An iterative clustering algorithm for the Contextual Stochastic Block Model with optimality guarantees

Guillaume Braun & Hemant Tyagi & Christophe Biernacki

November 29, 2021

Abstract

Real-world networks often come with side information that can help to improve the performance of network analysis tasks such as clustering. Despite a large number of empirical and theoretical studies conducted on network clustering methods during the past decade, the added value of side information and the methods used to incorporate it optimally in clustering algorithms are relatively less understood. We propose a new iterative algorithm to cluster networks with side information for nodes (in the form of covariates) and show that our algorithm is optimal under the Contextual Symmetric Stochastic Block Model. Our algorithm can be applied to general Contextual Stochastic Block Models and avoids hyperparameter tuning in contrast to previously proposed methods. We confirm our theoretical results on synthetic data experiments where our algorithm significantly outperforms other methods, and show that it can also be applied to signed graphs. Finally we demonstrate the practical interest of our method on real data.

1 Introduction

The Stochastic Block Model (SBM) is a popular generative model for random graphs – introduced by Holland et al. (1983) – which captures the community structures of networks often observed in the real world. Here, edges are independent Bernoulli random variables with the probability of connection between two nodes depending only on the communities to which they belong. It is typically used as a benchmark to measure the performance of clustering algorithms. However, real-world networks often come with side information in the form of nodes covariates which can often be used to improve clustering performance. The Contextual Stochastic Block Model (CSBM) is a simple extension of the SBM that incorporates such side information: each node is associated with a Gaussian vector of parameters depending only on the community to which the node belongs; see Section 2 for details.

Several variants of this model and clustering algorithms have been proposed in the literature. These methods include model-based approaches (Mele et al., 2019; Weng and Feng, 2016), spectral methods (Binkiewicz et al., 2017), modularity based optimization methods (Zhang et al., 2015) and semidefinite programming (SDP) based approaches (Yan and Sarkar, 2020). Even if some of these algorithms come with some theoretical guarantees, the added value of side information is not well understood. The recent works of Abbe et al. (2020), Lu and Sen (2020) and Ma and Nandy (2021) clarify the situation by establishing information theoretic thresholds for exact recovery and detection in a particular case. However, the algorithm presented in the former work is not likely to be extended to a general CSBM with more than two (possibly unequal-sized) communities, while the latter two results focus on detection rather than consistency.

Our contributions. We make the following contributions in this paper.

- We propose a new iterative algorithm for clustering networks that is fast and is applicable to various settings including the general CSBM and also signed weighted graphs as shown in experiments.
- The proposed algorithm is analyzed under the Contextual Symmetric SBM (CSSBM) and we show that its rate of convergence is statistically optimal. As a byproduct, we derive the threshold for exact recovery with K communities under the CSSBM, thus extending the recent result of Abbe et al. (2020) which was obtained in a slightly different setting for $K = 2$.

- We confirm the theoretical properties of our algorithm through experiments on simulated data showing that our method outperforms existing algorithms, not only under the CSBM but also under the Signed SBM the latter of which models community structure in signed networks (Cucuringu et al., 2019). Finally, we provided a real data application of our algorithm.

Related work. Our iterative method can be thought of as a Classification-EM algorithm (Celeux and Govaert, 1992), hereafter referred to as C-EM, where instead of using the likelihood we used a least squares criterion. Such ideas were first applied and analyzed under various models including associative SBM by Lu and Zhou (2016) and then extended to a general method by Gao and Zhang (2019). Recently, such ideas were also successfully applied to the Gaussian Tensor Block Model (Han et al., 2020) and a general Gaussian Mixture Model (GMM) (Chen and Zhang, 2021). However, the previously obtained results can not be directly applied to our framework and several adaptations are required due to dependencies arising in the SBM, along with the heterogeneity of the data. Iterative refinement algorithms can also be derived naturally from the Power Method (Wang et al., 2021; Ndaoud et al., 2019) or alternative optimization methods (Chi et al., 2019).

Notation. We use lowercase letters ($\epsilon, a, b, \ldots$) to denote scalars and vectors, except for universal constants that will be denoted by $c_1, c_2, \ldots$ for lower bounds, and $C_1, C_2, \ldots$ for upper bounds and some random variables. We will sometimes use the notation $a_n \lesssim b_n$ (or $a_n \gtrsim b_n$) for sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$ if there is a constant $C > 0$ such that $a_n \leq Cb_n$ (resp. $a_n \geq Cb_n$) for all n. If $a_n \lesssim b_n$ and $a_n \gtrsim b_n$, then we write $a_n \asymp b_n$. Matrices will be denoted by uppercase letters. The i-th row of a matrix A will be denoted as $A_{i:}$ and depending on the context can be interpreted as a column vector. The column j of A will be denoted by $A_{:j}$, and the (i, j)th entry by A_{ij}. The transpose of A is denoted by $A^\top$ and $A_{:j}^\top$ corresponds to the jth row of $A^\top$ by convention. I_k denotes the $k \times k$ identity matrix. For matrices, we use $||.||$ and $||.||_F$ to respectively denote the spectral norm (or Euclidean norm in case of vectors) and Frobenius norm. The number of non zero entries of a matrix A is denoted $nnz(A)$.

2 The statistical framework

The CSBM consists of a graph encoded in an adjacency matrix $A \in \{0,1\}^{n\times n}$ and nodes covariates forming a matrix $X = [X_1\ X_2 \cdots X_n]^\top \in \mathbb{R}^{n\times d}$ where d is the dimension of the covariate space. The graph and the covariates are generated as follows.

The graph part of the data is generated from a Stochastic Block Model (SBM) which is defined by the following parameters.
- The set of nodes $\mathcal{N} = [n]$.
- Communities $\mathcal{C}_1, \ldots, \mathcal{C}_K$, of respective sizes $n_1, \ldots, n_K$, forming a partition of $\mathcal{N}$.
- A membership matrix $Z \in \mathcal{M}_{n,K}$ where $\mathcal{M}_{n,K}$ denotes the class of membership matrices. Here, $Z_{ik} = 1$ if node i belongs to $\mathcal{C}_k$, and is 0 otherwise. Each membership matrix Z can be associated bijectively with a function $z : [n] \to [K]$ such that $z(i) = z_i = k$ where k is the unique column index satisfying $Z_{ik} = 1$. To each matrix $Z \in \mathcal{M}_{n,K}$ we can associate a matrix W by normalizing the columns of Z in the ℓ_1 norm: $W = ZD^{-1}$ where $D = \mathrm{diag}(n_1, \ldots, n_K)$. This implies that $W^\top Z = I_K = Z^\top W$.
- A symmetric, connectivity matrix of probabilities between communities

$$\Pi = (\pi_{kk'})_{k,k'\in[K]} \in [0,1]^{K\times K}.$$

We additionally assume that the communities are approximately well balanced, i.e.,

$$\frac{n}{\alpha K} \leq n_k \leq \frac{\alpha n}{K} \quad \forall k \in [K],$$

for some constant $\alpha > 1$. Denoting $P = (p_{ij})_{i,j\in[n]} := Z\Pi Z^T$, a graph $\mathcal{G}$ is distributed according to $\mathrm{SBM}(Z, \Pi)$ if the entries of the corresponding symmetric adjacency matrix A are generated by

$$A_{ij} \overset{\text{ind.}}{\sim} \mathcal{B}(p_{ij}), \quad 1 \leq i \leq j \leq n,$$

where $\mathcal{B}(p)$ denotes a Bernoulli distribution with parameter p. Hence the probability that two nodes are connected depends only on their community memberships. We will frequently use the notation E for the centered noise matrix defined as $E_{ij} = A_{ij} - p_{ij}$, and denote the maximum entry of P by $p_{max} = \max_{i,j} p_{ij}$. The latter can be interpreted as the sparsity level of the graph.

Assumption 1. *Throughout this work we will assume that $p_{max} \asymp \log n/n$.*

If the graph is denser, we are in the exact recovery regime and the problem is easy. If we are in a sparser regime, we would need to regularize the adjacency matrix to enforce concentration, but we prefer to avoid this additional technical difficulty.

For the analysis, we will also consider a special case of the SBM where the communities are equal sized, i.e., $n_k = n/K$ for all $k \in [K]$, and the connectivity matrix is given by

$$\Pi = \begin{pmatrix} p & q & \dots & q \\ q & p & \dots & q \\ \vdots & \vdots & \ddots & \vdots \\ q & q & \dots & p \end{pmatrix} \in [0,1]^{K \times K}.$$

We will further assume that $p = p' \frac{\log n}{n}$ and $q = q' \frac{\log n}{n}$ for constants p', q' such that $p' > q'$. This model will be referred to as the Symmetric SBM and denoted by $\mathrm{SSBM}(p, q, n, K)$.

The **nodes covariates** are generated by a Gaussian Mixture Model (GMM), independent of A conditionally on the partition Z. More formally, for each i,

$$X_i = \mu_{z_i} + \epsilon_i, \text{ where } \epsilon_i \overset{\text{ind.}}{\sim} \mathcal{N}(0, \sigma^2 I_d)$$

with $\mu_k \in \mathbb{R}^d$ for all $k \leq K$ and $\sigma > 0$. We assume that $d = O(n)$. For the ease of exposition we further assume that σ is known but our method can be extended to anisotropic GMM with unknown variance as in Chen and Zhang (2021).

The **misclustering rate** associated to an estimated partition $\hat{z}$ quantifies the number of nodes assigned to a wrong cluster and is formally defined by

$$r(\hat{z}, z) = \frac{1}{n} \min_{\pi \in \mathfrak{S}} \sum_{i \in [n]} \mathbf{1}_{\{\hat{z}(i) \neq \pi(z(i))\}},$$

where $\mathfrak{S}$ denotes the set of permutations on $[K]$. We say that we are in the exact recovery regime if $r(\hat{Z}, Z) = 0$ with probability $1 - o(1)$ as n tends to infinity. If $\mathbb{P}(r(\hat{Z}, Z) = o(1)) = 1 - o(1)$ as n tends to infinity then we are is the weak consistency regime. A more complete overview of the different types of consistency and the sparsity regimes where they occur can be found in Abbe (2018).

3 How to integrate heterogeneous sources of information?

The use of side information should intuitively help to recover clusters that are not well separated on each individual source of information. However, it is not well understood how to integrate two heterogeneous sources of information in the clustering process. Previous attempts (Binkiewicz et al., 2017; Yan and Sarkar, 2020) proceed by directly aggregating the adjacency matrix and a Gram matrix (or Kernel matrix) formed by the covariates, but a lot of information can be lost in the aggregation process. Moreover it is not clear what is the best linear combination of the two matrices. Here we propose a different approach based on a two step algorithm (see Algorithm 1) that fully exploits all information. In the first step we obtain a rough estimate of the model parameters from the previous estimate of the partition; the initialization methods that can be used are discussed in Section 3.2. Then, in the second step, we iteratively refine the partition, as further explained in Section 3.1. In Section 5.1 we illustrate via experiments that Algorithm 1 outperforms existing methods for cluster recovery in the setting where the clusters are insufficiently separated on a single source of information.

Algorithm 1 Iterative Refinement with Least Squares

Input: $A \in \mathbb{R}^{n\times n}$, $X \in \mathbb{R}^{n\times d}$,$K \in \mathbb{N}^*$, $\sigma > 0$, $Z^{(0)} \in \{0,1\}^{n\times K}$ a membership matrix and $T \geq 1$.

for $0 \leq t \leq T-1$ **do**

Given $Z^{(t)}$, estimate the model parameters: $n_k^{(t)} = |\mathcal{C}_k^{(t)}|$, $W^{(t)} = Z^{(t)}(D^{(t)})^{-1}$ where $D^{(t)} = \mathrm{diag}(n_k^{(t)})_{k\in[K]}$, $\Pi^{(t)} = W^{(t)\top} A W^{(t)}$, and $\mu_k^{(t)} = W_{k:}^{(t)\top} X$, for all $k \leq K$.

Refine the partition by solving for each $i \in [n]$

$$z_i^{(t+1)} = \arg\min_k ||(A_{i:}W^{(t)} - \Pi_{k:}^{(t)})\sqrt{\Sigma_k^{(t)}}||^2 + \frac{||X_i - \mu_k^{(t)}||^2}{\sigma^2}$$

where

$$\Sigma_k^{(t)} = \begin{cases} \mathrm{diag}(\frac{n_{k'}^{(t)}}{\Pi_{kk'}^{(t)}})_{k'\in[K]} & (\texttt{IR-LS}) \\ \frac{\min_{k'} n_{k'}^{(t)}}{\max_{k',k''} \Pi_{k'k''}^{(t)}} I_K & (\texttt{sIR-LS}) \\ \frac{n}{K(p^{(t)}-q^{(t)})} \log(\frac{p^{(t)}(1-q^{(t)})}{q^{(t)}(1-p^{(t)})}) I_K & (\texttt{IR-LSS}) \end{cases}$$

with $p^{(t)} = K^{-1}\sum_{k\in[K]} \Pi_{kk}^{(t)}$ and $q^{(t)} = (K^2-K)^{-1}\sum_{k\neq k'\in[K]} \Pi_{kk'}^{(t)}$.

Form the matrices $Z^{(t+1)}$ from $z^{(t+1)}$.

end for

Output: A partition of the nodes $Z^{(T)}$.

3.1 The refinement mechanism

At each step t, Algorithm 1 estimates the model parameters given a current estimate of the partition ($W^{(t)}$), then updates the partition by reassigning each node to its closest community. Here, the proximity of a node i to a community k is measured by the distance between its estimated (graph) connectivity profile ($A_{i:}W^{(t)}$) and its covariates (X_i) to the current estimate of the community parameters ($\Pi_{k:}^{(t)}, \mu_k^{(t)}$). Instead of using the Maximum A Priori (MAP) estimator as in C-EM algorithms, we use a least-square criterion. In a model-based perspective, this can be interpreted as a Gaussian approximation of the SBM. We will see later in the experiments that this doesn't lead to a loss of accuracy, and is also faster.

Different variants of our algorithm are possible depending on the way the variance of each community is estimated and integrated in the criterion used for the partition refinement. The general method will be referred as `IR-LS`, the simplified spherical version is denoted by `sIR-LS` and the version of the algorithm used for CSSBM is denoted by `IR-LSS`.

Computational cost. In each iteration, the complexity of estimating the parameters is $O(nnz(A)+nd)$ while that of estimating the partition is $O(nK(K+d))$. So the total cost of `IR-LS` is $O(T(nnz(A) + nK(\max(K,d)))$. In our setting A is sparse, hence $nnz(A) \asymp n\log n$.

Remark 1. *Algorithm 1 can also be used for clustering weighted signed graphs, as shown later in the experiments. Moreover, it is interesting to note that when there are no covariates, the algorithm can be applied to graphs generated from a general SBM. This is in contrast to the iterative algorithm proposed by Lu and Zhou (2016) that can only be applied to assortative SBMs (see appendix).*

3.2 Initialization

Different strategies can be adopted for initialization. If we assume that the communities are separated on each source of information and that the signal-to-noise ratio (SNR) is large enough to recover a sufficient proportion in each cluster, we can use a spectral method on one source of information (the graph for example). However, it is in general better to combine both sources of information. While one could use the methods proposed in Yan and Sarkar (2020) or Binkiewicz et al. (2017) that also come with some theoretical guarantees, we instead use Algorithm 2 to initialize the partition. This algorithm will be referred to as `EM-Emb`. In our experiments, we used the package `clusterR` (Mouselimis, 2021) for estimating the Gaussian mixture with an EM algorithm. This algorithm is fast, provides a sufficiently accurate estimate of the partition, and avoids hyperparameter tuning.

Algorithm 2 EM on graph embedding and covariates (`EM-Emb`)

Input: The number of communities K, the adjacency matrix A, covariates X.

1: Compute $U_K \in \mathbb{R}^{n\times K}$ the matrix formed by the eigenvectors associated with the top-K eigenvalues (in absolute order) of A.
2: Merge the columns of U_K with the columns of X to obtain a matrix X'.
3: Cluster the rows of X' by using an EM algorithm for GMM.

Output: A partition of the nodes $Z^{(0)}$.

4 Theoretical analysis

In this section, we analyze the variants `sIR-LS` and `IR-LSS` of Algorithm 1. While it is possible to extend the analysis to `IR-LS`, it would be considerably more technical and tedious due to its non- spherical structure. Hence, we will assume here that the covariance matrix $\Sigma_k^{(t)}$ in Algorithm 1 has the form $\lambda^{(t)} I_k$ where $\lambda^{(t)}$ is an appropriate scalar depending on whether we use `sIR-LS` or `IR-LSS`.

In Section 4.1 we will present the general principle for the analysis. Then we will specialize it for analyzing `IR-LSS` (under the CSSBM) in Section 4.2, and prove that the convergence rate obtained is optimal in Section 4.3. Finally we show that the same framework can be used to bound the convergence rate of `sIR-LS` (under the CSBM) in Section 4.4. The details of the proofs are outlined in the appendix.

4.1 Analysis principle

Our analysis is motivated by the general framework recently developed by Gao and Zhang (2019), and also borrows some decomposition techniques used for analyzing Gaussian tensors from Han et al. (2020). However, these results are not directly applicable to our setting due to dependencies arising from symmetry in the SBM. Moreover, we need a tighter control of the error terms then provided by these works.

We will assume w.l.o.g. that $\sigma = 1$ (since σ is assumed to be known in our framework) and that the permutation π that minimizes the distance between $z^{(0)}$ and z is the identity (if not, then replace z by $\pi^{-1}(z)$). Hence there is no label switching ambiguity in the community labels of $z^{(t)}$ because they are determined from $z^{(0)}$.

The first step is to analyze the event "after one refinement step, the node i will be incorrectly clustered given the current estimation of the partition $z^{(t)}$ at time t". This corresponds to the condition

$$a \neq \arg\min_k ||X_i - \mu_k^{(t)}||^2 + \hat{\lambda}^{(t)} ||A_{i:}W^{(t)} - \Pi_{k:}^{(t)}||^2$$

for a node i such that $z_i = a$. One can see that this condition is equivalent to the existence of $b \in [K] \setminus a$ such that

$$\underbrace{\langle \epsilon_i, \tilde{\mu}_a - \tilde{\mu}_b\rangle + \lambda\langle E_{i:}W, \tilde{\Pi}_{a:} - \tilde{\Pi}_{b:}\rangle}_{C_i(a,b)} \leq \frac{-\Delta^2(a,b)}{2} + Err_{ib}^{(t)}.$$

Here,

$$\Delta^2(a,b) = ||\mu_a - \mu_b||^2 + \lambda||\Pi_{a:} - \Pi_{b:}||^2,$$
$$\tilde{\mu}_k = X^\top W_{:a},\ \tilde{\Pi}_{k:} = W_{k:}^\top AW,$$
$$\text{and } \lambda = \frac{n_{min}}{p_{max}} \text{ or } \frac{n}{K(p-q)} \log\left(\frac{p(1-q)}{q(1-p)}\right)$$

depending whether we are analyzing `sIR-LS` or `IR-LSS`. Moreover, $Err_{ib}^{(t)}$ is an error term that can be further decomposed as a sum $F_{ib}^{(t)} + G_{ib}^{(t)} + H_{ib}^{(t)}$ of different kinds of error terms which will be controlled in different ways. If we ignore the error term, we obtain the condition corresponding to having an incorrect result after one iteration starting from the ground truth partition. The errors occurring in this way will be quantified by the **ideal oracle error**

$$\xi(\delta) = \sum_{i=1}^{n} \sum_{b\in[K]\setminus z_i} \Delta^2(z_i, b)\mathbf{1}_{\{C_i(a,b) \le \frac{-(1-\delta)\Delta^2(z_i,b)}{2}\}}.$$

Let us denote

$$\Delta_{min} = \min_{a\neq b\in[K]} \Delta(a, b)$$

to quantify the separation of the parameters associated with the different communities. If $\Delta_{min} = 0$, it would imply that at least two communities are indistinguishable and the model would not be identifiable. For $t \ge 1$, let

$$\tau^{(t)} = \max(\frac{7}{8}\tau^{(t-1)}, \tau),\ \delta^{(t)} = \max(\frac{2cK\tau^{(t-1)}}{n\Delta^2_{min}}, \delta)$$

be sequences where $\tau^{(0)}, c > 0$ and $\tau \ge 0$. Moreover, $\delta = \delta(n)$ converges to zero at a suitably slow rate.

In general the rate of decay of $\xi(\delta)$ leads to the convergence rate of iterative refinement algorithms, hence it is important to control this quantity.

Condition A (ideal error)**.** *Assume that*

$$\xi(\delta^{(t)}) \le \frac{3}{4}\tau^{(t-1)},\ \textit{for all } t \ge 1$$

holds with probability at least $1 - \eta_1$*.*

We now have to analyze the error terms and prove that their contribution is negligible compared to the ideal oracle error rate. Let

$$l(z, z') = \sum_{i\in[n]} \Delta^2(z_i, z'_i)\mathbf{1}_{\{z_i\neq z'_i\}}$$

be a measure of distance between two partitions $z, z' \in [K]^n$. We will control the error terms by showing that the following conditions are satisfied.

Condition B (F-error type)**.** *Assume that*

$$\max_{\{z^{(t)}:l(z,z^{(t)})\le\tau^{(0)}\}} \sum_{i=1}^{n} \max_{b\in[K]\setminus z_i} \frac{(F^{(t)}_{ib})^2}{\Delta^2(z_i, b)l(z, z^{(t)})} \le \frac{\delta^2}{256}$$

for all $t \ge 0$ *holds with probability at least* $1 - \eta_2$*.*

Condition C (GH-error type)**.** *Assume that*

$$\max_{i\in[n]} \max_{b\in[K]\setminus z_i} \frac{|H^{(t)}_{ib}| + |G^{(t)}_{ib}|}{\Delta(z_i, b)^2} \le \frac{\delta^{(t)}}{4}$$

holds uniformly on the event $\{z^{(t)} : l(z, z^{(t)}) \le \tau^{(t)}\}$ *for all* $t \ge 0$ *with probability at least* $1 - \eta_3$ *.*

We can now show under these conditions that there is a contraction of the error if the initial estimate of the partition is close enough to the ground truth partition.

Theorem 1. *Assume that*

$$l(z^{(0)}, z) \le \tau^{(0)}$$

for some constant $\tau^{(0)}$ *such that* $\delta^{(1)} < 1$ *and additionally assume that Conditions A, B, and C hold. Then with probability at least* $1 - \sum_{i=1}^{3} \eta_i$

$$l(z^{(t)}, z) \le \xi(\delta^{(t)}) + \frac{1}{8}l(z^{(t-1)}, z)\ \textit{ for all } t \ge 1.$$

Remark 2. *This is an adaptation of Theorem 3.1 in Gao and Zhang (2019) where we allow at each step to choose a different* $\delta^{(t)}$. *It allows us to obtain a weaker condition for initialization than the one used in (Gao and Zhang, 2019, Theorem 4.1). Indeed, they require* $l(z^{(0)}, z) = o(\frac{n\Delta_{min}^2}{K})$*, but we only need* $l(z^{(0)}, z) = O(\frac{n\Delta_{min}^2}{K})$.

Proof of Theorem 1. By choice of $\tau^{(0)}$ we can assume that $\delta^{(1)} < 1$. Since by definition $\tau^{(t)}$ is decreasing, so is $\delta^{(t)}$, hence $\delta^{(t)} < 1$ for all $t \geq 1$. Let $i \in [n]$ such that $z_i = a$ and assume that $l(z^{(t-1)}, z) \leq \tau^{(t-1)}$ for some given $t \geq 1$. Denoting $I_i^{(t)}(a, b) := \mathbf{1}_{\{C_i(a,b) \leq \frac{-(1-\delta^{(t)})\Delta^2(a,b)}{2}\}}$, observe that

$$
\begin{aligned}
&\mathbf{1}_{\{z_i^{(t)}=b\}} \\
&\overset{(1)}{\leq} \mathbf{1}_{\{C_i(a,b) \leq \frac{-\Delta^2(a,b)}{2} + F_{ib}^{(t-1)} + G_{ib}^{(t-1)} + H_{ib}^{(t-1)}\}} \\
&\overset{(2)}{\leq} I_i^{(t)}(a, b) + \mathbf{1}_{\{\frac{\delta^{(t)}}{2}\Delta^2(a,b) \leq F_{ib}^{(t-1)} + G_{ib}^{(t-1)} + H_{ib}^{(t-1)}\}} \\
&\overset{(3)}{\leq} I_i^{(t)}(a, b) + \mathbf{1}_{\{\frac{\delta^{(t)}}{4}\Delta^2(a,b) \leq F_{ib}^{(t-1)}\}} \\
&\leq I_i^{(t)}(a, b) + \mathbf{1}_{\{\frac{\delta}{4}\Delta^2(a,b) \leq F_{ib}^{(t-1)}\}} \\
&\overset{(4)}{\leq} I_i^{(t)}(a, b) + \frac{32(F_{ib}^{(t-1)})^2}{\delta^2 \Delta^4(a, b)}.
\end{aligned}
$$

The inequality (1) follows from the definition of $z_i^{(t)}$ and the error decomposition. Inequality (2) comes from a union bound while (3) uses Condition C. Finally, (4) follows from Markov inequality. Hence,

$$
\begin{aligned}
l(z^{(t)}, z) &= \sum_{i \in [n]} \sum_{b \in [K] \setminus \{z_i\}} \Delta^2(z_i, b) \mathbf{1}_{\{z_i^{(t)}=b\}} \\
&\leq \sum_{i \in [n]} \sum_{b \in [K] \setminus \{z_i\}} \Delta^2(z_i, b) \mathbf{1}_{\{C_i(a,b) \leq \frac{-(1-\delta^{(t)})\Delta^2(z_i,b)}{2}\}} \\
&+ \sum_{i \in [n]} \sum_{b \in [K] \setminus \{z_i\}} \Delta^2(z_i, b) \mathbf{1}_{\{z_i^{(t)}=b\}} \frac{32(F_{ib}^{(t-1)})^2}{\delta^2 \Delta^4(z_i, b)} \\
&\leq \xi(\delta^{(t)}) + \sum_{i \in [n]} \max_{b \in [K] \setminus \{z_i\}} \frac{32(F_{ib}^{(t-1)})^2}{\delta^2 \Delta^2(z_i, b)} \\
&\leq \xi(\delta^{(t)}) + \frac{1}{8} l(z^{(t-1)}, z).
\end{aligned}
$$

Using Condition A, we hence obtain

$$
l(z^{(t)}, z) \leq \xi(\delta^{(t)}) + \frac{1}{8}\tau^{(t-1)} \leq \frac{7}{8}\tau^{(t-1)}.
$$

Thus $\tau^{(t)}$ is an upper bound for $l(z^{(t)}, z)$ and the theorem is proved by induction. □

4.2 Convergence guarantees for `IR-LSS` under the CSSBM

Let us define the SNR

$$
\tilde{\Delta}^2 = \frac{1}{8} \min_{k \neq k'} ||\mu_k - \mu_{k'}||^2 + \frac{\log n}{K} (\sqrt{p'} - \sqrt{q'})^2.
$$

It is easy to see that $\tilde{\Delta} \asymp \Delta_{min}$. The following lemma shows that $\xi(\delta)$ decreases exponentially fast in $\tilde{\Delta}$ provided Δ_{min} is suitably large.

Lemma 1. *Assume that $K^{1.5}/\Delta_{min} \to 0$. Then with probability at least $1-\exp(-\tilde{\Delta})$, we have*

$$\xi(\delta) \leq n\exp(-(1+o(1))\tilde{\Delta}^2).$$

The following theorem shows that if $z^{(0)}$ is close enough to z, then the misclustering rate decreases exponentially fast with the SNR $\tilde{\Delta}$ after $O(\log n)$ iterations.

Theorem 2. *Assume that $K^{1.5}/\Delta_{min} \to 0$ and $\tilde{\Delta}^2 \asymp \log n/K$. Under the CSSBM$(p,q,n,K)$ assumption, if $z^{(0)}$ is such that*

$$l(z, z^{(0)}) \leq \frac{\epsilon n \Delta_{min}^2}{K}$$

for a constant ϵ small enough, then with probability at least $1-n^{-\Omega(1)}$ we have for all $t \gtrsim \log n$

$$r(z^{(t)}, z) \leq \exp(-(1+o(1))\tilde{\Delta}^2).$$

Sketch of proof. We first show that Conditions B and C are satisfied. Then we show that Condition A is satisfied for the sequences $\delta^{(t)}$ and $\tau^{(t)}$, hence Theorem 1 can be applied to obtain a contraction of the error at each step. □

Remark 3. *By assumption $\tilde{\Delta}^2 \gtrsim \log n/K$, and so the condition $\tilde{\Delta}^2 \asymp \log n/K$ is not very restrictive. Indeed, if the information provided by the GMM part was not of the same order as the graph part, it would not be useful to aggregate information. If $\tilde{\Delta}^2 \gg \log n$ then we would be in the exact recovery setting and the problem becomes easy.*

4.3 Minimax lower-bound for CSSBM

We are going to establish that the convergence rate established in Theorem 2 is optimal. Let

$$\Theta = \{(\mu_k)_{k\in[K]} \in \mathbb{R}^K, p, q \in [0,1] \text{ such that } p > q\}$$

be the admissible parameter space.

Theorem 3. *Under the assumption $\tilde{\Delta}/\log K \to \infty$, we have*

$$\inf_{\hat{z}} \sup_{\theta\in\Theta} \mathbb{E}(r(\hat{z}, z)) \geq \exp(-(1+o(1))\tilde{\Delta}^2).$$

If $\tilde{\Delta} + \log K = O(1)$, then $\inf_{\hat{z}} \sup_{\theta\in\Theta} \mathbb{E}(\frac{r(\hat{z},z)}{n}) \geq c$ for some positive constant c.

Remark 4. *This lower-bound shows that if $\tilde{\Delta}^2 < \log n$ then every estimator fails to achieve exact recovery with a probability bounded below from zero because $\sup_{\theta\in\Theta} \mathbb{E}(r(\hat{z}, z)) > n^{\epsilon}$ for some $\epsilon > 0$. On the other hand, Theorem 2 shows that when $\tilde{\Delta}^2 > \log n$ then `IR-LSS` achieves exact recovery. Hence the threshold for exact recovery is $\tilde{\Delta}^2/\log n$. When $K = 2$ and $\mu_1 = -\mu_2 = \mu$ this matches the result obtained by Abbe et al. (2020).*

Sketch of proof. We can use the same argument as in (Lu and Zhou, 2016, Theorem 3.3) to reduce the problem to a hypothesis testing problem. The solution of the latter is given by the maximum likelihood test according to the Neyman-Pearson lemma. Then the probability of error can be controlled by using concentration inequalities. □

4.4 Convergence guarantees for `sIR-LS`

The proof techniques used in the previous section can be extended in a straightforward way to obtain consistency results for `sIR-LS` under the CSBM. The main difference is that the specialized concentration inequality used to prove Lemma 1 can no longer be applied to this setting.

Theorem 4. *Assume that $K^{1.5}/\Delta_{min} \to 0$, $\Delta^2_{min} \asymp \log n/K$ and $\max_{a,b\in[K]} \Delta^2(a,b) \lesssim \Delta^2_{min}$. Under the CSBM with approximately balanced communities, if*

$$l(z, z^{(0)}) \leq \frac{\epsilon n \Delta^2_{min}}{K}$$

for some small enough constant $\epsilon > 0$, then with probability at least $1 - n^{-\Omega(1)}$ we have for all $t \gtrsim \log n$

$$r(z^{(t)}, z) \leq \exp\left(-\frac{1}{8}\Delta^2_{min}\right).$$

5 Numerical experiments

We now empirically evaluate our method on both synthetic and real data. Section 5.1 contains simulations for the CSBM and Section 5.2 contains results for clustering signed networks under a Signed SBM. In Section 5.3, we test our method on a dataset consisting of a (weighted) signed graph along with covariate information for the nodes.

5.1 CSBM with not well separated communities

In this experiment the graph is generated from a SBM with parameters $n = 1000$, $K = 3$, $Z_i \overset{i.i.d}{\sim}$ Multinomial$(1; 1/3, 1/3, 1/3)$, and

$$\Pi = 0.02 * \begin{pmatrix} 1.6 & 1.2 & 0.5 \\ 1.2 & 1.6 & 0.5 \\ 0.05 & 0.05 & 1.2 \end{pmatrix}.$$

The covariates are generated from a GMM with variance $\sigma^2 = 0.2$ and class centers $\mu_1 = (0,0,1), \mu_2 = (-1,1,0)$, $\mu_3 = (0,0,1)$. Note that $\mathcal{C}_1, \mathcal{C}_3$ cannot be separated by the covariate information, while $\mathcal{C}_1, \mathcal{C}_2$ are not well separated in the graph information (as seen from Π). Hence, one would expect in this example that using only a single source of information should not yield good clustering results. To demonstrate this, we use the Normalized Mutual Information (NMI) criterion to measure the quality of the resulting clusters. It is an information theoretic measure of similarity taking values in $[0, 1]$, with 1 denoting a perfect match, and 0 denoting the absence of correlation between partitions.

Performance comparison. We will use `K-SC` and `L-SC` to denote respectively the results obtained by applying spectral clustering on the Gaussian kernel matrix K formed from the covariates, and spectral clustering (SC) applied on the Laplacian of the graph. Additionally, `SDP-Comb` refers to the method proposed by Yan and Sarkar (2020); `IR-MAP` is similar to `IR-LS` but with the least-square criterion replaced by the MAP to update the partition; `ORL-SC` (Oracle Regularized Laplacian SC) corresponds to SC applied on $A + \lambda K$ where λ is chosen to maximize the NMI between the (oracle known) true partition and the one obtained by using SC on $A + \lambda K$. For the implementation of `SDP-Comb`, we used the Matlab code provided by Yan and Sarkar (2020) with the λ given by `ORL-SC`. Figure 1 shows that the three iterative methods considered (`IR-MAP`, `IR-LS`, `sIR-LS`), initialized with `EM-Emb`, provide significantly better clustering performance compared to the other methods. The variance of `sIR-LS` is a bit larger than `IR-LS` and it also seems that `sIR-LS` `IR-MAP` are more sensitive to initialization (i.e., more outliers). On the other hand, other methods based on the aggregation of the two sources of information (`SDP-Comb` and `ORL-SC`) lead to a limited improvement in clustering performance.

Computational cost. We took the average of CPU time (in seconds) over 20 repetitions. There is an important gain in speed obtained by replacing the MAP objective by a least square criterion. Moreover, the initialization obtained with `EM-Emb` is very fast. The results are gathered in Table 1.

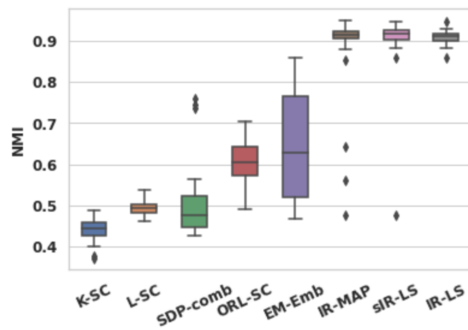

Figure 1: Average performances over 40 runs of different algorithms under CSBM.

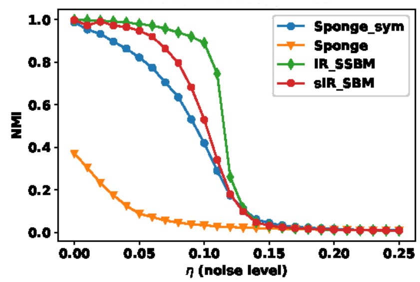

Figure 2: NMI versus η (noise) under signed SBM, $K = 20$, $n = 10000$, $p = 0.01$.

	L-SC	ORL-SC	EM-Emb	IR-LS	IR-MAP
Time	1.4	7.9	0.5	1.2	37.1
Ratio	2.7	15	1	2.3	70

Table 1: Comparison between computation times (averaged over 20 runs)

5.2 Signed SBM

A graph is generated from the Signed SBM as follows. First we generate an Erdös-Renyi graph where each edge appears with probability p and each edge takes the value 1 if both extremities are in the same community and -1 otherwise. Then we flip the sign of each edge independently with probability $\eta \in [0, 1/2)$. Our method `sIR-LS` can be directly applied to this setting, but we can also use the fact that the connectivity matrix Π is assortative to design a more specialised algorithm `IR-SSBM` (see appendix) that assigns a node to the community which maximizes its intra-connectivity estimated probability. For initialization, we use `Sponge-sym` (Cucuringu et al., 2019) for clustering signed graphs. Figure 2 shows that 20 refinement steps improves the clustering.

5.3 Australia Rainfall Dataset

We consider the time series data of historical rainfalls in locations throughout Australia, this was also studied in Cucuringu et al. (2019). Edge weights are obtained from the pairwise Pearson correlation, leading to a complete signed graph on $n = 306$ nodes. We use the longitude and latitude as covariates X, and `Sponge` (Cucuringu et al., 2019) to obtain an initial partition for `sIR-LS` and `Iter-SBM` (the version of `sIR-LS` without covariates). We exclude `IR-LS` here due to its relative instability on this dataset (see

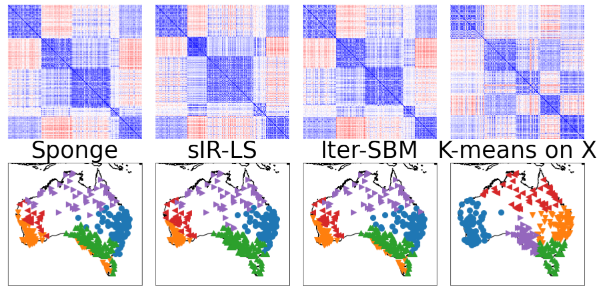

Figure 3: Sorted adjacency matrices and maps for Australian rainfall dataset ($K = 5$).

appendix). This shows that in some situations it can be better to use `sIR-LS` rather than `IR-LS`. Figure 3 illustrates the clustering obtained with `Sponge` (using only the graph), `sIR-LS` (integrating the covariates), `Iter-SBM` (refinement without covariates), and `K-means` applied on the covariates. The use of covariates in the refinement steps reinforces the geographical structure (orange points in the bottom right part of the map disappeared), increases the size of smallest cluster (the violet cluster on the three first maps), and strengthens the original clustering as seen in the sorted adjacency matrix, whereas `Iter-SBM` ignores the geography and `K-means` ignores the graph structure. Results for other choices of K are in the appendix.

6 Future work

We only analyzed `sIR-LS` and `IR-LSS` to reduce technicalities but we believe that the framework can be extended to analyze `IR-LS`. The principle we used to design our algorithm could also be applied to obtain a clustering method for bipartite graphs or multilayer networks. Another direction of research is to extend our method to more general graph models integrating common properties of real-life networks such as degree heterogeneity, mixed membership, presence of outliers and missing values.

Bibliography

E. Abbe. Community detection and stochastic block models. *Foundations and Trends® in Communications and Information Theory*, 14(1-2):1–162, 2018.

E. Abbe, J. Fan, and K. Wang. An ℓ_p theory of pca and spectral clustering, 2020.

A. S. Bandeira and R. van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *Ann. Probab.*, 44(4):2479–2506, 07 2016.

N. Binkiewicz, J. Vogelstein, and K. Rohe. Covariate-assisted spectral clustering. *Biometrika*, 104:361 – 377, 2017.

G. Celeux and G. Govaert. A classification em algorithm for clustering and two stochastic versions. *Computational Statistics & Data Analysis*, 14:315–332, 1992.

X. Chen and A. Zhang. Optimal clustering in anisotropic gaussian mixture models. *ArXiv*, 2021.

Y. Chi, Y. Lu, and Y. Chen. Nonconvex optimization meets low-rank matrix factorization: An overview. *IEEE Transactions on Signal Processing*, PP:1–1, 08 2019. doi: 10.1109/TSP.2019.2937282.

M. Cucuringu, P. Davies, A. Glielmo, and H. Tyagi. Sponge: A generalized eigenproblem for clustering signed networks. In *AISTATS*, 2019.

C. Gao and A. Zhang. Iterative algorithm for discrete structure recovery. *arXiv: Statistics Theory*, 2019.

C. Gao, Z. Ma, A. Zhang, and H. Zhou. Community detection in degree-corrected block models. *Annals of Statistics*, 46, 07 2016.

R. Han, Y. Luo, M. Wang, and A. Zhang. Exact clustering in tensor block model: Statistical optimality and computational limit. *ArXiv*, 2020.

P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2): 109 – 137, 1983.

J.-B. Léger. Blockmodels: A r-package for estimating in latent block model and stochastic block model, with various probability functions, with or without covariates. *arXiv: Computation*, 2016.

C. Lu and S. Sen. Contextual stochastic block model: Sharp thresholds and contiguity, 2020.

Y. Lu and H. H. Zhou. Statistical and computational guarantees of lloyd's algorithm and its variants. *ArXiv*, 2016.

Z. Ma and S. Nandy. Community detection with contextual multilayer networks, 2021.

A. Mele, L. Hao, J. Cape, and C. Priebe. Spectral inference for large stochastic blockmodels with nodal covariates. *SSRN Electronic Journal*, 2019.

L. Mouselimis. *ClusterR: Gaussian Mixture Models, K-Means, Mini-Batch-Kmeans, K-Medoids and Affinity Propagation Clustering*, 2021.

M. Ndaoud, S. Sigalla, and A. Tsybakov. Improved clustering algorithms for the bipartite stochastic block model. *arXiv: Statistics Theory*, 2019.

P. Wang, H. Liu, Z. Zhou, and A. M.-C. So. Optimal non-convex exact recovery in stochastic block model via projected power method. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 10828–10838. PMLR, 18–24 Jul 2021.

H. Weng and Y. Feng. Community detection with nodal information. *arXiv: Methodology*, 2016.

B. Yan and P. Sarkar. Covariate Regularized Community Detection in Sparse Graphs. *Journal of the American Statistical Association*, 0(0):1–12, 2020.

A. Y. Zhang and H. H. Zhou. Minimax Rates of Community Detection in Stochastic Block Models. *The Annals of Statistics*, 44(5):2252–2280, sep 2016. ISSN 00905364.

Y. Zhang, E. Levina, and J. Zhu. Community detection in networks with node features. *ArXiv*, 2015.

Supplementary Material

The proof of Lemma 1 in presented in Section A. Theorem 2 is proved in Section B and Theorem 3 is proved in Section C. The technical lemmas used in the proofs are gathered in Section D. Finally, Section E contains additional experiments results.

A Proof of Lemma 1

Let $\delta, \bar{\delta} > 0$. The ideal oracle error term can be upper bounded as follows

$$\xi(\delta) \leq \underbrace{\sum_{i=1}^{n} \sum_{b \in [K] \setminus z_i} \Delta^2(z_i, b) \mathbf{1}_{\{\langle \epsilon_i, \mu_{z_i} - \mu_b \rangle + \lambda \langle E_{i:} W, \Pi_{z_i:} - \Pi_{b:} \rangle \leq \frac{-(1-\delta-\bar{\delta})\Delta^2(z_i,b)}{2}\}}}_{M_1}$$
$$+ \underbrace{\sum_{i=1}^{n} \sum_{b \in [K] \setminus z_i} \Delta^2(z_i, b) \mathbf{1}_{\{\langle \epsilon_i, \tilde{\mu}_{z_i} - \mu_{z_i} \rangle + \lambda \langle E_{i:} W, \tilde{\Pi}_{z_i:} - \Pi_{z_i:} \rangle \leq \frac{-\bar{\delta}\Delta^2(z_i,b)}{4}\}}}_{M_2}$$
$$+ \underbrace{\sum_{i=1}^{n} \sum_{b \in [K] \setminus z_i} \Delta^2(z_i, b) \mathbf{1}_{\{\langle -\epsilon_i, \tilde{\mu}_{b} - \mu_{b} \rangle - \lambda \langle E_{i:} W, \tilde{\Pi}_{b:} - \Pi_{b:} \rangle \leq \frac{-\bar{\delta}\Delta^2(z_i,b)}{4}\}}}_{M_3}.$$

We will first obtain upper bounds for each $\mathbb{E}(M_i)$, $i = 1 \dots 3$. In particular we will show that the dominant term is $\mathbb{E}(M_1)$. Then, we will use Markov inequality to control $\xi(\delta)$ with high probability.

Upper bound of $\mathbb{E}(M_1)$. Let us denote for any given $i \in [n]$ and $b \in [K] \setminus z_i$ the event

$$\Omega_1 = \left\{ \langle \epsilon_i, \mu_{z_i} - \mu_b \rangle + \lambda \langle E_{i:} W, \Pi_{z_i:} - \Pi_{b:} \rangle \leq \frac{-(1-\delta-\bar{\delta})\Delta^2(z_i,b)}{2} \right\}.$$

By using an analogous argument as the one presented in Lemma 6 we obtain

$$\mathbb{P}(\Omega_1) \leq \exp(-(1+o(1))\tilde{\Delta}^2).$$

Thus by taking $\delta = \bar{\delta}$ going to zero as $n \to \infty$, we obtain

$$\begin{aligned}
\mathbb{E}(M_1) &\leq \sum_{i=1}^{n} \sum_{b \in [K] \setminus z_i} \Delta^2(z_i, b) \exp(-(1+o(1))\tilde{\Delta}^2) \\
&\leq nK \exp(-(1+o(1))\tilde{\Delta}^2) \\
&\leq n \exp(-(1+o(1))\tilde{\Delta}^2).
\end{aligned}$$

In the second line we used the fact that $\Delta^2(z_i, b) = \Delta_{min} \lesssim \tilde{\Delta}^2$ for all $z_i \neq b$ and $\tilde{\Delta} \asymp \sqrt{\log n/K} \to \infty$. In the third line we used the assumption $\tilde{\Delta}^2/\log(K) \to \infty$.

Upper bound of $\mathbb{E}(M_2)$. Let us denote for any given $i \in [n]$ and $b \in [K] \setminus z_i$ the events

$$\Omega_2 = \left\{ \langle \epsilon_i, \tilde{\mu}_{z_i} - \mu_{z_i} \rangle + \lambda \langle E_{i:} W, \tilde{\Pi}_{z_i:} - \Pi_{z_i:} \rangle \leq \frac{-\bar{\delta}\Delta^2(z_i,b)}{4} \right\},$$

$$\Omega_2' = \left\{ \langle \epsilon_i, \tilde{\mu}_{z_i} - \mu_{z_i} \rangle \leq \frac{-\bar{\delta}\Delta^2(z_i,b)}{8} \right\},$$

and

$$\Omega_2'' = \left\{ \lambda \langle E_{i:} W, \tilde{\Pi}_{z_i:} - \Pi_{z_i:} \rangle \leq \frac{-\bar{\delta}\Delta^2(z_i, b)}{8} \right\}.$$

Clearly $\mathbb{P}(\Omega_2) \leq \mathbb{P}(\Omega_2') + \mathbb{P}(\Omega_2'')$ by a union bound argument.

Let us first upper bound $\mathbb{P}(\Omega_2')$. Recall that $n_k = n/K$ under the CSSBM by assumption, let us also define $n_{min} := \min_k n_k$. We keep this general notation because it is shorter and indicates how the proof can be generalized to the unbalanced setting. By definition $\tilde{\mu}_{z_i} - \mu_{z_i} = \sum_{j \in \mathcal{C}_{z_i}} \frac{\epsilon_j}{n_{z_i}}$, hence

$$\langle \epsilon_i, \tilde{\mu}_{z_i} - \mu_{z_i} \rangle = \frac{||\epsilon_i||^2 + \epsilon_i^\top \sum_{\substack{j \in \mathcal{C}_{z_i} \\ j \neq i}} \epsilon_j}{n_{z_i}}.$$

This last quantity is lower bounded by $\epsilon_i^\top \eta_i$ where $\eta_i = \frac{\sum_{j \in \mathcal{C}_{z_i}, j \neq i} \epsilon_j}{n_{z_i}}$. In particular ϵ_i and η_i are independent and their entries are also independent. Moreover η_i is a centered gaussian random variables with independent entries such that $\mathrm{Var}((\eta_i)_k) \leq 1/n_k$. So by Bernstein inequality, it holds for all $x > 0$ that

$$\mathbb{P}\left(||\eta_i||^2 \geq \frac{1}{n_k}(K + 2\sqrt{Kx} + 2x) \right) \leq \exp(-x)$$

which in turn implies

$$\begin{aligned}
\mathbb{P}\left(\epsilon_i^\top \eta_i \leq -\frac{\bar{\delta}\Delta_{min}^2}{8} \right) &\leq \mathbb{P}\left(\epsilon_i^\top \eta_i \leq -\frac{\bar{\delta}\Delta_{min}^2}{8} \,\middle|\, ||\eta_i||^2 \leq \frac{1}{n_k}(K + 2\sqrt{Kx} + 2x) \right) \\
&\quad + \mathbb{P}\left(||\eta_i||^2 \geq \frac{1}{n_k}(K + 2\sqrt{Kx} + 2x) \right) \\
&\leq \exp\left(-c \frac{n_k(\bar{\delta}\Delta_{min}^2)^2}{K + 2\sqrt{Kx} + 2x} \right) + \exp(-x).
\end{aligned}$$

Setting $x = \sqrt{n_k}\bar{\delta}\Delta_{min}^2$ we obtain $\mathbb{P}(\Omega_2') \leq 2\exp(-C\bar{\delta}\sqrt{n_k}\Delta_{min}^2)$. Since $\bar{\delta} \to 0$ (as $n \to \infty$) at a suitably slow rate, we have $\bar{\delta}\sqrt{n/K} \to +\infty$. Consequently,

$$\mathbb{P}(\Omega_2') = o(\exp(-(C + o(1))\Delta_{min}^2) = o(\exp(-(1 + o(1))\tilde{\Delta}^2)).$$

Let us now bound $\mathbb{P}(\Omega_2'')$. Since $\tilde{\Pi}_{kk'} - \Pi_{kk'} = \sum_{i \in \mathcal{C}_k, j \in \mathcal{C}_{k'}} \frac{E_{ij}}{n_k n_{k'}} = W_{:k}^\top E W_{:k'}$, we obtain the decomposition

$$\lambda \langle E_{i:} W, \tilde{\Pi}_{z_i:} - \Pi_{z_i:} \rangle = \lambda \langle E_{i:} W, W_{z_i:}^\top E^{(i)} W \rangle + \lambda \langle E_{i:} W, W_{z_i:}^\top E^{(-i)} W \rangle$$

where $E^{(i)}$ is obtained from E by only keeping the ith row and column, and E_{-i} is the matrix obtained from E by replacing the ith row and column by zero. In particular, $E^{(-i)}$ is independent from $E_{i:}$. The second term can be controlled by using the same techniques as before. Indeed, the entries of $W_{z_i:}^\top E^{(-i)} W$ are independent and $\mathrm{Var}(W_{z_i:}^\top E^{(-i)} W_{:k}) \leq C \frac{p_{max}}{n_{min}^2}$ for all k. Denoting $\eta'_{ik} = W_{z_i:}^\top E^{(-i)} W_{:k}$ and $\eta'_i = (\eta'_{ik})_{k \in [K]}$, this implies

$$\begin{aligned}
\mathbb{P}\left(\lambda \langle E_{i:} W, \eta'_i \rangle \leq -\frac{\bar{\delta}\Delta_{min}^2}{16} \right) &\leq \mathbb{P}\left(\lambda \langle E_{i:} W, \eta'_i \rangle \leq -\frac{\bar{\delta}\Delta_{min}^2}{16} \,\middle|\, \forall k,\ |\eta'_{ik}|^2 \leq C \frac{p_{max}}{n_k^2}(K + x) \right) \\
&\quad + \mathbb{P}\left(\exists k,\ |\eta'_{ik}|^2 \geq C \frac{p_{max}}{n_{min}^2}(K + x) \right) \\
&\leq K \exp\left(-C \frac{n_{min}^3 (\bar{\delta}\Delta_{min}^2)^2}{p_{max}\lambda^2 (K + x)} \right) + K \exp(-x).
\end{aligned}$$

Here, we used a union bound argument for the first inequality. The second inequality uses Lemma 4 – which provides a concentration bound for binomial random variables – along with the fact $\lambda \asymp \frac{n}{K p_{max}}$.

Setting $x = C\sqrt{n_k p_{max}}\bar{\delta}\Delta_{min}^2$ we obtain

$$\mathbb{P}\left(\lambda\langle E_{i:}W, \eta_i'\rangle \leq -\frac{\bar{\delta}\Delta_{min}^2}{16}\right) \leq 2\exp(-c\sqrt{n_k p_{max}}\bar{\delta}\Delta_{min}^2) = o(\exp(-\tilde{\Delta}^2))$$

since $\bar{\delta}$ can be chosen such that $\sqrt{n_k p_{max}}\bar{\delta} \to +\infty$ (because by assumption $np_{max} \asymp \log n \gg K$) and $\Delta_{min} \asymp \tilde{\Delta}$.

It remains to control $\langle E_{i:}W, W_{z_i:}^\top E^{(i)}W\rangle$. Using the fact

$$(W_{z_i:}^\top E^{(i)}W)_k = \begin{cases} \sum_{j'\in\mathcal{C}_k} \frac{E_{ij'}}{n_{z_i}n_k} & \text{if } k \neq z_i \\ 2\sum_{j'\in\mathcal{C}_k} \frac{E_{ij'}}{n_{z_i}^2} & \text{if } k = z_i \end{cases}$$

we have

$$\begin{aligned}\langle E_{i:}W, W_{z_i:}^\top E^{(i)}W\rangle &= \sum_{k\neq z_i}\sum_{\substack{j\in\mathcal{C}_k\\ j'\in\mathcal{C}_k}} \frac{E_{ij}}{n_k}\frac{E_{ij'}}{n_{z_i}n_k} + 2n_{z_i}\left(\sum_{j\in\mathcal{C}_{z_i}}\frac{E_{ij'}}{n_{z_i}^2}\right)^2 \\ &= \frac{1}{n_{z_i}}\left(\sum_{j\in\mathcal{C}_k}\frac{E_{ij}}{n_k}\right)^2 + 2n_{z_i}\left(\sum_{j\in\mathcal{C}_{z_i}}\frac{E_{ij'}}{n_{z_i}^2}\right)^2 \\ &\geq 0.\end{aligned}$$

Consequently, $\mathbb{P}(\Omega_2'')$ can be bounded as

$$\mathbb{P}(\Omega_2'') \leq \mathbb{P}\left(\lambda\langle E_{i:}W, \eta_i'\rangle \leq -\frac{\bar{\delta}\Delta_{min}^2}{16}\right) = o(\exp(-\tilde{\Delta}^2)).$$

Upper bound of $\mathbb{E}(M_3)$. Let us denote for any given $i \in [n]$ and $b \in [K] \setminus z_i$ the event

$$\Omega_3 = \left\{\langle -\epsilon_i, \tilde{\mu}_b - \mu_b\rangle - \lambda\langle E_{i:}W, \tilde{\Pi}_{b:} - \Pi_{b:}\rangle \leq \frac{-\bar{\delta}\Delta^2(z_i, b)}{4}\right\}.$$

First observe that

$$\langle \epsilon_i, \tilde{\mu}_b - \mu_b\rangle = \frac{\epsilon_i^\top \sum_{j\in\mathcal{C}_b}\epsilon_j}{n_{z_i}},$$

therefore this term can be handled in the same way as before. Moreover, we have

$$\lambda\langle E_{i:}W, \tilde{\Pi}_{b:} - \Pi_{b:}\rangle = \lambda\langle E_{i:}W, W_{b:}^\top E^{(i)}W\rangle + \lambda\langle E_{i:}W, W_{b:}^\top E^{(-i)}W\rangle.$$

The second term can be handled in the same way as before by using a conditioning argument. Now observe that

$$\langle E_{i:}W, W_{b:}^\top E^{(i)}W\rangle = \frac{1}{n_{z_i}^2 n_b}\left(\sum_{j\in\mathcal{C}_b} E_{ij}\right)\left(\sum_{j'\in\mathcal{C}_{z_i}} E_{ij'}\right)$$

where $\sum_{j\in\mathcal{C}_b} E_{ij}$ and $\sum_{j'\in\mathcal{C}_{z_i}} E_{ij'}$ are independent subgaussian random variables. Thus this term can also be controlled by using the same conditioning argument as before.

Conclusion. The previously obtained upper bounds imply

$$\mathbb{E}(\xi(\delta)) \leq 3\mathbb{E}(M_1) \leq n\exp(-(1+o(1))\tilde{\Delta}^2).$$

Finally, by Markov inequality, we obtain

$$\mathbb{P}(\xi(\delta) \geq \exp(\tilde{\Delta})\mathbb{E}\xi(\delta)) \leq \exp(-\tilde{\Delta}).$$

But since

$$\exp(\tilde{\Delta})\mathbb{E}\xi(\delta)) \leq n\exp(-(1+o(1))\tilde{\Delta}^2)$$

we obtain that with probability at least $1 - \exp(-\tilde{\Delta})$

$$\xi(\delta) \leq n\exp(-(1+o(1))\tilde{\Delta}^2).$$

B Proof of Theorem 2

The general proof strategy has been presented in Section 4.1. In Section B.1 we will make the error decomposition explicit. Then, we will control the different error terms in Sections B.2, B.3 and B.4. Finally, we will conclude by applying Theorem 1 in Section B.5.

B.1 Error decomposition for the one-step analysis of `IR-LSS`

We will assume without lost of generality that $\sigma = 1$ to simplify the exposition. Let $i \in [n]$ and $a \in [K]$ be such that $z_i = a$, and let

$$\lambda^{(t)} = \frac{n}{K(p^{(t)} - q^{(t)})} \log\left(\frac{p^{(t)}(1-q^{(t)})}{q^{(t)}(1-p^{(t)})}\right)$$

denote the scalar corresponding to the diagonal entry of the inverse covariance matrix $\Sigma_k^{(t)}$. Similarly, let us denote

$$\lambda = \frac{n}{K(p-q)} \log\left(\frac{p(1-q)}{q(1-p)}\right).$$

Given the current estimator of the partition $Z^{(t)}$, node i will be incorrectly estimated after one refinement step if

$$a \neq \arg\min_k ||X_i - \mu_k^{(t)}||^2 + \hat{\lambda}^{(t)}||A_{i:}W^{(t)} - \Pi_{k:}^{(t)}||^2$$

or equivalently, if there exists $b \in [K] \backslash a$ such that

$$||X_i - \mu_b^{(t)}||^2 + \hat{\lambda}^{(t)}||A_{i:}W^{(t)} - \Pi_{b:}^{(t)}||^2 \leq ||X_i - \mu_a^{(t)}||^2 + \hat{\lambda}^{(t)}||A_{i:}W^{(t)} - \Pi_{a:}^{(t)}||^2.$$

The above inequality is equivalent to

$$\langle \epsilon_i, \tilde{\mu}_a - \tilde{\mu}_b \rangle + \lambda \langle E_{i:}W, \tilde{\Pi}_{a:} - \tilde{\Pi}_{b:} \rangle \leq \frac{-\Delta^2(a,b)}{2} + F_{ib}^{(t)} + G_{ib}^{(t)} + H_{ib}^{(t)}$$

where

$$\Delta^2(a,b) = ||\mu_a - \mu_b||^2 + \lambda ||\Pi_{a:} - \Pi_{b:}||^2, \ \tilde{\mu}_k = X^\top W_{:a}, \text{ and } \tilde{\Pi}_{k:} = W_{k:}^\top AW$$

for all $k \in [K]$. Furthermore, the terms $F_{ib}^{(t)}, G_{ib}^{(t)}$ and $H_{ib}^{(t)}$ are given by

$$\begin{aligned}
F_{ib}^{(t)} &= \langle \epsilon_i, (\tilde{\mu}_a - \mu_a^{(t)}) - (\tilde{\mu}_b - \mu_b^{(t)}) \rangle + \lambda^{(t)} \langle E_{i:}W^{(t)}, (\tilde{\Pi}_{a:} - \Pi_{a:}^{(t)}) - (\tilde{\Pi}_{b:} - \Pi_{b:}^{(t)}) \rangle \\
&\quad + \lambda^{(t)} \langle E_{i:}(W - W^{(t)}), \tilde{\Pi}_{a:} - \tilde{\Pi}_{b:} \rangle + (\lambda - \lambda^{(t)}) \langle E_{i:}W, \tilde{\Pi}_{a:} - \tilde{\Pi}_{b:} \rangle, \\
2G_{ib}^{(t)} &= (||\mu_a - \mu_a^{(t)}||^2 - ||\mu_a - \tilde{\mu}_a||^2) - (||\mu_a - \mu_b^{(t)}||^2 - ||\mu_a - \tilde{\mu}_b||^2) \\
&\quad + \lambda^{(t)}(||P_{i:}W^{(t)} - \Pi_{a:}^{(t)}||^2 - ||P_{i:}W^{(t)} - W_{a:}^\top AW^{(t)}||^2) \\
&\quad - \lambda^{(t)}(||P_{i:}W^{(t)} - \Pi_{b:}^{(t)}||^2 - ||P_{i:}W^{(t)} - W_{b:}^\top AW^{(t)}||^2) \\
\text{and} \quad 2H_{ib}^{(t)} &= ||\mu_a - \tilde{\mu}_a||^2 - ||\mu_a - \tilde{\mu}_b||^2 + ||\mu_a - \mu_b||^2 \\
&\quad + \lambda^{(t)}(||P_{i:}W^{(t)} - W_{a:}^\top AW^{(t)}||^2 - ||P_{i:}W^{(t)} - W_{b:}^\top AW^{(t)}||^2 + ||\Pi_{a:} - \Pi_{b:}||^2) \\
&\quad + (\lambda - \lambda^{(t)})||\Pi_{a:} - \Pi_{b:}||^2.
\end{aligned}$$

The main term in this decomposition is

$$\langle \epsilon_i, \tilde{\mu}_a - \tilde{\mu}_b \rangle + \lambda \langle E_{i:}W, \tilde{\Pi}_{a:} - \tilde{\Pi}_{b:} \rangle \leq \frac{-\Delta^2(a,b)}{2}$$

and corresponds to the error when the current estimation of the partition is the ground truth partition. It is controlled by Lemma 1

The three error terms will be controlled in different ways. The error term $F_{ib}^{(t)}$ depends in a crucial way on i and t, it will be controlled with a l_2-type norm (see Condition B). The square of the error terms $G_{ib}^{(t)}$ and $|H_{ib}^{(t)}|$ will be controlled uniformly (see Condition C).

B.2 Error term $F_{ib}^{(t)}$

We need to upper-bound of

$$F = \max_{\{z^{(t)}: l(z,z^{(t)}) \leq \tau^{(0)}\}} \sum_{i=1}^{n} \max_{b \in [K] \setminus z_i} \frac{(F_{ib}^{(t)})^2}{\Delta(z_i,b)^2 l(z,z^{(t)})}.$$

To this end, we can decompose $F_{ib}^{(t)} = F_{ib}^{1,(t)} + F_{ib}^{2,(t)}$ where

$$F_{ib}^{1,(t)} = \langle \epsilon_i, (\tilde{\mu}_a - \mu_a^{(t)}) - (\tilde{\mu}_b - \mu_b^{(t)}) \rangle$$

is the error arising from the GMM part of the model and

$$F_{ib}^{2,(t)} = \lambda^{(t)} \langle E_{i:} W^{(t)}, (\tilde{\Pi}_{a:} - \Pi_{a:}^{(t)}) - (\tilde{\Pi}_{b:} - \Pi_{b:}^{(t)}) \rangle + \lambda^{(t)} \langle E_{i:}(W - W^{(t)}), \tilde{\Pi}_{a:} - \tilde{\Pi}_{b:} \rangle + (\lambda - \lambda^{(t)}) \langle E_{i:} W, \tilde{\Pi}_{a:} - \tilde{\Pi}_{b:} \rangle$$

is the error coming from the SBM part of the model. We have

$$F \leq 2 \max_{\{z^{(t)}: l(z,z^{(t)}) \leq \tau^{(0)}\}} \underbrace{\sum_{i=1}^{n} \max_{b \in [K] \setminus z_i} \frac{(F_{ib}^{1,(t)})^2}{\Delta(z_i,b)^2 l(z,z^{(t)})}}_{F_1^{(t)}} + 2 \max_{\{z^{(t)}: l(z,z^{(t)}) \leq \tau^{(0)}\}} \underbrace{\sum_{i=1}^{n} \max_{b \in [K] \setminus z_i} \frac{(F_{ib}^{2,(t)})^2}{\Delta(z_i,b)^2 l(z,z^{(t)})}}_{F_2^{(t)}}$$

and it is sufficient to individually control each term.

Control of F_1. We follow the same steps as in Gao and Zhang (2019), the only difference is that we use a different definition for Δ. To begin with,

$$\begin{aligned}
F_1^{(t)} &\leq \sum_{i=1}^{n} \sum_{b \in [K] \setminus z_i} \frac{\langle \epsilon_i, (\tilde{\mu}_{z_i} - \mu_{z_i}^{(t)}) - (\tilde{\mu}_b - \mu_b^{(t)}) \rangle^2}{\Delta(z_i,b)^2 l(z,z^{(t)})} \\
&\leq \sum_{b \in [K]} \sum_{a \in [K] \setminus b} \sum_{i \in \mathcal{C}_a} \frac{\langle \epsilon_i, (\tilde{\mu}_a - \mu_a^{(t)}) - (\tilde{\mu}_b - \mu_b^{(t)}) \rangle^2}{\Delta(a,b)^2 l(z,z^{(t)})} \\
&\leq \sum_{b \in [K]} \sum_{a \in [K] \setminus b} || \sum_{i \in \mathcal{C}_a} \epsilon_i \epsilon_i^\top || \frac{||(\tilde{\mu}_a - \mu_a^{(t)}) - (\tilde{\mu}_b - \mu_b^{(t)})||^2}{\Delta(a,b)^2 l(z,z^{(t)})}.
\end{aligned}$$

We first need to control $|| \sum_{i \in \mathcal{C}_a} \epsilon_i \epsilon_i^\top ||$ which can be done using the following lemma.

Lemma 2. *Let $\epsilon_i \overset{i.i.d}{\sim} \mathcal{N}(0, I_d)$. With probability at least $1 - \exp(-0.5n)$, we have*

$$|| \sum_{i \in [n]} \epsilon_i \epsilon_i^\top || \lesssim n + d.$$

Proof. See Lemma A.2 in Lu and Zhou (2016). ☐

Next, we need to control $||\tilde{\mu}_a - \mu_a^{(t)}||^2$ for all $a \in [K]$, this can be done with the following lemma.

Lemma 3. *Under the assumptions of Theorem 2, the following inequalities holds with probability at least $1 - n^{-\Omega(1)}$.*

1. $\max_{k \in [K]} ||\tilde{\mu}_k - \mu_k|| \lesssim \sqrt{\frac{K(d+\log n)}{n}}$,
2. $\max_{k \in [K]} ||\mathbb{E}(X)^\top (W_{:k}^{(t)} - W_{:k})|| \lesssim \frac{K}{n\Delta_{min}} l(z^{(t)}, z)$,
3. $\max_{k \in [K]} ||(X - \mathbb{E}(X))^\top W_{:k}^{(t)}|| \lesssim \frac{K\sqrt{(d+n) l(z^{(t)},z)}}{n\Delta_{min}} + \frac{K\sqrt{K(d+\log n)} l(z^{(t)},z)}{n\sqrt{n}\Delta_{min}^2}$.

Moreover, under the condition $l(z^{(t)}, z) \leq \tau^{(0)} \leq \frac{\epsilon n \Delta_{min}^2}{K}$, the previous inequalities imply

$$||\tilde{\mu}_k - \mu_k^{(t)}|| \leq C_3 \frac{K\sqrt{(d+n)l(z^{(t)}, z)}}{n\Delta_{min}}.$$

Proof. Straightforward adaptation the proof of Lemma 4.1 in Gao and Zhang (2019). □

By combining the different bounds, we can now conclude that with high probability,

$$\max_{\{z^{(t)}: l(z^{(t)}, z) \leq \tau^{(0)}\}} F_1^{(t)} \lesssim \frac{K^2(Kd/n+1)}{\Delta_{min}^2}\left(1 + \frac{K(d/n+1)}{\Delta_{min}^2}\right).$$

This quantity goes to zero when $\Delta_{min}^2/K^3 \to +\infty$.

Control of $F_2^{(t)}$. Here we can not directly apply the framework developed by Gao and Zhang (2019). Different changes are necessary and we need to deal with additional dependencies.

Let $b \in [K] \neq z_j$, we then have the bound

$$\begin{aligned}
(F_{ib}^{2,(t)})^2 &\leq 3(\lambda^{(t)}\langle E_{i:}W^{(t)}, (\tilde{\Pi}_{a:} - \Pi_{a:}^{(t)}) - (\tilde{\Pi}_{b:} - \Pi_{b:}^{(t)})\rangle)^2 \\
&+ 3(\lambda^{(t)}\langle E_{i:}(W - W^{(t)}), \tilde{\Pi}_{a:} - \tilde{\Pi}_{b:}\rangle)^2 \\
&+ 3(\lambda - \lambda^{(t)})^2\langle E_{i:}W, \tilde{\Pi}_{a:} - \tilde{\Pi}_{b:}\rangle^2 \\
&= F_{21}^2 + F_{22}^2 + F_{33}^2.
\end{aligned}$$

We drop the superscript (t) in the notation for the terms F_{21}, F_{22} and F_{23} for convenience, but clearly they depend on t as well. We will now bound each of the terms F_{2i} for $i = 1 \ldots 3$ separately. Starting with F_{21}, first note that

$$|\langle E_{i:}W^{(t)}, (\tilde{\Pi}_{a:} - \Pi_{a:}^{(t)}) - (\tilde{\Pi}_{b:} - \Pi_{b:}^{(t)})\rangle|^2 \leq 4||E_{i:}W^{(t)}||^2 \max_k ||\tilde{\Pi}_{k:} - \Pi_{k:}^{(t)}||^2.$$

The term $||\tilde{\Pi}_{k:} - \Pi_{k:}^{(t)}||^2$ can be bounded as

$$\begin{aligned}
||\tilde{\Pi}_{k:} - \Pi_{k:}^{(t)}||^2 &\leq 2||(W_{:k} - W_{:k}^{(t)})^\top AW||^2 + 2||W_{:k}^{(t)\top} A(W - W^{(t)})||^2 \\
&\lesssim \left(\frac{K^2\sqrt{p_{max}}l(z^{(t)}, z)}{n^{1.5}\Delta_{min}}\right)^2 && \text{(by Lemma 10)} \\
&\lesssim \frac{K^4 p_{max} l(z^{(t)}, z)^2}{n^3\Delta_{min}^2}.
\end{aligned}$$

Since $\lambda^{(t)} \lesssim \lambda$ by Lemma 12 we have

$$\begin{aligned}
\sum_{i=1}^n \max_{b\in[K]\setminus z_i} \frac{F_{21}^2}{\Delta^2(z_i, b)l(z, z^{(t)})} &\lesssim \lambda^2 \sum_i ||E_{i:}W^{(t)}||^2 \frac{K^4 p_{max} l(z^{(t)}, z)^2}{n^3\Delta_{min}^2} \\
&\lesssim \lambda^2 ||EW^{(t)}||_F^2 \frac{K^4 p_{max} l(z^{(t)}, z)}{n^3\Delta_{min}^4} \\
&\lesssim \lambda^2 K ||EW^{(t)}||^2 \frac{K^4 p_{max} l(z^{(t)}, z)}{n^3\Delta_{min}^4} \\
&\lesssim \lambda^2 K^2 p_{max} \frac{K^4 p_{max} l(z^{(t)}, z)}{n^3\Delta_{min}^4} \\
&\lesssim \frac{K^4 l(z^{(t)}, z)}{n\Delta_{min}^4} && \left(\text{ since } \lambda \lesssim \tfrac{n}{Kp_{max}}\right) \\
&\lesssim \frac{K^3}{\Delta_{min}^2} \to 0
\end{aligned}$$

where we used the fact $\frac{Kl(z^{(t)},z)}{n\Delta_{min}^2} \leq \epsilon$ for the last line.

We also have

$$F_{22}^2 \lesssim \lambda^2 ||E_{i:}(W - W^{(t)})||^2 ||\tilde{\Pi}_{z_i:} - \tilde{\Pi}_{b:}||^2 \lesssim \lambda^2 ||E_{i:}(W - W^{(t)})||^2 ||\Pi_{z_i:} - \Pi_{b:}||^2$$

hence

$$\begin{aligned}
\sum_{i=1}^{n} \max_{b \in [K]\setminus z_i} \frac{F_{22}^2}{\Delta(z_i,b)^2 l(z,z^{(t)})} &\lesssim \lambda \sum_i ||E_{i:}(W - W^{(t)})||^2 \frac{1}{l(z^{(t)},z)} \max_{b \in [K]\setminus z_i} \frac{\lambda ||\Pi_{z_i:} - \Pi_{b:}||^2}{\Delta(z_i,b)^2} \\
&\lesssim ||E(W - W^{(t)})||_F^2 \frac{\lambda}{l(z^{(t)},z)} \qquad \text{(because } \Delta(z_i,b)^2 \geq \lambda \Delta_2^2(z_i,b)\text{)} \\
&\lesssim K ||E(W - W^{(t)})||^2 \frac{\lambda}{l(z^{(t)},z)} \\
&\lesssim \lambda K n p_{max} \frac{K^3 l(z^{(t)},z)}{n^3 \Delta_{min}^4} \qquad \text{(by Lemma 10)} \\
&\lesssim \frac{K^3}{\Delta_{min}^2} \to 0.
\end{aligned}$$

Finally, using the bound

$$(\langle E_{i:}W, \tilde{\Pi}_{a:} - \tilde{\Pi}_{b:} \rangle)^2 \lesssim K p_{max} \Delta_2^2(a,b)$$

and since by Lemma 12

$$|\lambda^{(t)} - \lambda| \lesssim \lambda \frac{K^2 l(z^{(t)},z)}{\sqrt{np_{max}} n \Delta_{min}},$$

we obtain the bound

$$\begin{aligned}
\sum_{i=1}^{n} \max_{b \in [K]\setminus z_i} \frac{F_{23}^2}{\Delta(z_i,b)^2 l(z,z^{(t)})} &\lesssim \lambda^2 \left(\frac{K^2 l(z^{(t)},z)}{\sqrt{np_{max}} n \Delta_{min}} \right)^2 \sum_i ||E_{i:}W||^2 \max_{b \in [K]\setminus z_i} ||\tilde{\Pi}_{z_i:} - \tilde{\Pi}_{b:}||^2 \frac{1}{\Delta(z_i,b)^2 l(z^{(t)},z)} \\
&\lesssim \lambda \frac{K^4 l(z^{(t)},z)}{n^3 p_{max} \Delta_{min}^2} ||EW||_F^2 \max_{b \in [K]\setminus z_i} \frac{\lambda ||\Pi_{z_i:} - \Pi_{b:}||^2}{\Delta(z_i,b)^2} \\
&\lesssim \frac{n}{K p_{max}} \frac{K^4 l(z^{(t)},z)}{n^3 p_{max} \Delta_{min}^2} K p_{max} \\
&\lesssim \frac{K^4 l(z^{(t)},z)}{n^2 p_{max} \Delta_{min}^2} \\
&\lesssim \frac{K^3}{n p_{max}} \\
&\lesssim \frac{K^3}{\Delta_{min}^2} \to 0
\end{aligned}$$

under the assumption $l(z^{(t)},z) \leq \epsilon n \Delta_{min}^2 / K$.

Consequently, we have established that Condition B holds for all $\delta = o(1)$ such that $\delta^2 = \omega(K^3/\Delta_{min}^2)$.

B.3 Error term $G_{ib}^{(t)}$

As for $F_{ib}^{(t)}$ we can split $G_{ib}^{(t)}$ into

$$G_{ib}^{1,(t)} = ||\mu_a - \mu_a^{(t)}||^2 - ||\mu_a - \tilde{\mu}_a||^2) - (||\mu_a - \mu_b^{(t)}||^2 - ||\mu_a - \tilde{\mu}_b||^2)$$

and

$$G_{ib}^{2,(t)} = \lambda^{(t)}(||P_{i:}W^{(t)} - \Pi_{a:}^{(t)}||^2 - ||P_{i:}W^{(t)} - W_{a:}AW^{(t)}||^2) - \lambda^{(t)}(||P_{i:}W^{(t)} - \Pi_{b:}^{(t)}||^2 - ||P_{i:}W^{(t)} - W_{b:}AW^{(t)}||^2).$$

By the proof of Lemma 4.1 in Gao and Zhang (2019) (last inequality of page 46, equations (115) and (118)), we have

$$
\begin{aligned}
\frac{|G_{ib}^{1,(t)}|}{\Delta^2(a,b)} &\lesssim \left(\frac{Kl(z^{(t)},z)}{n\Delta_{min}} + K\sqrt{\frac{Kl(z^{(t)},z)}{n\Delta_{min}}}\right)^2 \Delta_{min}^{-2} \\
&+ \left(\frac{Kl(z^{(t)},z)}{n\Delta_{min}} + K\sqrt{\frac{Kl(z^{(t)},z)}{n\Delta_{min}}}\right) \frac{Kl(z^{(t)},z)}{n\Delta_{min}} \Delta_{min}^{-2} + \left(\frac{Kl(z^{(t)},z)}{n\Delta_{min}} + K\sqrt{\frac{Kl(z^{(t)},z)}{n\Delta_{min}}}\right) \Delta_{min}^{-1} \\
&\lesssim \frac{Kl(z^{(t)},z)^2}{n^2\Delta_{min}^4} + K\frac{l(z^{(t)},z)}{n\Delta_{min}^2} \\
&\lesssim K\frac{l(z^{(t)},z)}{n\Delta_{min}^2} \\
&\leq \max\left(\frac{\delta}{16}, Kc_1\frac{\tau^{(t)}}{8n\Delta_{min}^2}\right)
\end{aligned}
$$

for some universal constant c_1.

To bound $G_{ib}^{2,(t)}$ we will adapt the method developed in Han et al. (2020). We have by direct calculation

$$
\begin{aligned}
\frac{G_{ib}^{2,(t)}}{\lambda^{(t)}} &= (||\Pi_{a:}^{(t)} - W_{a:}^\top AW^{(t)}||^2 - ||\Pi_{b:}^{(t)} - W_{b:}AW^{(t)}||^2) + \langle P_{i:}W^{(t)} - W_{a:}AW^{(t)}, W_{:a}^\top AW^{(t)} - \Pi_{a:}^{(t)}\rangle \\
&- \langle P_{i:}W^{(t)} - W_{b:}AW^{(t)}, W_{:b}^\top AW^{(t)} - \Pi_{b:}^{(t)}\rangle \\
&\leq \left| ||\Pi_{a:}^{(t)} - W_{:a}^\top AW^{(t)}||^2 - ||\Pi_{b:}^{(t)} - W_{:b}^\top AW^{(t)}||^2 \right| + 4\max_{k\in[K]} \left|\langle W_{:a}^\top EW^{(t)}, (W_{:b} - W_{:b}^{(t)})^\top AW^{(t)}\rangle\right| \\
&+ 2\left|\langle (\Pi_{a:} - \Pi_{b:})Z^\top W^{(t)}, (W_{:b} - W_{:b}^{(t)})^\top AW^{(t)}\rangle\right| \\
&= G_{21} + G_{22} + G_{23}.
\end{aligned}
$$

We drop the superscript (t) in the notation for the terms G_{21}, G_{22} and G_{23} for convenience, but clearly they depend on t as well. First observe that

$$
G_{21} \leq \max_{k\in[K]} ||\Pi_{a:}^{(t)} - W_{:a}^\top AW^{(t)}||^2 = \max_{k\in[K]} ||(W_{a:}^{(t)} - W_{:a})^\top AW^{(t)}||^2 \lesssim \frac{K^3 p_{max} l(z^{(t)},z)^2}{n^3\Delta_{min}^2}
$$

where we used Lemma 10 for the last inequality. This implies by Lemma 11 that

$$
\max_{b\in[K]\setminus z_i} \frac{\lambda^{(t)}G_{21}}{\Delta(z_i,b)^2} \lesssim \frac{K^2 l(z^{(t)},z)^2}{n^2\Delta_{min}^4} \leq \max\left(\frac{\delta}{32}, Kc_2\frac{\tau^{(t)}}{16n\Delta_{min}^2}\right)
$$

for some constant $c_2 > 0$. Next, G_{22} can be bounded as

$$
\begin{aligned}
G_{22} &\leq 4\max_{a\in[K]} ||W_{:a}^\top EW^{(t)}|| \max_{a\in[K]} ||(W_{:b} - W_{:b}^{(t)})^\top AW^{(t)}|| \\
&\lesssim \frac{K\sqrt{p_{max}}}{\sqrt{n}} \frac{K^{1.5}\sqrt{p_{max}} l(z^{(t)},z)}{n^{1.5}\Delta_{min}} && \text{(by Lemma 10)} \\
&\lesssim K^{2.5}\frac{p_{max}}{n}\frac{l(z^{(t)},z)}{n\Delta_{min}}
\end{aligned}
$$

which in turn implies

$$
\begin{aligned}
\max_{b\in[K]\setminus z_i} \frac{\lambda^{(t)}G_{22}}{\Delta(z_i,b)^2} &\lesssim \frac{\sqrt{K}}{\Delta_{min}}\frac{Kl(z^{(t)},z)}{n\Delta_{min}^2} && \text{(because } \lambda^{(t)} \lesssim n/(Kp_{max})\text{)} \\
&= o(\delta)
\end{aligned}
$$

since $\sqrt{K}/\Delta_{min} \to 0$ and $K\frac{l(z^{(t)},z)}{n\Delta_{min}^2} \leq 1$.

Finally, using the bound

$$\begin{aligned} G_{23} &\lesssim ||\Pi_{a:} - \Pi_{b:})Z^\top W^{(t)}|| \max_{b\in[K]} ||(W_{:b} - W_{:b}^{(t)})^\top AW^{(t)}|| \\ &\lesssim \Delta_2(a,b)\frac{K^{1.5}\sqrt{p_{max}}l(z^{(t)},z)}{n^{1.5}\Delta_{min}} \end{aligned}$$

we obtain

$$\begin{aligned} \max_{b\in[K]\setminus z_i} \frac{\lambda^{(t)}G_{23}}{\Delta(z_i,b)^2} &\lesssim \lambda\frac{\Delta_2(z_i,b)}{\Delta(z_i,b)}\frac{K^{1.5}\sqrt{p_{max}}l(z^{(t)},z)}{n^{1.5}\Delta_{min}^2} \\ &\lesssim \sqrt{\lambda}\frac{\sqrt{Kp_{max}}}{\sqrt{n}}\frac{Kl(z^{(t)},z)}{n\Delta_{min}^2} \\ &\leq \max\left(\frac{\delta}{32}, c_3K\frac{\tau^{(t)}}{16n\Delta_{min}^2}\right) \end{aligned}$$

for some constant $c_3 > 0$.

B.4 Error term $H_{ib}^{(t)}$

As before, we can split $H_{ib}^{(t)} = H_{ib}^{1,(t)} + H_{ib}^{2,(t)}$ where

$$2H_{ib}^{1,(t)} = ||\mu_a - \tilde{\mu}_a||^2 - ||\mu_a - \tilde{\mu}_b||^2 + ||\mu_a - \mu_b||^2$$

and

$$2H_{ib}^{2,(t)} = \lambda^{(t)}(||P_{i:}W^{(t)} - W_{a:}^\top AW^{(t)}||^2 - ||P_{i:}W^{(t)} - W_{b:}^\top AW^{(t)}||^2 + ||\Pi_{a:} - \Pi_{b:}||^2) + (\lambda - \lambda^{(t)})||\Pi_{a:} - \Pi_{b:}||^2.$$

We obtain by an immediate adaptation of Lemma 4.1 in Gao and Zhang (2019)

$$\frac{|H_{ib}^{1,(t)}|}{\Delta(z_i,b)^2} \lesssim \frac{K(d+\log n)}{n\Delta_{min}^2} + \sqrt{\frac{K(d+\log n)}{n\Delta_{min}^2}} \to 0$$

as long as $K/\Delta_{min}^2 \to 0$.

It remains to control uniformly $H_{ib}^{2,(t)}$, let us split it[1] as $H_{ib}^{2,(t)} = \lambda^{(t)}H_1 + H_2$. First note that by Lemma 12

$$H_2 := (\lambda - \lambda^{(t)})||\Pi_{a:} - \Pi_{b:}||^2 \lesssim \lambda\frac{Kl(z^{(t)},z)}{n\Delta_{min}^2}\Delta_2^2(a,b)$$

so we have

$$\frac{|H_2|}{\Delta(z_i,b)^2} \lesssim \frac{Kl(z^{(t)},z)}{n\Delta_{min}^2} \leq \max\left(\frac{\delta}{32}, Kc_4\frac{\tau^{(t)}}{16n\Delta_{min}^2}\right)$$

for some constant $c_4 > 0$.

Now observe that

$$\begin{aligned} H_1 :&= ||P_{i:}W^{(t)} - W_{a:}^\top AW^{(t)}||^2 - ||P_{i:}W^{(t)} - W_{b:}^\top AW^{(t)}||^2 + ||\Pi_{a:} - \Pi_{b:}||^2 \\ &= ||W_{a:}^\top EW^{(t)}||^2 + (||\Pi_{a:} - \Pi_{b:}||^2 - ||P_{i:}W^{(t)} - W_{b:}^\top PW^{(t)}||^2) \\ &- (||P_{i:}W^{(t)} - W_{b:}^\top AW^{(t)}||^2 - ||P_{i:}W^{(t)} - W_{b:}^\top PW^{(t)}||^2) \\ &= (||\Pi_{a:} - \Pi_{b:}||^2 - ||P_{i:}W^{(t)} - W_{b:}^\top PW^{(t)}||^2) + (||W_{a:}^\top EW^{(t)}||^2 - ||W_{b:}^\top EW^{(t)}||^2) \\ &+ 2\langle P_{i:}W^{(t)} - W_{b:}^\top PW^{(t)}, W_{:b}^\top EW^{(t)}\rangle \\ &= H_{11}^{(t)} + H_{12}^{(t)} + H_{13}^{(t)}. \end{aligned}$$

[1]We drop the superscript (t) in the notation for H_1, H_2 for convenience, but clearly they both depend on t as well.

By writing

$$P_{i:}W^{(t)} - W_{b:}^{\top} P W^{(t)} = (\Pi_{a:} - \Pi_{b:}) Z^{\top} W^{(t)},$$

we can bound $|H_{13}^{(t)}|$ as

$$|H_{13}^{(t)}| \lesssim ||P_{i:}W^{(t)} - W_{b:}^{\top} P W^{(t)}|| ||W_{:b}^{\top} E W^{(t)}|| \lesssim ||\Pi_{a:} - \Pi_{b:}|| \frac{K\sqrt{p_{max}}}{\sqrt{n}} \lesssim ||\Pi_{a:} - \Pi_{b:}||^2 \frac{K}{\sqrt{np_{max}}}.$$

In particular,

$$\frac{\lambda^{(t)} |H_{13}^{(t)}|}{\Delta^2(z_i, b)} \lesssim \frac{K}{\sqrt{np_{max}}} \to 0.$$

Next observe that

$$\left| ||W_{a:}^{\top} E W^{(t)}||^2 - ||W_{b:}^{\top} E W^{(t)}||^2 \right| \leq \max_{k \in [K]} ||W_{k:}^{\top} E W^{(t)}||^2 \lesssim \frac{K^2 p_{max}}{n}$$

where the last inequality uses Lemma 10. This implies

$$\frac{\lambda^{(t)} |H_{12}^{(t)}|}{\Delta^2(z_i, b)} \lesssim \frac{K}{\Delta_{min}^2} \to 0.$$

Finally, it remains to bound $|H_{11}^{(t)}|$. To begin with,

$$|H_{11}^{(t)}| := \left| ||\Pi_{a:} - \Pi_{b:}||^2 - ||P_{i:}W^{(t)} - W_{b:}^{\top} P W^{(t)}||^2 \right| = \left| ||\Pi_{a:} - \Pi_{b:}||^2 - ||(\Pi_{a:} - \Pi_{b:}) Z^{\top} W^{(t)}||^2 \right|.$$

Using the fact

$$\left| ||\Pi_{a:} - \Pi_{b:}||^2 - ||(\Pi_{a:} - \Pi_{b:}) Z^{\top} W^{(t)}||^2 \right| \leq \left(\left\| Z^{\top} W^{(t)} - I \right\|^2 + 2 \left\| Z^{\top} W^{(t)} - I \right\| \right) \|\Pi_{a:} - \Pi_{b:}\|^2,$$

we obtain by the proof of part 1 of Lemma 10

$$|H_{11}^{(t)}| \lesssim ||\Pi_{a:} - \Pi_{b:}||^2 \frac{K l(z^{(t)}, z)}{n \Delta_{min}^2}.$$

This implies that

$$\frac{\lambda^{(t)} |H_{11}^{(t)}|}{\Delta^2(z_i, b)} \lesssim \frac{K l(z^{(t)}, z)}{n \Delta_{min}^2} \leq \max\left(\frac{\delta}{32}, \frac{K c_5 \tau^{(t)}}{8 n \Delta_{min}^2} \right)$$

for some constant c_5.

By summing all these inequalities we see that $G_{ib}^{(t)}$ and $H_{ib}^{(t)}$ satisfy Condition C for some constant $c > 0$.

B.5 Conclusion

In order to apply Theorem 1, we need to show that Condition A is satisfied. In particular, we need to control

$$\Omega(\delta^{(t)}) = \{C_i(a, b) \leq \frac{1 - \delta^{(t)}}{2} \Delta^2(a, b)\}.$$

But this is an immediate adaptation of Lemma 6: we just need to replace $\Delta^2/4$ by $(1 - \delta^{(t)})\Delta^2/4$ in the last step of the lemma to obtain

$$\mathbb{P}(\Omega(\delta^{(t)})) \leq \exp(-(1 - \delta^{(t)} + o(1))\tilde{\Delta}).$$

Then we can easily adapt Lemma 1 to obtain that

$$\xi(\delta^{(t)}) \leq n \exp(-(1 - \delta^{(t)} + o(1))\tilde{\Delta}).$$

So $\xi(\delta^{(t)}) \le \tau \le \tau^{(t-1)}$ as long as

$$\tau \ge n\exp(-(1-\delta^{(0)}+o(1))\tilde{\Delta}).$$

Hence we can choose

$$\tau = \max\left(n\exp(-(1-\delta^{(0)}+o(1))\tilde{\Delta}), 1\right).$$

Since $\tau^{(0)} \le n\Delta_{min}^2/K \lesssim n\log n$, after $O(\log n)$ iterations $\tau^{(t)} = \tau$ due to the geometric decay of $(\tau^{(t)})$. It implies that $\delta^{(t)}$ also reaches δ after $O(\log n)$ iterations since by definition of τ, $\frac{2cK\tau}{n\Delta_{min}^2} = o(\delta)$. Consequently, we get for all $t \gtrsim \log n$ by Theorem 1 and Lemma 1

$$\begin{aligned} l(z^{(t)}, z) &\le \xi(\delta) + \frac{1}{8} l(z^{(t-1)}, z) \\ &\le n\exp(-(1+o(1))\tilde{\Delta}) + \frac{1}{8} l(z^{(t-1)}, z). \end{aligned}$$

Since $l(z^{(t-1)}, z) \le n\Delta_{min}^2/K = O(n^2)$, we get after $O(\log n)$ additional iterations that

$$l(z^{(t)}, z) \le \frac{8}{7} n\exp(-(1+o(1))\tilde{\Delta}) + o(1) = n\exp(-(1+o(1))\tilde{\Delta}).$$

C Proof of Theorem 3

Proof. The proof follows the same lines as in (Lu and Zhou, 2016, Theorem 3.3), only the last part needs to be changed. For the sake of completeness, we reproduce the arguments below. Let us denote

$$h(z', z'') = \sum_{i\in[n]} \mathbf{1}_{\{z'(i)\neq z''(i)\}}$$

to be the unnormalized Hamming distance between $z', z'' \in [K]^n$. Without lost of generality we can assume that

$$\min_{k,k'} ||\mu_k - \mu_{k'}|| = ||\mu_1 - \mu_2||.$$

For each $k \in [K]$, let T_k a subset of $\mathcal{C}_k$ with cardinality $\frac{3n}{4K}$. Define $T = \cup_{k=1}^K T_k$ and

$$\mathcal{Z} = \{\hat{z} : \hat{z}_i = z_i \text{ for all } i \in T\}.$$

For all $\hat{z} \neq \tilde{z} \in \mathcal{Z}$ we have

$$\frac{h(\hat{z}, \tilde{z})}{n} \le \frac{1}{4}$$

and for all permutations $\sigma \in \mathfrak{S}_K, \sigma \neq Id$ (where Id denotes the identity permutation) we have

$$\frac{h(\sigma(\hat{z}), \tilde{z})}{n} \ge \frac{1}{2}.$$

Thus $r(\hat{z}, \tilde{z}) = \frac{h(\hat{z},\tilde{z})}{n}$. Then following the same arguments as in the proof of Theorem 2 in Gao et al. (2016) we can obtain

$$\inf_{\hat{z}} \sup_{\theta\in\Theta} \mathbb{E}(r(\hat{z}, z)) \ge \frac{1}{6|T^c|} \sum_{i\in T^c} \frac{1}{2K^2} \inf_{\hat{z}_i} (\mathbb{P}_1(\hat{z}_i = 2) + \mathbb{P}_2(\hat{z}_i = 1)) \tag{C.1}$$

where $\mathbb{P}_k$ denotes the probability distribution of the data when $z_i = k$. By the Neyman Pearson Lemma, the infinimum of the right hand side of (C.1) is achieved by the likelihood ratio test. From (Zhang and Zhou, 2016, Section 3.1), the log-likelihood of the SBM part can be rewritten as

$$\log\left(\frac{p(1-q)}{q(1-p)}\right) \sum_{i<j} A_{ij} \mathbf{1}_{\{z_i=z_j\}} + f(A)$$

where $f(A)$ doesn't depend on z. Consequently,

$$\begin{aligned}
&\frac{1}{2}\inf_{\hat{z}_i}\left(\mathbb{P}_1(\hat{z}_i=2)+\mathbb{P}_2(\hat{z}_i=1)\right)\\
&=\mathbb{P}\left(-0.5||\epsilon_i||^2+\log\left(\frac{p(1-q)}{q(1-p)}\right)\sum_{j\in\mathcal{C}_1}A_{ij}\le -0.5||\mu_1+\epsilon_i-\mu_2||^2+\log\left(\frac{p(1-q)}{q(1-p)}\right)\sum_{j\in\mathcal{C}_2}A_{ij}\right) \qquad \text{(C.2)}
\end{aligned}$$

Let us denote $Z_i=\log(\frac{p(1-q)}{q(1-p)})(\sum_{j\in\mathcal{C}_2}A_{ij}-\sum_{j\in\mathcal{C}_1}A_{ij})$, this is a random variable independent of ϵ_i. So we get

$$\begin{aligned}
(C.2)&=\mathbb{P}(0.5||\mu_1-\mu_2||^2-Z_i\le\langle\epsilon_i,\mu_1-\mu_2\rangle)\\
&\ge\mathbb{P}(0.5||\mu_1-\mu_2||^2-Z_i\le\langle\epsilon_i,\mu_1-\mu_2\rangle\mid Z_i>0)\mathbb{P}(Z_i>0)\\
&\ge\mathbb{P}(||\mu_1-\mu_2||^2\le 2\langle\epsilon_i,\mu_1-\mu_2\rangle)\mathbb{P}(Z>0)\\
&\ge\exp(-\frac{\Delta_1^2}{8})\exp(-n\frac{(1+o(1))(\sqrt{p}-\sqrt{q})^2}{K})\\
&\ge\exp(-(1+o(1))\tilde{\Delta}^2).
\end{aligned}$$

Here we used for the penultimate inequality a result from the proof of (Lu and Zhou, 2016, Theorem 3.3) and also use (Zhang and Zhou, 2016, Lemma 5.2). □

D Technical Lemmas

D.1 General concentration inequalities

Lemma 4 (Chernoff simplified bound)**.** *Let $X_1,\dots,X_n\overset{ind.}{\sim}\mathcal{B}(p)$. Then*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i\in[n]}X_i-p\right|\ge t\right)\le\exp(-2nt^2).$$

Lemma 5. *Assume that $A\sim SBM(Z,\Pi)$. Let $E=A-\mathbb{E}(A)$. Then with probability at least $1-n^{-\Omega(1)}$ the following holds.*

1. $||E||\le\sqrt{np_{max}}$,
2. $||EW||_F^2\lesssim K^2p_{max}$.

Proof. The first inequality is a classical result used for SBM in the relatively sparse regime $p_{max}=\omega(\log n)$. It can obtained as a consequence of Remark 3.13 in Bandeira and van Handel (2016). The second inequality follows from

$$\begin{aligned}
||EW||_F^2&\le K||EW||^2\\
&\le K||E||^2||W||^2\\
&\lesssim K^2p_{max}.
\end{aligned}$$

□

D.2 Concentration rate for the ideal oracle error under CSSBM

Lemma 6. *We have*

$$\mathbb{P}(\Omega_1)\le\exp(-(1+o(1))\tilde{\Delta}^2)$$

where

$$\tilde{\Delta}^2=\frac{1}{8}\min_{k,k'}||\mu_k-\mu_{k'}||^2+\frac{\log n}{K}(\sqrt{p'}-\sqrt{q'})^2.$$

Proof. We are going to bound the m.g.f of $Z = \langle \epsilon_i, \mu_{z_i} - \mu_b \rangle + \lambda \langle E_{i:}W, \Pi_{z_i:} - \Pi_{b:} \rangle$ and use Chernoff method. We have for all $t < 0$

$$\begin{aligned}
\log \mathbb{E}e^{tZ} &\leq \log \mathbb{E}e^{t\langle \epsilon_i, \mu_{z_i} - \mu_b \rangle} + \log \mathbb{E}e^{t\lambda \langle E_{i:}W, \Pi_{z_i:} - \Pi_{b:} \rangle} && \text{(by independence)} \\
&\leq ||\mu_{z_i} - \mu_b||^2 \frac{t^2}{2} + \frac{n}{K} \log(pe^{t\lambda(p-q)K/n} + 1 - p)(qe^{-t\lambda(p-q)K/n} + 1 - q) \\
&- t\lambda(p-q)^2.
\end{aligned}$$

By choosing $t = -1/2$ we get

$$e^{t\lambda(p-q)K/n} = \sqrt{\frac{q(1-p)}{p(1-q)}}$$

and thus

$$\log(pe^{t\lambda(p-q)K/n} + 1 - p)(qe^{-t\lambda(p-q)K/n} + 1 - q) = \log(pq + (1-p)(1-q) + 2\sqrt{pq}\sqrt{(1-p)(1-q)}).$$

This last quantity is equal to $-(1 + o(1))(\sqrt{p} - \sqrt{q})^2$. We can now conclude by remarking that

$$\mathbb{P}(\Omega_1) = \mathbb{P}\left(-\frac{1}{2}\langle \epsilon_i, \mu_{z_i} - \mu_b \rangle - \frac{1}{2}\lambda \langle E_{i:}W, \Pi_{z_i:} - \Pi_{b:} \rangle \geq \frac{\Delta_{min}^2}{4} \right)$$

hence

$$\begin{aligned}
\mathbb{P}(\Omega_1) &\leq \mathbb{E}e^{-\frac{Z}{2} - \frac{\Delta_{min}^2}{4}} \\
&\leq \exp(\frac{||\mu_{z_i} - \mu_b||^2}{8} - \frac{n}{K}(1 + o(1))(\sqrt{p} - \sqrt{q})^2 + \lambda(p-q)^2 - \frac{\Delta^2}{4}) \\
&\leq \exp(-(1 + o(1))\tilde{\Delta}^2),
\end{aligned}$$

since $\Delta_{min}^2 = ||\mu_{z_i} - \mu_b||^2 + 2\lambda(p-q)^2$. □

D.3 Concentration rate for the ideal oracle error under the general setting

In the general setting it is more difficult to derive a sharp concentration inequality for the oracle error. Here we use gaussian approximation, but it leads to a slightly sub-optimal convergence rate.

Lemma 7. *We have*

$$\mathbb{P}(\Omega_1) \leq \exp\left(-\frac{1}{8}\Delta_{min}^2 \right)$$

for some constant $c > 0$.

Proof. First observe that

$$\begin{aligned}
t\lambda \langle E_{i:}W, \Pi_{z_i:} - \Pi_{b:} \rangle &= t\lambda \sum_{k \in [K]} (\Pi_{z_i k} - \Pi_{bk}) \frac{\sum_{j \in \mathcal{C}_k} E_{ij}}{n_k} \\
&= \sum_{k \in [K]} t\lambda (\Pi_{z_i k} - \Pi_{bk}) \frac{\sum_{j \in \mathcal{C}_k} A_{ij} - \Pi_{z_i k}}{n_k}.
\end{aligned}$$

The sum over k involves independent random variables so in order to bound the $m.g.f.$ of $\langle E_{i:}W, \Pi_{z_i:} - \Pi_{b:} \rangle$ it is sufficient to control the $m.g.f.$ of $\sum_{j \in \mathcal{C}_k} A_{ij} - \Pi_{z_i k}$ for each k. Setting $t' = \lambda t \frac{|\Pi_{z_i k} - \Pi_{bk}|}{n_k}$, we have

$$\begin{aligned}
\log \mathbb{E}(e^{t' \sum_{j \in \mathcal{C}_k} (A_{ij} - \Pi_{z_i k})}) &= \log(\Pi_{z_i k} e^{t'} + 1 - \Pi_{z_i k}) - n_k t' \Pi_{z_i k} \\
&\leq n_k \Pi_{z_i k}(e^{t'} - t' - 1) \\
&\leq n_k \Pi_{z_i k} \frac{e(t')^2}{2} && \text{(by Taylor-Lagrange inequality)} \\
&\leq 1.5 n_k p_{max} (\lambda t \frac{\Pi_{z_i k} - \Pi_{bk}}{n_k})^2 \\
&\leq 1.5\lambda |\Pi_{z_i k} - \Pi_{bk}|^2 t^2.
\end{aligned}$$

For the second inequality we used the fact that for $0 < x < 1$, $\log(1-x) \leq -x$.

Consequently,

$$\log \mathbb{E} e^{tZ} \leq ||\mu_{z_i} - \mu_b||^2 \frac{t^2}{2} + 1.5\lambda ||\Pi_{z_i:} - \Pi_{b:}||^2 t^2$$

and

$$\mathbb{P}(\Omega_1) \leq e^{||\mu_{z_i} - \mu_b||^2 \frac{t^2}{2} + 1.5\lambda ||\Pi_{z_i:} - \Pi_{b:}||^2 t^2 - \frac{\Delta^2(z_i, b)}{4}}.$$

For $t = 1/2$ we then obtain

$$\mathbb{P}(\Omega_1) \leq e^{-\frac{||\mu_{z_i} - \mu_b||^2}{8} - \frac{\lambda}{8} ||\Pi_{z_i:} - \Pi_{b:}||^2} \leq e^{-\frac{1}{8}\Delta^2_{min}}.$$

□

D.4 Useful inequalities to control the error terms

Lemma 8. *For all $z, z' \in [K]^n$ we have*

$$h(z, z') \leq \frac{l(z, z')}{\Delta^2_{min}}.$$

Proof.

$$\sum_{i \in [n]} \mathbf{1}_{z_i \neq z'_i} \leq \sum_{i \in [n]} \frac{\Delta(z_i, z'_i)^2}{\Delta^2_{min}} \mathbf{1}_{z_i \neq z'_i} = \frac{l(z, z')}{\Delta^2_{min}}.$$

□

Lemma 9. *Assume that for some $\alpha > 1$*

$$\frac{n}{\alpha K} \leq n_k \leq \frac{\alpha n}{K}.$$

If $l(z, z^{(t)}) \leq n\Delta^2_{min}/(2\alpha K)$ then for all $k \in [K]$

$$\frac{n}{2\alpha K} \leq n_k^{(t)} \leq \frac{2\alpha n}{K}.$$

Proof. Since for all $k \in [K]$ we have $n/(\alpha K) \leq n_k \leq \alpha n / K$,

$$\begin{aligned}
\sum_{i \in \mathcal{C}_k^{(t)}} 1 \geq \sum_{i \in \mathcal{C}_k \cap \mathcal{C}_k^{(t)}} 1 &\geq \sum_{i \in \mathcal{C}_k} 1 - \sum_{i \in [n]} \mathbf{1}_{z_i \neq z_i^{(t)}} \\
&\geq \frac{n}{\alpha K} - h(z, z^{(t)}) \overset{\text{Lemma 8}}{\geq} \alpha \frac{n}{K} - \frac{l(z, z^{(t)})}{\Delta^2_{min}} \\
&\geq \frac{\alpha n}{2K}
\end{aligned}$$

by assumption. The other inequality is proved in a similar way. □

Lemma 10. *Assume that $A \sim SBM(Z, \Pi)$ with equal size communities. Under the assumption that $\frac{l(z^{(t)}, z)}{n\Delta^2_{min}} \leq \epsilon / K$, with probability at least $1 - n^{-\Omega(1)}$ the following holds.*

1. $\max_{k \in [K]} ||W^{(t)}_{:k} - W_{:k}|| \lesssim \frac{K^{1.5}}{n^{1.5}\Delta^2_{min}} l(z^{(t)}, z)$,
2. $\max_{k \in [K]} ||(W^{(t)}_{:k} - W_{:k})^\top A W|| \lesssim \frac{K^{1.5}\sqrt{p_{max}}}{n^{1.5}\Delta_{min}} l(z, z^{(t)})$,
3. $\max_{k \in [K]} ||W^{(t)\top}_{:k} A (W - W^{(t)})|| \lesssim \frac{K^2 \sqrt{p_{max}} l(z^{(t)}, z)}{n^{1.5}\Delta_{min}}$,
4. $\max_{k \in [K]} ||(W^{(t)}_{:k} - W_{:k})^\top A W^{(t)}|| \lesssim \frac{K^{1.5}\sqrt{p_{max}} l(z^{(t)}, z)}{n^{1.5}\Delta_{min}}$,
5. $||Z^\top W^{(t)}|| \lesssim 1$.

Proof. This is a rather straightforward adaptation of Lemma 4 in Han et al. (2020), but for completeness we include a proof adapted to our setting with our notations.

Proof of 1. First observe that Z is rank K and $\lambda_K(Z) = \sqrt{n_{min}}$ so that

$$||W^{(t)}_{:k} - W_{:k}|| \leq n_{min}^{-1/2} ||I - Z^\top W^{(t)}||.$$

For any $k \in [K]$, denote $\delta_k = 1 - (Z^\top W^{(t)})_{kk}$. Since for all $k, k' \in [K]$

$$(Z^\top W^{(t)})_{kk'} = \frac{\sum_{i \in \mathcal{C}_k} \mathbf{1}_{z_i^{(t)} = k'}}{n^{(t)}_{k'}},$$

we have

$$0 \leq \delta_k \leq 1, \qquad \sum_{k' \in [K] \setminus k} (Z^\top W^{(t)})_{k'k} = \delta_k.$$

Therefore,

$$\begin{aligned}
||Z^\top W^{(t)} - I|| &= \sqrt{\sum_{k \in [K]} \left(\delta_k^2 + \sum_{k' \in [K] \setminus k} (Z^\top W^{(t)})^2_{k'k} \right)} \\
&\leq \sqrt{\sum_{k \in [K]} \left(\delta_k^2 + \left(\sum_{k' \in [K] \setminus k} (Z^\top W^{(t)})_{k'k} \right)^2 \right)} \\
&\leq \sqrt{2 \sum_{k \in [K]} \delta_k^2} \leq \sqrt{2} \sum_{k \in [K]} \delta_k \\
&= \sqrt{2} \sum_{k \in [K]} \frac{\sum_{i \in \mathcal{C}^{(t)}_k} \mathbf{1}_{z_i \neq k}}{n^{(t)}_k} \\
&\leq \sqrt{2} \max_k (n^{(t)}_k)^{-1} \sum_{i \in [n]} \mathbf{1}_{z_i \neq z_i^{(t)}} \\
&\overset{\text{Lemma 9}}{\lesssim} \frac{K}{n} h(z, z^{(t)}) \overset{\text{Lemma 8}}{\lesssim} K \frac{l(z, z^{(t)})}{n \Delta^2_{min}}.
\end{aligned} \tag{D.1}$$

Proof of 2. Observe that with probability at least $1 - n^{-\Omega(1)}$ we have

$$\begin{aligned}
\max_{k \in [K]} ||(W^{(t)}_{:k} - W_{:k})^\top A W|| &\leq \max_{k \in [K]} ||(W^{(t)}_{:k} - W_{:k})^\top P W|| + \max_{k \in [K]} ||(W^{(t)}_{:k} - W_{:k})^\top E W|| \\
&\leq \max_{k \in [K]} ||(W^{(t)}_{:k} - W_{:k})^\top Z \Pi|| + ||EW|| \max_{k \in [K]} ||(W^{(t)}_{:k} - W_{:k})|| \\
&\leq ||\Pi_{b:} - \sum_{j \in \mathcal{C}^{(t)}_b} \frac{\Pi_{z_j:}}{n^{(t)}_b}|| + C \sqrt{K p_{max}} \max_{k \in [K]} ||(W^{(t)}_{:k} - W_{:k})|| \\
&\lesssim ||\Pi_{b:} - \sum_{j \in \mathcal{C}^{(t)}_b} \frac{\Pi_{z_j:}}{n^{(t)}_b}|| + \frac{K^2 \sqrt{p_{max}}}{n^{1.5} \Delta^2_{min}} l(z^{(t)}, z).
\end{aligned}$$

Moreover,

$$
\begin{aligned}
||\Pi_{b:} - \sum_{j\in\mathcal{C}_b^{(t)}} \frac{\Pi_{z_j:}}{n_b^{(t)}}|| &= ||\sum_{\substack{j\in\mathcal{C}_b^{(t)}\\ b'\in[K]\setminus b}} \frac{\mathbf{1}_{\{z_j=b'\}}}{n_b^{(t)}}(\Pi_{b:}-\Pi_{b':})|| \\
&\le C\frac{K}{n}\sum_{\substack{j\in\mathcal{C}_b^{(t)}\\ b'\in[K]\setminus b}} \max_{b,b'} \Delta_2(b,b')\mathbf{1}_{\{z_j=b'\}} \\
&\le C\frac{K}{n}\max_{b,b'}\Delta_2(b,b')h(t,t^{(t)}) \\
&\le C\frac{K\Delta_{min}}{\sqrt{\lambda}n\Delta_{min}^2} l(z,z^{(t)}) \qquad \text{(since } \max_{b,b'}\Delta_2(b,b') \lesssim \tfrac{\Delta_{min}}{\sqrt{\lambda}} \text{ for SBM)} \\
&\le C\frac{K^{1.5}\sqrt{p_{max}}}{n^{1.5}\Delta_{min}} l(z,z^{(t)}).
\end{aligned}
$$

Consequently, by summing the previous bounds and using the first inequality of the Lemma we get

$$
\max_{k\in[K]} ||(W_{:k}^{(t)} - W_{:k})^\top AW|| \lesssim \frac{K^{1.5}\sqrt{p_{max}}}{n^{1.5}\Delta_{min}} l(z,z^{(t)}) + \frac{K^2\sqrt{p_{max}}}{n^{1.5}\Delta_{min}^2} l(z^{(t)},z).
$$

In our setting $\Delta_{min}^2 \asymp \log n$ so the first term is dominant.

Proof of 3. First let's bound $\max_{k\in[K]} ||W_{:k}^{(t)\top}P(W-W^{(t)})||$. By Lemma 9 we have $||W_k^{(t)}|| \lesssim \sqrt{K/n}$, so

$$
\begin{aligned}
\max_{k\in[K]} ||W_{:k}^{(t)\top}P(W-W^{(t)})|| &\le \max_{k\in[K]} ||W_{:k}^{(t)\top}Z||||\Pi Z^\top(W-W^{(t)})|| \\
&\lesssim ||\Pi Z^\top(W-W^{(t)})||_F \\
&\lesssim \sqrt{K}\max_{k\in[K]} ||(W_{:k}^{(t)} - W_{:k})^\top Z\Pi|| \\
&\lesssim \frac{K^2\sqrt{p_{max}}}{n^{1.5}\Delta_{min}} l(z,z^{(t)}). \qquad \text{(by the proof of part 2)}
\end{aligned}
$$

We now give an upper bound for $\max_{k\in[K]} ||W_{:k}^{(t)\top}E(W-W^{(t)})||$. By triangle inequality,

$$
||W_{:k}^{(t)\top}E(W-W^{(t)})|| \le ||W_{:k}^\top E(W-W^{(t)})|| + ||(W_{:k}^{(t)} - W_{k:})^\top E(W-W^{(t)})||.
$$

First we have

$$
\begin{aligned}
||W_{:k}^\top E(W-W^{(t)})|| &\le ||W_{:k}||||E||||(W-W^{(t)})|| \\
&\lesssim \frac{K^2\sqrt{p_{max}}}{n^{1.5}\Delta_{min}^2} l(z^{(t)},z).
\end{aligned}
$$

On the other hand, we also have

$$
\begin{aligned}
||(W_{:k}^{(t)} - W_{k:})^\top E(W-W^{(t)})|| &\le ||W_{:k} - W_{:k}^{(t)}||||E(W-W^{(t)})|| \\
&\le ||E||\sqrt{K}\max_k ||W_{:k} - W_{:k}^{(t)}||^2 \\
&\lesssim \frac{K^{3.5}\sqrt{np_{max}}l(z^{(t)},z)}{n^2\Delta_{min}^2}\frac{l(z^{(t)},z)}{n\Delta_{min}^2} \\
&\lesssim \frac{K^{3.5}\sqrt{p_{max}}l(z^{(t)},z)}{n^{1.5}\Delta_{min}^2}
\end{aligned}
$$

where the last inequality comes from the fact that by assumption $l(z,z^{(t)}) \le \tau \le \epsilon\frac{n\Delta_{min}^2}{K}$.

Thus it follows that

$$
\max_{k\in[K]} ||W_{:k}^{(t)\top}P(W-W^{(t)})|| \lesssim \frac{K^2\sqrt{p_{max}}l(z^{(t)},z)}{n^{1.5}\Delta_{min}}.
$$

Proof of 4. First note that

$$\begin{aligned}||(W_{:k}^{(t)} - W_{:k})^\top PW^{(t)}|| &\le ||(W_{:k}^{(t)} - W_{:k})^\top Z\Pi|| ||Z^\top W^{(t)}|| \\ &\lesssim \frac{K^{1.5}\sqrt{p_{max}} l(z^{(t)}, z)}{n^{1.5}\Delta_{min}}.\end{aligned}$$

Furthermore, by the same argument as before,

$$\begin{aligned}||(W_{:k}^{(t)} - W_{:k})^\top EW^{(t)}|| &\le ||(W_{:k}^{(t)} - W_{:k})^\top E(W^{(t)} - W)|| + ||(W_{:k}^{(t)} - W_{:k})^\top EW|| \\ &\lesssim K||E|| \max_k ||W_{:k} - W_{:k}^{(t)}||^2 + \frac{K^2\sqrt{p_{max}}}{n^{1.5}\Delta_{min}^2} l(z^{(t)}, z) \\ &\lesssim \frac{K^2\sqrt{p_{max}}}{n^{1.5}\Delta_{min}^2} l(z^{(t)}, z).\end{aligned}$$

We obtain the result by triangle inequality.

Proof of 5. Since $Z^\top W = I_K$ we have

$$\begin{aligned}||Z^\top W^{(t)}|| &\le 1 + ||Z^\top (W^{(t)} - W)|| \\ &\lesssim 1 + ||I - Z^\top W^{(t)}|| \\ &\lesssim 1 + K\frac{l(z^{(t)}, z)}{n\Delta_{min}^2} && \text{(by Equation (D.1))} \\ &\lesssim 1. && \text{(by assumption)}\end{aligned}$$

□

Lemma 11. *For* `sIR-LS` *we have with probability at least* $1 - n^{-\Omega(1)}$,

$$\max_{k\in[K]} |n_k^{(t)} - n_k| \lesssim \frac{K\sqrt{K}}{n\Delta_{min}^2} l(z^{(t)}, z), \quad |\lambda^{(t)} - \lambda| \lesssim \lambda\frac{Kl(z^{(t)}, z)}{n\Delta_{min}^2}.$$

Proof. First observe that

$$\max_{k\in[K]} |n_k^{(t)} - n_k| = \max_{k\in[K]} ||\mathbf{1}_n^\top (W_{:k}^{(t)} - W_{:k})|| \le \sqrt{n} \max_{k\in[K]} ||W_{:k}^{(t)} - W_{:k}|| \le \frac{K\sqrt{K}}{n\Delta_{min}^2} l(z^{(t)}, z)$$

by Lemma 10.

Then note that we have

$$\begin{aligned}|p_{max}^{(t)} - p_{max}| &\le \max_{k,k'} ||(W_{:k}^{(t)})^\top AW_{:k'}^{(t)} - W_{:k}^\top PW_{:k'}|| && \text{(the max is 1-Lipschitz)} \\ &\le \max_{k,k'} (||(W_{:k}^{(t)})^\top EW_{:k'}^{(t)}|| + ||(W_{:k}^{(t)} - W_{:k})^\top PW_{:k'}^{(t)}|| + ||W_{:k}^\top P(W_{:k'}^{(t)} - W_{:k'})||) \\ &\lesssim \max_k ||W_{:k}^{(t)}||^2||E|| + \frac{K^{1.5}\sqrt{p_{max}} l(z^{(t)}, z)}{n^{1.5}\Delta_{min}} && \text{(by the proof of Lemma 10)} \\ &\lesssim \frac{K\sqrt{p_{max}}}{\sqrt{n}} + \frac{K^{1.5}\sqrt{p_{max}} l(z^{(t)}, z)}{n^{1.5}\Delta_{min}} \\ &\lesssim \frac{K^{1.5}\sqrt{p_{max}} l(z^{(t)}, z)}{n^{1.5}\Delta_{min}} \\ &\lesssim \sqrt{\frac{K}{np_{max}}} \frac{Kl(z^{(t)}, z)}{n\Delta_{min}} p_{max} \\ &\lesssim \frac{Kl(z^{(t)}, z)}{n\Delta_{min}^2} p_{max}\end{aligned}$$

since $\Delta_{min} \asymp \sqrt{\log n/K}$. Consequently,

$$\begin{aligned}
\left|\frac{\lambda^{(t)}}{\lambda} - 1\right| &\leq \left|\frac{n_{min}^{(t)}}{n_{min}}\frac{p_{max}}{p_{max}^{(t)}} - 1\right| \\
&\leq \left|\frac{n_{min}^{(t)} - n_{min}}{n_{min}}\frac{p_{max}}{p_{max}^{(t)}}\right| + \left|\frac{p_{max} - p_{max}^{(t)}}{p_{max}^{(t)}}\right| \\
&\lesssim \frac{\sqrt{K}}{n^2\Delta_{min}^2} l(z^{(t)}, z) + \frac{Kl(z^{(t)}, z)}{n\Delta_{min}^2} \\
&\lesssim \frac{Kl(z^{(t)}, z)}{n\Delta_{min}^2}.
\end{aligned}$$

□

Lemma 12. *For* `IR-LSS`*, we have*

$$|\lambda^{(t)} - \lambda| \lesssim \lambda \frac{Kl(z^{(t)}, z)}{n\Delta_{min}^2}$$

with probability at least $1 - n^{-\Omega(1)}$.

Proof. By a similar argument as the one used in Lemma 11 we have with probability at least $1 - n^{-\Omega(1)}$ that

$$|p^{(t)} - p| \lesssim \frac{Kl(z^{(t)}, z)}{n\Delta_{min}^2} p, \quad |q^{(t)} - q| \lesssim \frac{Kl(z^{(t)}, z)}{n\Delta_{min}^2} q.$$

This implies

$$\frac{p^{(t)} - q^{(t)}}{p - q} = 1 + O\left(\frac{(p+q)Kl(z^{(t)}, z)}{(p-q)n\Delta_{min}^2}\right) = 1 + O\left(\frac{Kl(z^{(t)}, z)}{n\Delta_{min}^2}\right)$$

because $p - q \gtrsim p$. Thus,

$$\left|\log\left(\frac{p^{(t)}}{q^{(t)}}\right) - \log\left(\frac{p}{q}\right)\right| = \left|\log\left(\frac{p^{(t)}}{p}\frac{q}{q^{(t)}}\right)\right| = 2\log\left(1 + O\left(\frac{Kl(z^{(t)}, z)}{n\Delta_{min}^2}\right)\right) = O\left(\frac{Kl(z^{(t)}, z)}{n\Delta_{min}^2}\right).$$

Hence

$$\left|\frac{\log(\frac{p^{(t)}}{q^{(t)}})}{\log(\frac{p}{q})} - 1\right| = O\left(\frac{K^{1.5}l(z^{(t)}, z)}{n\Delta_{min}^2}\right)$$

since $\log(p/q)$ is bounded above by assumption. Consequently,

$$\frac{\lambda^{(t)}}{\lambda} - 1 = O\left(\frac{Kl(z^{(t)}, z)}{n\Delta_{min}^2}\right).$$

□

E Additional experiments

E.1 Heterophilic SBM

If we disregard the covariates, our algorithm can be used for inference on SBM and extend the method proposed by Lu and Zhou (2016) to general the general setting. In particular our algorithm also works for networks with heterophilic communities. The following experiment illustrate the gain in term of accuracy provided `IR-LS` initialized with `A-SC`, spectral clustering on the adjacency matrix. It also shows the interest of using more than one iteration in the refinement step with the MAP (this corresponds to `IR-MAP(1)`).

The model used for this experiment is the following: $n = 1000, K = 3, Z_i \overset{i.i.d}{\sim} \text{Multinomial}(1; 1/3, 1/3, 1/3)$ and

$$\Pi = \begin{pmatrix} 0.2 & 0.05 & 0.1 \\ 0.05 & 0.15 & 0.05 \\ 0.1 & 0.05 & 0.03 \end{pmatrix}.$$

We took an average of the NMI over 40 repetitions. We also considered the VEM algorithm implemented in the R package `blockmodels` (Léger, 2016), but the running time was prohibitive (approximately one hour whereas our algorithm take few seconds). It nevertheless returned the exact partition as `IR-LS`.

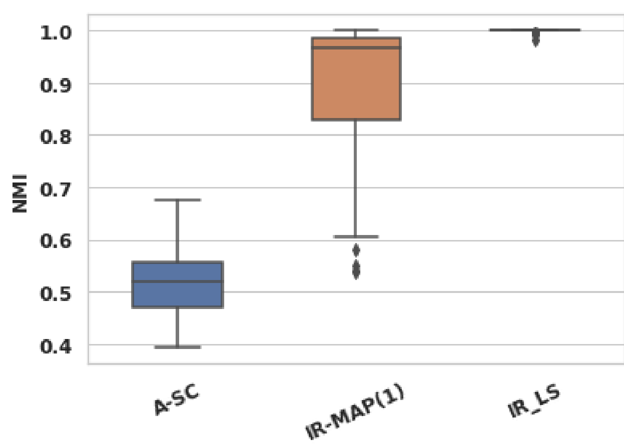

Figure 4: Average performances of different algorithms on an heterophilic SBM sorted by mean.

E.2 Australian Rainfall

We reproduced the experiment presented in Section 5 for $K = 10$ where the effect of `sIR-LS` is more visible. Figure 5 shows more clearly than $K = 5$ that `sIR-LS` increases the size of the smallest communities obtained with `Sponge-sym`, strengthen the community structure on the adjacency matrix and provide more localised clusters. On the other hand, `IR-LS` reduced the size of some small clusters (here we used 10 iterative refinement steps, but if we increased this number, some clusters could disappear), increase the size of other clusters and almost merge two clusters. Similar observations can be made for other choice of K as shown in Figure 6 and Figure 7.

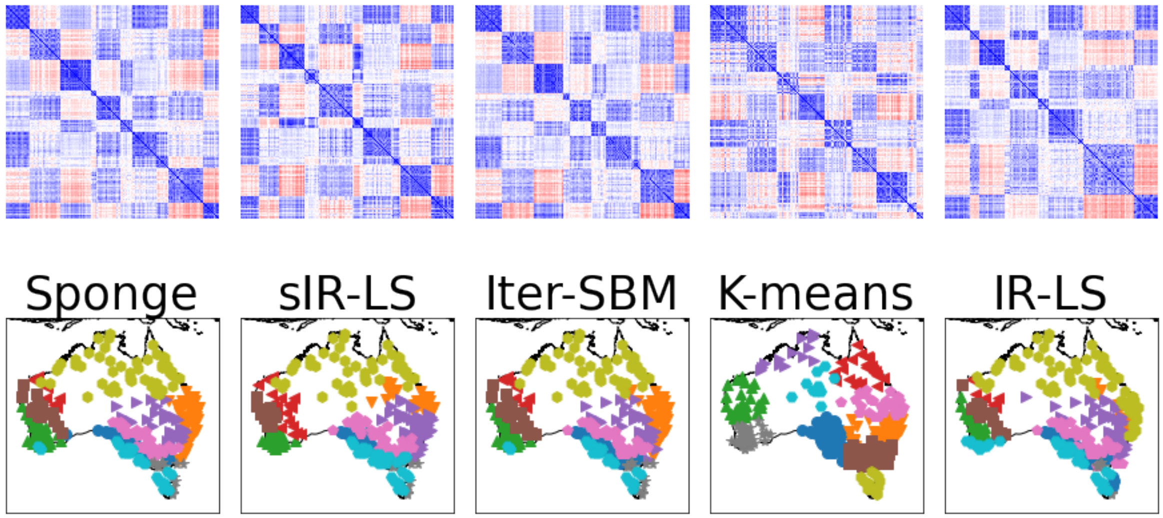

Figure 5: Sorted adjacency matrices of the Australian rainfall data set and corresponding maps for $K = 10$.

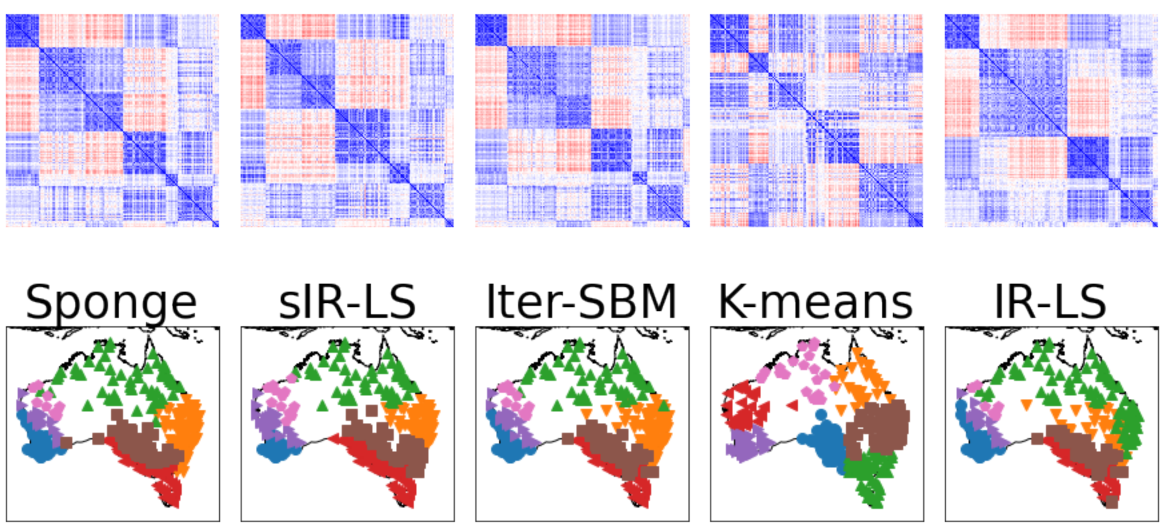

Figure 6: Sorted adjacency matrices of the Australian rainfall data set and corresponding maps for $K = 7$.

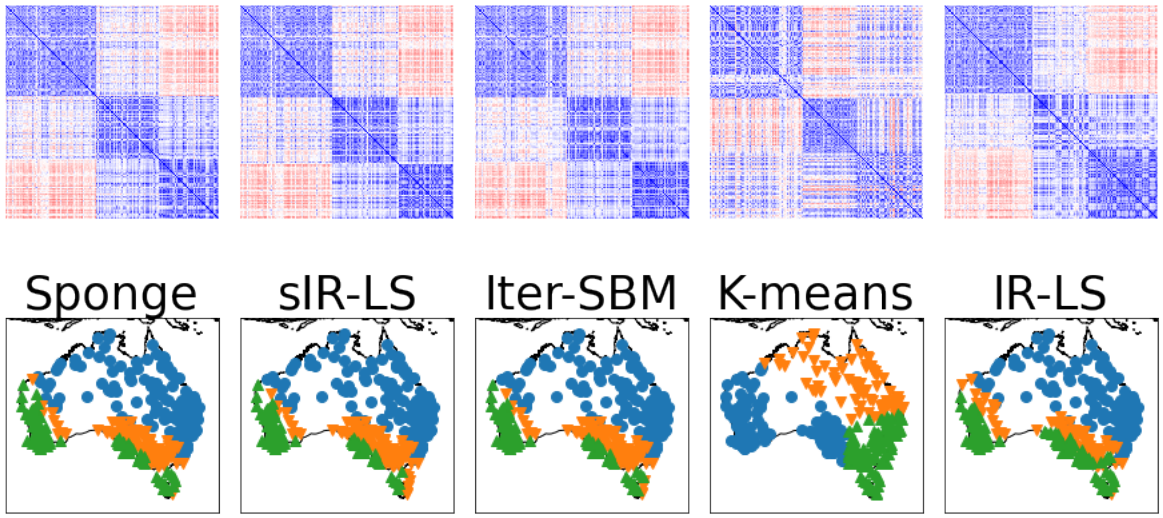

Figure 7: Sorted adjacency matrices of the Australian rainfall data set and corresponding maps for $K = 3$.

E.3 Signed SBM

We reproduced the experiment presented in Section 5.2 at a different sparsity level $p = 0.03$. The performances of the methods are similar to the ones previously observed. It seems that there is a threshold above which no algorithm can succeed. We conjecture that `IR-SSBM` is optimal an attain this threshold.

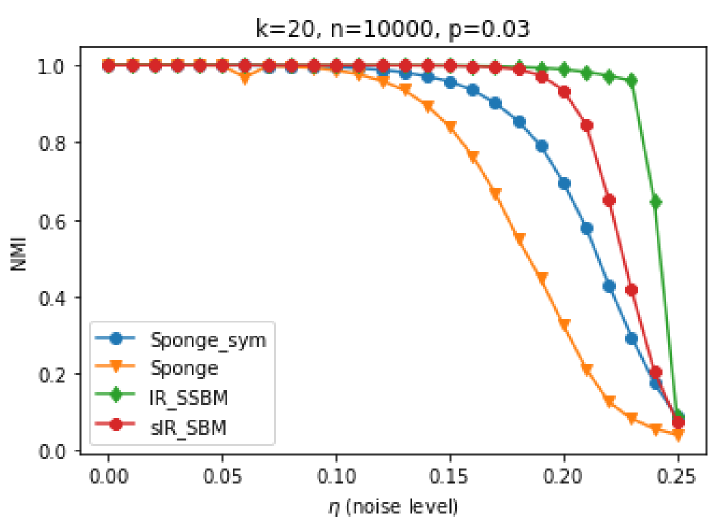

Figure 8: NMI for variying noise η.

Algorithm 3 `IR-SSBM`

Input: The number of communities K, initial partition $z^{(0)}$, $T \geq 1$.

for $0 \leq t \leq T-1$ **do**

  Compute $W^{(t)} = Z^{(t)}(D^{(t)})^{-1}$ where $D^{(t)} = \operatorname{diag}(n_k^{(t)})_{k \in [K]}$, and $C^{(t)} = AW^{(t)}$.

  Update the partition for each $i \leq n$

$$z_i^{(t+1)} = \arg\max_k C_{ik}^{(t)}$$

end for

Output: A partition of the nodes $z^{(T)}$.

E.4 Not distinguishable community

We repeated the first experiment of Section 5 with the rank deficient connectivity matrix

$$\Pi = 0.02 * \begin{pmatrix} 1.5 & 1.5 & 0.5 \\ 1.5 & 1.5 & 0.5 \\ 0.05 & 0.05 & 1.5 \end{pmatrix}$$

and the same covariates parameters. Without surprise, we obtained similar results. The main difference is that the performances of `ORL-SC` worsened.

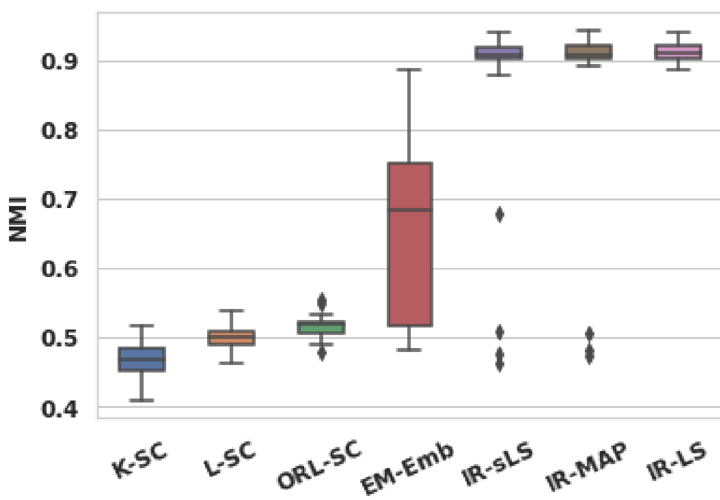

Figure 9: Performance of different algorithms on CSBM. Results are sorted by mean NMI and obtained over 40 runs.