$\mathbb{P}\left(n\alpha_a^0 - |(z^*)^{-1}(a)| \geq 0.5n\alpha_a^0\right) \leq 2e^{-n\alpha_a^0/16}.$$

Thus,

$$\mathbb{P}\left(\min_a |(z^*)^{-1}(a)| \leq 0.5n \min_a \alpha_a^0\right) \leq 2ke^{-n\min_a \alpha_a^0/16}.$$

Therefore, the event $\Omega = \left\{\min_{a,b} n_{a,b}(z^*) \geq n^2 \min_a (\alpha_a^0)^2/5\right\}$ holds with probability going to 1.

Now, we show that on the event Ω, with large probability, $\boldsymbol{Q}(z^*) \in [c_{min}\rho_n/2, 2c_{max}\rho_n]^{k\times k}$. Recall that for any a, b, conditionally on z^*, $n_{ab}(z^*)\boldsymbol{Q}(z^*)_{ab}$ is a sum of $n_{ab}(z^*)$ independent Bernoulli random variables with mean $\rho_n \boldsymbol{Q}^0_{ab}$. Then, Bernstein's inequality 4 implies that for any $t > 0$

$$\mathbb{P}\left(\left|n_{ab}(z^*)\boldsymbol{Q}(z^*)_{ab} - n_{ab}(z^*)\rho_n\boldsymbol{Q}^0_{ab}\right| \geq \sqrt{2tn_{ab}(z^*)\rho_n\boldsymbol{Q}^0_{ab}} + \frac{2t}{3}\right) \leq 2e^{-t}.$$

Choosing $t = n_{ab}(z^*)\rho_n\boldsymbol{Q}^0_{ab}/16$ yields

$$\mathbb{P}\left(\left|n_{ab}(z^*)\boldsymbol{Q}(z^*)_{ab} - n_{ab}(z^*)\rho_n\boldsymbol{Q}^0_{ab}\right| \geq 0.5n_{ab}(z^*)\rho_n\boldsymbol{Q}^0_{ab}\right) \leq 2e^{-n_{ab}(z^*)\rho_n\boldsymbol{Q}^0_{ab}/16}.$$

On the event Ω, this implies that

$$\mathbb{P}\left(\left|n_{ab}(z^*)\boldsymbol{Q}(z^*)_{ab} - n_{ab}(z^*)\rho_n\boldsymbol{Q}^0_{ab}\right| \geq 0.5n_{ab}(z^*)\rho_n\boldsymbol{Q}^0_{ab}\right) \leq 2e^{-n^2\rho_n\boldsymbol{Q}^0_{ab}(\min_a \alpha^0_a)^2/80}.$$

A union bound yields

$$\mathbb{P}\left(\boldsymbol{Q}(z^*) \notin [c_{min}\rho_n/2, 2c_{max}\rho_n]^{k\times k}\right) \leq 2k^2 e^{-n^2\rho_n \min_{a,b}\boldsymbol{Q}^0_{ab}(\min_a \alpha^0_a)^2/80}$$

on the event Ω. Since $\mathbb{P}(\Omega) \to 1$ and $n^2\rho_n \to +\infty$, this shows that

$$\mathbb{P}\left(\boldsymbol{Q}(z^*) \in [c_{min}\rho_n/2, 2c_{max}\rho_n]^{k\times k}\right) \to 1.$$

Now, Equation (27) shows that with probability going to 1, $\widehat{z} \sim z^*$. Thus, $Q(\widehat{z}) \in [c_{min}\rho_n/2, 2c_{max}\rho_n]^{k\times k}$ with probability going to one, and the maximum likelihood estimator of the probabilities of connections between nodes coincides with the restricted maximum likelihood estimator. This concludes the proof of Equation (28).

B Further informations on the numerical experiments

B.1 Simulation protocol

In this section, we provide details on the simulation protocol for Section 4.1. The numerical experiments where conducted using R version 4.0.3, the package softImpute version 1.4.1, and the package missSBM version 0.3.0.

Dense stochastic block model The parameters used for the simulations are the following : $\alpha^{assort.} = \alpha^{disassort.} = (1/3, 1/3, 1/3)$, $\alpha^{mix.} = (0.1, 0.3, 0.6)$, and

$$\boldsymbol{Q}^{assort.} = \begin{pmatrix} 0.5 & 0.2 & 0.2 \\ 0.2 & 0.5 & 0.2 \\ 0.2 & 0.2 & 0.5 \end{pmatrix}, \ \boldsymbol{Q}^{disassort.} = \begin{pmatrix} 0.2 & 0.5 & 0.5 \\ 0.5 & 0.2 & 0.5 \\ 0.5 & 0.5 & 0.2 \end{pmatrix}, \ \boldsymbol{Q}^{mix.} = \begin{pmatrix} 0.1 & 0.5 & 0.3 \\ 0.5 & 0.2 & 0.4 \\ 0.3 & 0.4 & 0.6 \end{pmatrix}.$$

For each model and each number of nodes, we simulate 100 networks. For each networks, entries of the adjacency matrix are observed independently from one another with probability 1/2. Then, the matrix of connection probabilities $\boldsymbol{\Theta}^*$ is estimated using each method (variational approximation to the maximum likelihood estimator, missSBM, and softImpute). The oracle estimator is obtained as

$$\forall a < k \text{ and } b < k, \widehat{\boldsymbol{Q}}^*_{ab} \triangleq \frac{\sum\limits_{i\in(z^*)^{-1}(a), j\in(z^*)^{-1}(b), i\neq j} \boldsymbol{X}_{ij}\boldsymbol{A}_{ij}}{\sum\limits_{i\in(z^*)^{-1}(a), j\in(z^*)^{-1}(b), i\neq j} \boldsymbol{X}_{ij}}$$

Sparse stochastic block model The parameters $(\alpha, \boldsymbol{Q})$ of the stochastic block model are given by $\alpha = (1/3, 1/3, 1/3)$, and

$$\boldsymbol{Q} = \rho \begin{pmatrix} 0.5 & 0.2 & 0.2 \\ 0.2 & 0.5 & 0.2 \\ 0.2 & 0.2 & 0.5 \end{pmatrix}$$

for ρ ranging between 0.05 and 1. For each sparsity, we simulate 100 networks with 500 nodes. For each networks, entries of the adjacency matrix are observed independently from one another with probability 1/2. Then, the matrix of connection probabilities $\boldsymbol{\Theta}^*$ is estimated using each method (variational approximation to the maximum likelihood estimator, missSBM, softImpute, the oracle estimator and the naive estimator).

Stochastic block model with missing observations The parameters $(\alpha, \boldsymbol{Q})$ of the stochastic block model are given by
$\alpha = (1/3, 1/3, 1/3)$, and

$$\boldsymbol{Q} = \begin{pmatrix} 0.5 & 0.2 & 0.2 \\ 0.2 & 0.5 & 0.2 \\ 0.2 & 0.2 & 0.5 \end{pmatrix}$$

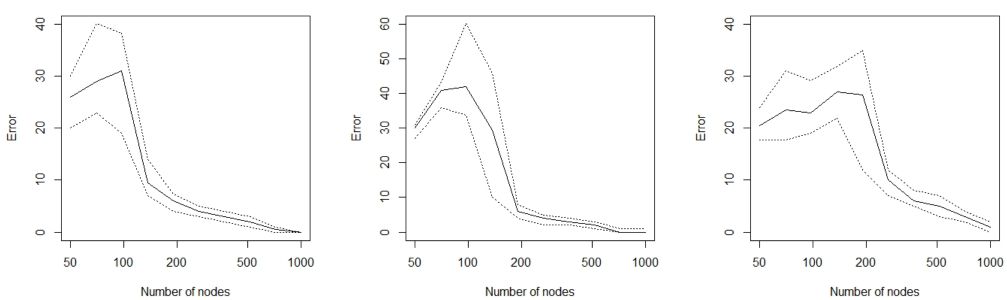
The proportion of observed entries p varies between 0.02 and 1. For each p, we simulate 100 networks with 500 nodes. For each networks, entries of the adjacency matrix are observed independently from one another with probability p. Then, the matrix of connection probabilities $\boldsymbol{\Theta}^*$ is estimated using each method (variational approximation to the maximum likelihood estimator, missSBM, softImpute, the oracle estimator and the naive estimator).

B.2 Empirical strong consistency of the variational estimator

We illustrate the empirical strong consistency of the variational estimator. Using the parameters chosen for simulating dense stochastic block models, we compute the number of misclassified nodes, defined as

$$\min_{z \sim \hat{z}} \left\{ \sum_i \mathbb{1}\left\{ z^*(i) \neq z(i) \right\} \right\}.$$

The total classification error for the assortative, dissasortartive and mixed models are presented in Figure 2. These simulations confirm that the variational estimator achieves strong recovery of the labels, even in unbalanced setting when neither assortative or disassortative behaviour are observed.



(a) Assortative SBM. (b) Disassortative SBM. (c) Mixed SBM.

Figure 2: Number of nodes misclassified by the variational estimator in the assortative SBM with balanced communities (left), in the disassortative SBM with balanced communities (middle), and in the mixed SBM with unbalanced communities (right). The full lines indicate the median of the number of misclassified nodes over 100 repetitions, while the dashed lines indicate its 25% and 75% quantiles.

B.3 Prediction of interactions within an elementary school

To compare the errors in term of link prediction of the methods missSBM and softImpute with that of our estimator, we plot the precision-recall curves of these estimators. More precisely, for any estimator $\widehat{\boldsymbol{\Theta}}$ of the matrix of connection probabilities $\boldsymbol{\Theta}^*$, and all thresholds $t \in [0, 1]$, one can define the link-prediction estimator $\widehat{A}$ as follows : $\widehat{A}_{ij} = 1$ if and only if $\widehat{\boldsymbol{\Theta}}_{ij} \geq t$, that is, we predict that there exists a link between nodes i and j is the estimated probability that these nodes are connected is larger than the threshold t. The recall-precision curves obtained by varying this threshold is presented in Figure 3. We also represent the mean precision-recall curve of the baseline estimator obtained by predicting edges independently at random with an increasing probability.

The three methods used for link prediction obtain quite similar precision-recall curves. No single method is better across all sensitivity levels.

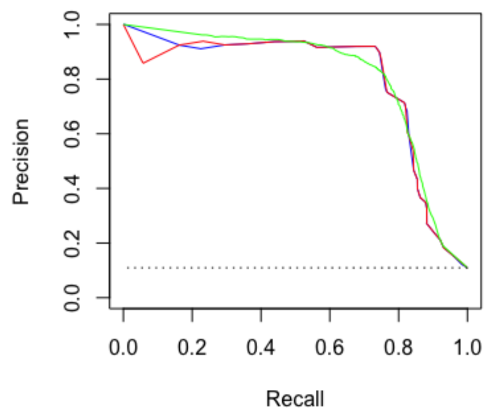


Figure 3: **Precision-recall curves for link prediction in the network of interactions within a school**: Precision-recall curves of the estimator obtained using `missSBM` (in red), of the estimator obtained using `softImpute` (in green), and of the variational approximation to the maximum likelihood estimator (in blue). The dotted black line represents the precision of the baseline estimator.

B.4 Prediction of collaboration in the co-authorship network

Similarly, we plot the precision-recall curves of the link-prediction methods obtained by using our new estimator, missSBM and softImpute. We also represent the mean precision-recall curve of the baseline estimator obtained by predicting edges independently at random with an increasing probability. The recall-precision curves is presented in Figure 4.

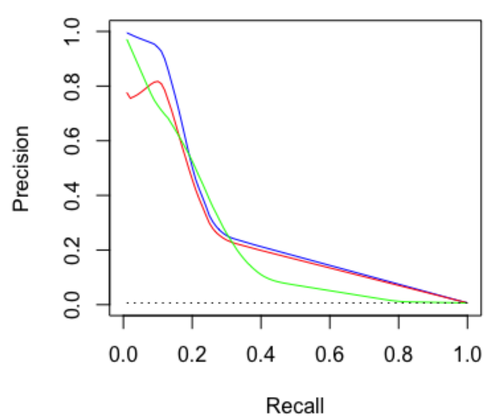


Figure 4: **Precision-recall curves for link prediction in the network co-authorship**: Precision-recall curves of the estimator obtained using `missSBM` (in red), of the estimator obtained using `softImpute` (in green), and of the variational approximation to the maximum likelihood estimator (in blue). The dotted black line represents the precision of the baseline estimator.

The precision-recall curve of the variational approximation to the maximum likelihood estimator is equivalent to or better than the other estimators across all sensitivity levels.

C EM algorithm for the variational estimator

The expectation - maximization (EM) algorithm derived in [48] can be used to iteratively compute the variational estimator. This algorithm alternates between the following two steps :

- Estimation Step: given parameters $(\alpha, \boldsymbol{Q})$, the variational parameter τ maximizing $\mathcal{J}_{\boldsymbol{X}}(\boldsymbol{A}; \tau, \alpha, \boldsymbol{Q})$ is given by the fixed point equation :

  $$\tau_a^i = c_i \alpha_a \prod_{j \neq i : \boldsymbol{X}_{ij}=1} \prod_{b \leq k} \left(\boldsymbol{Q}_{ab}^{\boldsymbol{A}_{ij}} \left(1 - \boldsymbol{Q}_{ab}\right)^{1-\boldsymbol{A}_{ij}} \right)^{\tau_b^j} \quad \text{where } c_i \text{ is a normalizing constant;}$$

- Maximization Step: given parameter τ, the parameters $(\alpha, \boldsymbol{Q})$ maximizing $\mathcal{J}_{\boldsymbol{X}}(\boldsymbol{A}; \tau, \alpha, \boldsymbol{Q})$ are given by

  $$\alpha_a = \frac{\sum_i \tau_a^i}{n}\ , \boldsymbol{Q}_{ab} = \frac{\sum_{i \neq j} \boldsymbol{X}_{ij} \tau_a^i \tau_b^j \boldsymbol{A}_{ij}}{\sum_{i \neq j} \boldsymbol{X}_{ij} \tau_a^i \tau_b^j}.$$

Note that this algorithm is not guaranteed to converge to a global maximum. To circumvent this problem, the authors of [48] suggest to initialize the weights τ_a^i using a first clustering step.